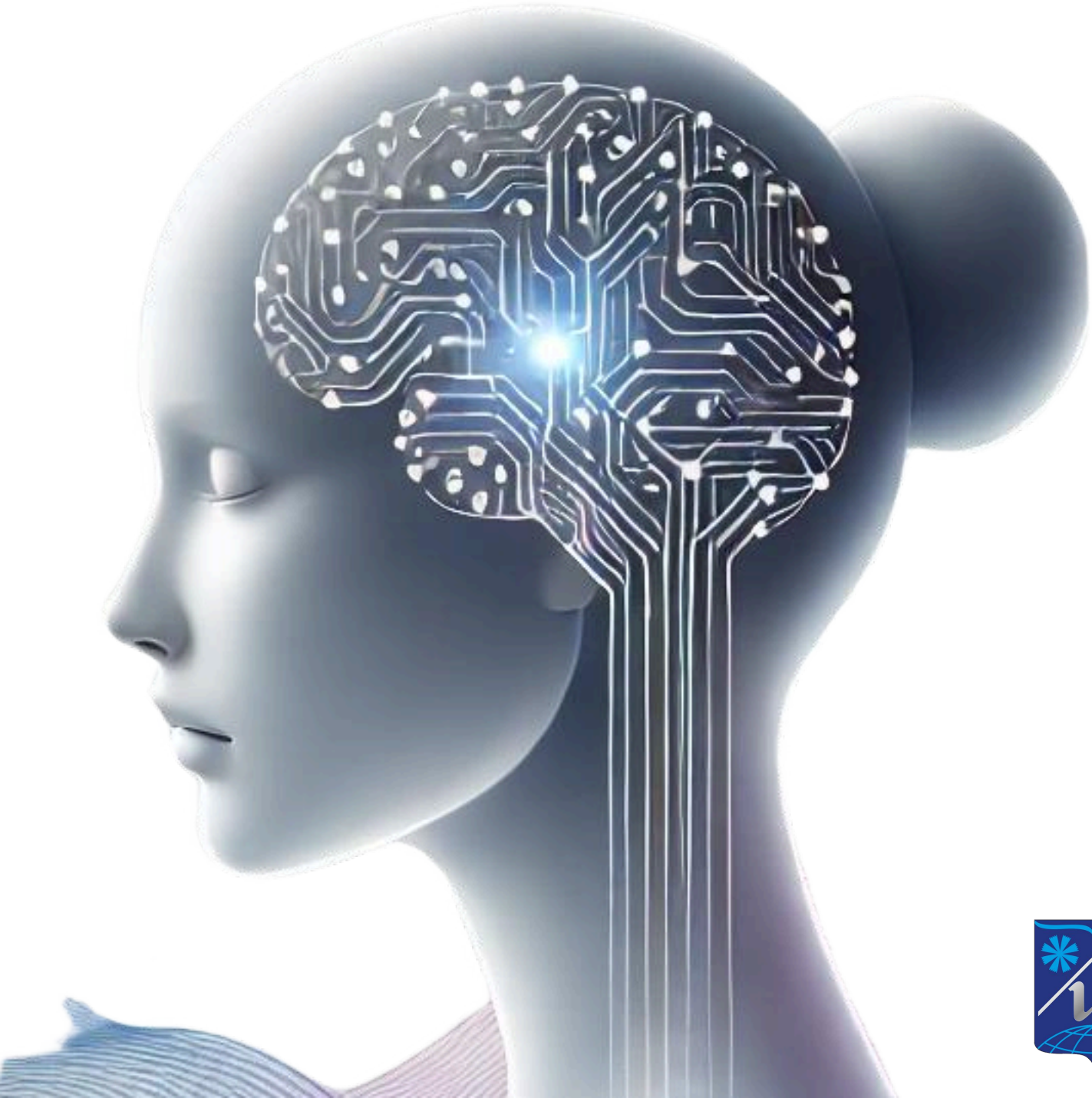
# BIG DATA
## And Advanced Analytics

XI Международная научно-практическая конференция



Министерство образования Республики Беларусь
Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»
Национальная академия наук Беларуси
Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси»
Иранская исследовательская организация по науке и технологиям
Ташкентский университет информационных технологий имени Мухаммада ал-Хоразмий
МИРЭА – Российский технологический университет
Университет инженерных технологий Туркменистана имени Огуз-хана
ФГАОУ ВО «Северо-Кавказский федеральный университет»
Евразийский национальный университет имени Л. Н. Гумилева

# BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ

# BIG DATA AND ADVANCED ANALYTICS

СБОРНИК НАУЧНЫХ СТАТЕЙ
XI МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

(Республика Беларусь, Минск, 23–24 апреля 2025 года)

Минск БГУИР 2025

УДК 004.65(082)
ББК 32.972.134
Б59

**Редакционная коллегия:**

**В. А. Богуш** – ректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р физ.-мат. наук, профессор, Республика Беларусь;

**С. К. Дик** – заведующий кафедрой инженерной и компьютерной графики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. физ.-мат. наук, доцент, Республика Беларусь;

**Д. В. Лихачевский** – декан факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент, Республика Беларусь;

**Т. В. Казак** – заведующий кафедрой инженерной психологии и эргономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р психол. наук, профессор, Республика Беларусь;

**Г. А. Пискун** – заместитель декана факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент, Республика Беларусь

**Рецензенты:**

**Т. Б. Тебуева** – заведующий кафедрой прикладной математики и компьютерной безопасности ФГАОУ ВО «Северо-Кавказский федеральный университет» (г. Ставрополь), д-р физ.-мат. наук, доцент, Российская Федерация;

**Д. А. Тусупов** – заведующий кафедрой информационных систем Евразийского национального университета имени Л. Н. Гумилева, д-р физ.-мат. наук, профессор, Республика Казахстан;

**Д. В. Лихачевский** – декан факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент, Республика Беларусь;

**А. Косари** – доцент кафедры электротехники и информационных технологий Иранской исследовательской организации по науке и технологиям, канд. техн. наук, доцент, Исламская Республика Иран

**BIG DATA** и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. XI Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 23–24 апреля 2025 года) / редкол.: В. А. Богуш [и др.]. – Минск : БГУИР, 2025. – 498 с.

Б59

ISBN 978-985-543-814-5.

Опубликованы результаты научных исследований и разработок в области BIG DATA and Advanced Analytics для оптимизации IT- и бизнес-решений, а также тематических исследований в области медицины, образования и экологии.

**УДК 004.65(082)**
**ББК 32.972.134**

**ISBN 978-985-543-814-5**

© УО «Белорусский государственный университет информатики и радиоэлектроники», 2025

# ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

**Председатель**
**Богуш В. А.,** ректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р физ.-мат. наук, профессор

## ЧЛЕНЫ ОРГАНИЗАЦИОННОГО КОМИТЕТА

**Араш Косари,** доцент кафедры электротехники и информационных технологий Иранской исследовательской организации по науке и технологиям, канд. техн. наук, доцент, Исламская Республика Иран

**Батура М. П.,** научный руководитель НИЛ 8.1 БГУИР, д-р техн. наук, профессор, академик Международной академии наук высшей школы, заслуженный работник образования Республики Беларусь

**Давыдов М. В.,** первый проректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент

**Джуманов Ж. Х.,** профессор Ташкентского университета информационных технологий имени Мухаммада ал-Хоразмий, д-р техн. наук, профессор, Республика Узбекистан

**Дик С. К.,** заведующий кафедрой инженерной и компьютерной графики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. физ.-мат. наук, доцент

**Казак Т. В.,** заведующий кафедрой инженерной психологии и эргономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р психол. наук, профессор

**Кругликов С. В.,** генеральный директор государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси», д-р воен. наук, канд. техн. наук, доцент

**Кудж С. А.,** ректор МИРЭА – Российский технологический университет, д-р техн. наук

**Лихачевский Д. В.,** декан факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент

**Мезилов Г. А.,** ректор Университета инженерных технологий Туркменистана имени Огуз-хана, д-р техн. наук

**Пархименко В. А.,** заведующий кафедрой экономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. экон. наук, доцент

**Пискун Г. А.,** заместитель декана факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент

**Стемпицкий Р. А.,** проректор по научной работе учреждения образования «Белорусский государственный университет информатики радиоэлектроники», канд. техн. наук, доцент

**Тебуева Т. Б.,** заведующий кафедрой прикладной математики и компьютерной безопасности ФГАОУ ВО «Северо-Кавказский федеральный университет» (г. Ставрополь), д-р физ.-мат. наук, доцент, Российская Федерация

**Тузиков А. В.,** заведующий лабораторией математической кибернетики государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси», д-р физ.-мат. наук, профессор

**Тусупов Д. А.,** заведующий кафедрой информационных систем Евразийского национального университета имени Л. Н. Гумилева, д-р физ.-мат. наук, профессор, Республика Казахстан

**Ульянов Н. И.**, декан факультета компьютерных систем и сетей учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент

**Шнейдеров Е. Н.,** проректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент

# ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ

Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»

Национальная академия наук Беларуси

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси»

Ministry of Science, Research and Technology (MSRT)
Iranian Research Institute for Information Science and Technology (IranDoc)

Иранская исследовательская организация по науке и технологиям, Исламская Республика Иран

TASHKENT UNIVERSITY OF INFORMATION TECHNOLOGIES

Ташкентский университет информационных технологий имени Мухаммада ал-Хоразмий, Республика Узбекистан

РТУ МИРЭА

МИРЭА – Российский технологический университет, Российская Федерация

Университет инженерных технологий Туркменистана имени Огуз-хана, Туркменистан

СКФУ

ФГАОУ ВО «Северо-Кавказский федеральный университет», Российская Федерация

EURASIAN NATIONAL UNIVERSITY

Евразийский национальный университет имени Л. Н. Гумилева, Республика Казахстан

# ОГЛАВЛЕНИЕ

**E.K. Samandarov** Issues and modern methods of intellectual data analysis in education ..... 11

**D.M. Rahel** Grouping and classification features of financial data certain types ..... 14

**S.S. Beknazarova** Processing object in a video stream using Big Data ..... 18

**Н.В. Щербина** Big Data в повышении промышленной безопасности ..... 27

**A.J. Seytov, O.N. Abduraxmonov** Применение больших данных при моделировании водохозяйственных объектов ..... 31

**D.M. Rahel** Evaluation of indicators clustering methods in interrelated counterparties financial assessment ..... 35

**Н.В. Павлов** Опыт применения в образовании чат-ботов, основанных на большой языковой модели ..... 38

**И.П. Кобяк** Математическое ожидание распределения вероятностей пропуска ошибки при наблюдении векторов переходов ..... 42

**И.В. Тимошкевич** Интеллектуальные системы мониторинга инженерных сетей: применение интернета вещей и аналитики больших данных для предотвращения протечек в многоквартирных домах ..... 48

**П.Н. Бибило, Н.А. Кириенко, В.И. Романов** Экспериментальное исследование новых алгоритмов выделения подсистем из совместных многоуровневых представлений систем булевых функций ..... 54

**Е.А. Алуев** Преодоление проблем Big Data с помощью Data Lake ..... 66

**Е. Клицунова, М.М. Лукашевич** Сравнительный анализ методов балансировки данных для задач машинного обучения ..... 74

**П.Ю. Бранцевич** Декомпозиция цифровых сигналов электроэнцефалограмм при их сравнительном анализе ..... 85

**M.B. Bekiyeva, G.O. Orazdurdyyeva** Statistical method and graph theory-based algorithm for detecting network traffic ..... 92

**Г.Г. Турантаева, Н.С. Давыдова, В.Е. Васюк, К.И. Татарко** Алгоритм коррекции траектории движения спортсменов на базе матриц поворота ..... 97

**В.В. Шаталова, Т.В. Казак** Big Data в HR-аналитике ..... 103

**Yi Tang** Analysis and review of multimodal trajectory prediction methods in complex dynamic scenes: evolution from classical models to deep learning ..... 108

**Е.А. Бугаев** Использование метода ранжирования при определении списка потенциальных знакомых пользователя ..... 117

**Н.Ч. Реджепмурадова** Искусственный интеллект в образовании ..... 124

**В.В. Синицына** Технологии больших данных в обработке изображений для людей с ахроматопсией ..... 131

**М.Т. Мырадов** Архитектура Transformer для преобразования голоса в текст ..... 135

**П.С. Мырадов** Big Data и продвинутый анализ данных: современные тенденции, технологии и перспективы развития ..... 139

**А.Х.Оразгелдиев, О.Д.Чарыева, Д. Я.Караджаева** Нейронные сети: математические модели, используемые для решения сложных задач, таких как распознавание образов, классификация и прогнозирование и методы обработки и анализа звуковых данных,

включая распознавание речи, классификацию звуков, выделение признаков ..... 142

**С.С. Шайымов, А.Б. Чарыев** Технологии Big Data в автоматизации технологических процессов ..... 147

**А.Д. Русецкий** Принятие решений на основе данных в сельском хозяйстве ..... 152

**В.В. Дрозд, Е.А. Холод, В.В. Гоза, Я.В. Кипцевич** Способ предоставления данных V2X для дальнейшей обработки системами искусственного интеллекта ..... 161

**К.И. Свистунова, С.В. Абламейко** Классификация глазных заболеваний по снимкам глазного дна с помощью нейронной сети Efficientnet ..... 168

**В.А. Вишняков, И. Ся** Аппаратная поддержка обработки данных в ИТ диагностике ..... 174

**В.А. Вишняков, Ч. Юй** Подход к аппаратной поддержке данных текста с использованием ПЛИС ..... 178

**В.К. Раюшкин** Большие данные (Big Data) как совокупность объектов гражданских прав ..... 182

**И.В. Марахина** Воронка продаж как инструмент веб-аналитики ..... 187

**С.С. Шайымов, А.Б. Чарыев, Б.Н. Атанепесов** Нейротехнологии и Big Data в образовании ..... 190

**B.A. Al-Nami** The role of Big Data in graphic design and user interfaces a data-driven approach from idea to implementation ..... 194

**Л.И. Архипова** Бизнес-экосистемы и их роль в цифровой трансформации ..... 202

**Д.И. Черемисинов, Л.Д. Черемисинова** Отладка параллельных программ решения логико-комбинаторных задач ..... 208

**O.M. Abbasova** Big Data from multi-omics technologies is advansing medicine and healthcare ..... 215

**Т.В. Казак, А.Н. Василькова** Возможные перспективы генеративного искусственного интеллекта в Республике Беларусь ..... 220

**М.Б. Таджиева** Раскрытие потенциала искусственного интеллекта для аналитики больших данных ..... 229

**Ю.Р. Пигарева, Б.А. Аль-Нами** Влияние больших данных на создание интерфейсов ..... 234

**О.М. Аббасова, Г.Г. Аннасапаров** Роль больших данных в анализе геномных данных ..... 239

**В.В. Венгеренко, М.М. Лукашевич** Выбор функции активации и функции потерь для задачи классификации активностей человека ..... 243

**Э.Аралиев** Оптимизация IT-инфраструктуры с помощью Big Data: от мониторинга до прогнозирования сбоев ..... 253

**А.В. Деркач, И.И. Фролов** Исследование влияния входных параметров на классификацию изображений продуктов ..... 256

**Е.М. Косарева, А.И. Полуян** Применение больших данных в сфере проката автомобилей ..... 263

**С.Тилевов** Актуальность проблемы увеличения пропускной способности ВОЛС в связи с

ростом объема передаваемых данных ..... 267

**А. В. Сидоренко, Н.А. Солодухо, П.П. Королевич, М.М. Скугаревская** Анализ нелинейных параметров электроэнцефалограмм женщин при патологии в виде депрессии ..... 269

**И.А. Бабкин, Е.И. Алаторцев** Аспекты применения методов глубокого обучения для оптимизации энергопотребления на нефтеперерабатывающем заводе ..... 276

**М.Д. Михальков, С.Н. Нестеренков** Модель интеллектуальной системы поддержки принятия решений и процесс ее интеграции с персонализированной рекомендательной системой ..... 281

**Н.А. Шиманский, А.В. Баглов, Л.С. Хорошко** Статистический анализ и моделирование свойств наноматериалов методами large language lodels, agentic AI & machine learning ... 290

**В.С. Князькова** Методические подходы к оценке эффективности внедрения технологий Big Data ..... 296

**В.Ф. Алексеев, Д.В. Лихачевский, Г.А. Пискун** Оценка эффективности генеративной политики при оптимизации проектных решений ..... 300

**Ф.М. Алимова** Роль KDD в развитии инструментов анализа и управления данными ..... 310

**А.Г. Бокун, А.В. Герчик** Оптимизация времени сбора данных о транзакциях в сети Solana ..... 316

**В.Л.Кожуховский, А.С.Солонар, А.Ю.Брынчикова** Признаки радиолокационного распознавания баллистических объектов ..... 323

**Е.И. Ковшер, Н.В. Лапицкая** Оценка значимости параметров обучения модели нейронной сети на примере классификатора объектов на снимках земной поверхности ..... 329

**А.Н. Голубович, П.В. Камлач, А.В. Чураков, В.М. Бондарик** Модель сложных биологических тканей для исследования воздействия электрического тока ..... 336

**Guo Qicheng** Vehicle and label detection method based on yolov11 ..... 340

**Д.В. Краснопрошин, М.И. Вашкевич** Анализ подходов к построению систем распознавания эмоций по речи с использованием методов глубокого обучения ..... 343

**Д.А. Макар, М.И. Вашкевич** Анализ подходов к построению автоматизированных систем чтения по губам ..... 354

**А.А. Навроцкий, А.О. Шведко, В.И. Сакович** Проблематика машинного обучения: влияние гиперпараметров, переобучение и недообучение нейронных сетей ..... 361

**L.I. Arkhipova** Digital maturity categories in business: integrating and assessment ..... 365

**Т.А. Ализаде, А.А. Мустафаев** Обработка больших массивов измерительной информации с применением параллельных вычислений ..... 372

**В.В. Верняховская, О.М. Раптунович** Инновационные подходы к анализу больших данных: от теории к практике ..... 380

**А.Д. Джейранян, Л.Н. Лядова** Преодоление ограничений современных инструментов визуализации данных: новая методология ..... 386

**А.А. Карпук, Л.С. Лазута** Выделение и поиск функциональных зависимостей между атрибутами в системах Big Data ..... 397

**E.A. Alooeff, M.V. Bulycheva** Leveraging Big Data and machine learning for optimizing business decisions in e-commerce ..... 406

**С. А. Берговин, А. Ю. Саливон, Ф.В. Усенко, А.М. Прудник** Обработка и распознавание геометрических паттернов в игровых данных ..... 412

**Т.А. Васяева, М.А. Рябко** Применение машинного обучения для персонализации адаптивного дизайна ..... 418

**В.Г. Евдокимов, А.А. Навроцкий** Адаптивная фильтрация многомерных данных для оптимизации бизнес-решений на основе нелинейных метрик взаимозависимости признаков ..... 424

**Д.А. Кечик, А.С. Кожухов, Ю.Д. Кацер** Развитие подхода к оценке параметров механической системы методами машинного обучения ..... 430

**И.П. Логинова** Построение графической оболочки для проведения экспериментов на базе системы логической оптимизации функционально-структурных описаний дискретных устройств ..... 436

**В.С. Лянцевич, Н.В. Марковская** Анализ медицинских данных с помощью алгоритма Apriori на языке R ..... 452

**Ю.Е. Седельников, Д.А. Веденькин, А.Ф. Гильфанова** Применение мультифокусированных антенн в задачах медицинской диагностики ..... 460

**Л.Ю. Шилин, К.Д. Сморщок, Д.Д. Хмыз** Оптимизация алгоритмов глубокого обучения для повышения точности распознавания белорусских дорожных знаков в условиях дисбаланса данных ..... 464

**Yang Shixin, S.A. Pauliukavets, A.A. Velchenko** Autonomous mobile robot navigation and motion control problems and analysis of existing methods ..... 467

**Е.И. Лещевич, С.И. Сидоров, И.И. Ревинская, С.К. Дик, Г.Д. Ситник** Оценка качества передачи информации в системе диспетчеризации на базе MQTT-архитектуры ..... 477

**А.Н. Осипов, С.П. Урбанович, М.П. Батура, А.А. Ивановский, И.О. Хазановский, А.В. Короткевич** Автоматизированная система для проведения исследований медицинских имплантируемых приборов ..... 481

**В.Ф. Алексеев, Д.В. Лихачевский, Г.А. Пискун** Определение целевых функций и их ограничений в задачах оптимизации при генеративном проектировании ..... 487

**Г.П. Колбанов, Д.А. Романов, Е.С. Белоусова** Методика повышения безопасности дистрибутива ОС Linux с биометрической аутентификацией ..... 495

UDC 519.71

# ISSUES AND MODERN METHODS OF INTELLECTUAL DATA ANALYSIS IN EDUCATION

***E.K. Samandarov***
*PhD student at IFAR under TIIAME,*
*Tashkent, Uzbekistan*
samandaroverka09@gmail.com

***E.K. Samandarov***

*Graduated from national university of Uzbekistan named after Mirzo Ulugbek. The area of scientific interests is related to implementation the machine learning, artificial intelligence and big data for the educational sphere.*

**Abstract**. Currently, intelligent data analysis is used in many fields. Many methods are used in the intellectual analysis of data. In this article, the intellectual analysis of data in the educational system is considered. Educational data mining combines many methods of data analysis in education sphere. In addition, in this paper the most methods of data analysis are considered. The application of each method in the field of education is considered as well as in the field of education, data is intellectually analyzed and the results are analyzed.

**Keywords**: Big Data, artificial intelligence, machine learning, artificial neural network, educational data mining

**Introduction.** As a result of the rapid development of modern information technologies and the penetration of these technologies into all areas of society, a large amount of information (Big Data) has been created. This situation requires new solutions for analyzing large volumes of data, extracting the necessary information from it, and making decisions based on the extracted data. Intelligent data analysis is important in solving complex problems in various fields, based on modern technologies and methods. Technologies such as artificial intelligence (AI), machine learning, artificial neural networks, data science and deep data analysis (deep learning) are the most effective methods used to solve complex problems in this situation.

Intelligent analysis methods allow you to identify hidden relationships between data in a dataset, make predictions, and make decisions. These methods are used to solve problems in many fields, including business, economics, science and other fields. It is also widely used in the field of education. For example, in the field of education, through intellectual analysis, it is possible to analyze the level of knowledge of students, evaluate the academic results of students, organize personalized teaching processes according to the level of knowledge of each student, and predict the level of knowledge of students.

The intellectual analysis of data also has its own disadvantages. The large volume of data, the quality of data, their diversity, and the process of choosing analysis methods are among the problematic issues. This article examines the problems encountered in the field of intellectual analysis, the methods and technologies used to achieve them. Additionally, we will consider the application of these methods to the educational system, their results and prospects.

The importance of intellectual analysis of data is relevant not only in the development of science and technology, but also in all spheres of social life.

**Methods of intellectual analyze data.** We will consider in detail each of the methods highlighted in the above section and widely used in the field of education. Clustering method. This method is used in machine learning. Clustering divides data into groups based on data similarity. When applying this method to the field of education, it is possible to divide the interests of students and the level of knowledge of students into groups. As a result, it is possible to create customized training programs for groups that are grouped together. The next method is the classification method. Sorts data into predefined classes.

As for its application in the field of education, this method is used in situations such as dividing the knowledge level of students into high, average and low classes, and identifying students at risk of academic failure. Regression methods are used to make predictions based on data. In education, it is used in situations such as determining the aspects that affect the educational results, predicting the academic results in the future.

Association Rule Mining method detects hidden relationships between data. Based on the students' learning process, it is used to recommend suitable educational resources to them, to study the relationship between subjects according to the curriculum. Neural Networks is one of the most widely used methods. It is widely used in evaluating student performance, predicting student success, and classifying students. The key method is the Decision Tree algorithm. With the help of this algorithm, it is used to determine the main factors affecting the results of students' studies, to recommend which subjects students should master successfully. Big Data Analytics is a method of analyzing large amounts of data. It is a method of analyzing the data of a large amount of students related to the educational process. It is used in making strategic decisions for educational institutions. Principal Component Analysis (PCA). It is widely used in the educational system to determine the factors influencing students' mastery of subjects and to optimize educational programs. Machine Learning Algorithms is a component of artificial intelligence. In the field of education, ML algorithms are used to classify students' knowledge levels into classes, and to predict students' knowledge levels. In addition, it is used in the creation of personalized educational systems and the development of interactive educational resources. Similarity Analysis method is widely used in presenting educational materials in accordance with students' interests, and in offering advanced methods for increasing academic potential.

Educational Data Mining (EDM) is a branch of computer science focused on analyzing educational data sets, extracting useful knowledge and patterns from them using the methods discussed above. Using modern technologies and methods of intellectual data analysis, EDM provides an opportunity to understand, optimize and personalize the educational process according to the knowledge potential of each student. The main goal of EDM is to monitor and analyze student activity, personalize the learning process, and improve the quality of education. Technologies used in EDM are classification and clustering algorithms, forecasting techniques, association rules, machine learning and artificial intelligence, big data analytics. EDM can be used at all levels of education. EDM is mainly used in education in the following contexts: learning process analysis, personalized learning, academic risk identification, and resource management and optimization.

Through the methods discussed above, we can effectively use information in the educational process, optimize the educational process, and increase the possibilities of personalized education. Figure 1 illustrates educational data mining.

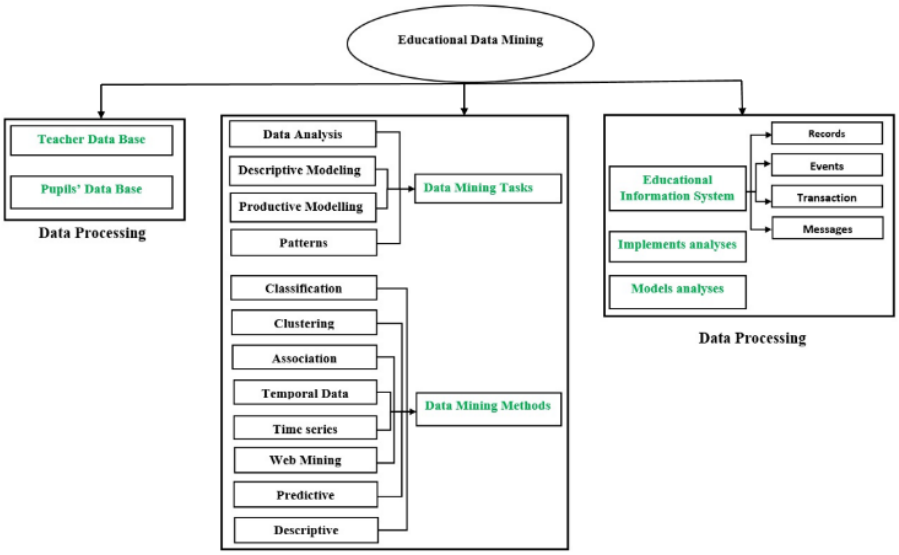



*Figure 1.* Educational Data Mining

**Intelektual tahlil qilish usullarini ta'lim tizimiga qo'llash va uning natijalari.** Many countries allocate millions of dollars annually to their education systems, representing a significant portion of their national budgets. Statistics show that global expenditure on school education is substantial. However, the cost of developing an educational platform using the proposed architecture would be several times lower than the yearly expenses of traditional offline education. Notably, the number of internet users worldwide is projected to rise by 47%, from 5.35 billion in 2024 to 7.9 billion in 2029. This growth enables students in internet-accessible areas to utilize such platforms based on artificial intelligence for learning. The such platforms are highly adaptable, making it suitable for all levels of education. An educational platform based on artificial intelligence can accurately assess student performance and address various educational challenges. Additionally, the platform proves particularly effective during pandemics.

**Conclusion.** Considering the above methods, we can come to the following conclusion regarding the intellectual analysis of data in education. The intellectual analysis of data in education is very important for the management of modern educational processes and the further development of the educational process, with the help of the above methods, the possibility of predicting the level of knowledge, interests and success of students and dividing them into classes increases. Algorithms of intellectual analysis provide recommendations on identifying the shortcomings of students in the educational process and eliminating the identified problems. There is a high possibility of making accurate predictions in the educational process through deep learning and machine learning methods.

## Reference list

[1] Samandarov E. The Architecture of educational platform based on machine learning //Journal of Mathematics, Mechanics and Computer Science. – 2024. – Т. 124. – №. 4. – С. 86-102.

[2] Romero C., Ventura S. Educational data mining: A survey from 1995 to 2005 //Expert systems with applications. – 2007. – Т. 33. – №. 1. – С. 135-146..

[3] Sekeroglu B., Dimililer K., Tuncal K. Student performance prediction and classification using machine learning algorithms //Proceedings of the 2019 8th international conference on educational and information technology. – 2019. – С. 7-11.

## Authors' contribution

**Erkaboy Samandarov** – analysis of the methods of intellectual analysis of data in the field of education, formation of the structure of the paper.

UDC 336.774.3

# GROUPING AND CLASSIFICATION FEATURES OF FINANCIAL DATA CERTAIN TYPES

***D.M. Rahel***
*PhD in Economics, Associate Professor of the Department of Economics of the Belarusian State University of Informatics and Radioelectronics*
ragel@bsuir.by

***D. M. Rahel***

*In 2000 he graduated from the Faculty of Economics of the Belarusian State University of Informatics and Radioelectronics with a degree in Economic Informatics. In 2016 he graduated from the postgraduate course of the Academy of Management under the President of the Republic of Belarus. In 2018 he defended his PhD thesis. Research interests: data mining in marketing, process modeling, data analysis, statistical forecasting, macroeconomics.*

**Annotation.** The article presents an approach to analyzing the distribution of various types of financial data. An approach is proposed that allows analyzing a set of similar data based on constructing a distribution series and analyzing values in different intervals of the series. The distribution of values will allow us to draw conclusions about the normality of the activities of the counterparties under study and, taking this into account, conduct further data analysis.

**Keywords:** data analysis, financial data, big data, data array, forecasting, analytics, distribution series, Sturges method, data distribution, financial characteristics, predictive analytics, financial analysis.

Based on the results of a sample study of 100 similar enterprises, data on the volume of their assets as of 09.01.2023 was obtained (Table 1). In order to operate with these data or draw initial conclusions regarding the business entities under consideration, it is necessary to carry out an initial classification in order to divide them into groups for further research. As the main solution for classifying the obtained data, you can use the construction of an interval variation series using the Sturges method.

Table 1. Volume of assets of 100 surveyed enterprises

| Initial data as of 09.01.2023, million rubles. | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0,86 | 0,77 | 0,54 | 2,22 | 2,7 | 2,26 | 2,996 | 2,29 | 2,25 | 2,98 |
| 0,97 | 4,14 | 1,77 | 2,31 | 2,42 | 2,07 | 2,41 | 6,55 | 2,29 | 2,44 |
| 0,74 | 2,11 | 3,61 | 2,99 | 2,64 | 2,32 | 2,28 | 2,25 | 2,2 | 2,93 |
| 2,97 | 2,51 | 2,94 | 2,1 | 2,61 | 2,45 | 2,44 | 2,19 | 2,37 | 3,44 |
| 3,55 | 2,92 | 4,1 | 3,12 | 2,93 | 2,55 | 4,64 | 3,34 | 3,12 | 3,22 |
| 2,34 | 1,97 | 0,91 | 4,34 | 2,92 | 2,91 | 5,55 | 4,44 | 4,65 | 5,15 |
| 2,71 | 1,86 | 0,95 | 2,9 | 2,65 | 2,98 | 4,36 | 6,41 | 6,1 | 5,16 |
| 2,94 | 2,94 | 0,76 | 2,99 | 2,44 | 2,39 | 4,32 | 5,33 | 4,33 | 3,34 |
| 1,89 | 2,14 | 0,61 | 1,33 | 2,5 | 2,42 | 3,37 | 5,59 | 4,11 | 4,92 |
| 1,45 | 2,89 | 1,12 | 1,56 | 2,38 | 2,94 | 3,02 | 5,91 | 4,69 | 4,2 |

To carry out a general classification and interval distribution of the obtained values, it is necessary to understand the width of the sample values, and then divide it into intervals that will allow drawing conclusions regarding the distribution of values and, on the basis of this, making a selection for further research. To do this, for the data array under consideration, we will determine the minimum value to determine the general lower limit of the entire considered interval, in this case Xmin = 0.54 million rubles. and maximum, as the upper limit for the considered set of values: Xmax = 6.55 million rubles.

Taking this into account, the range of variation in the data of the sample under consideration will be R=6.01. For further grouping and understanding the nature of the distribution, it is necessary to divide this segment into intervals. To split and determine the number of intervals, you can use the Sturges formula (1).

$$H = R/(1+log_2 n) = R/(1+3{,}322lgn) \quad (1)$$

where n is the sample size, in our case, we are talking about 100 observations.

In the case of our sample of values, we obtain the duration of the interval H = 0.7863, which indicates 8 intervals to which the values of our sample for the assets of the companies under consideration will be assigned.

Taking into account the obtained value of the interval length, we obtain the following set of intervals according to the proposed method of partitioning and determining their number (Table 2).

Table 2. Distribution of intervals for classification of the data

| Interval sequence number | Upper line | Bottom line |
|---|---|---|
| 1 | 0,147 | 1,326 |
| 2 | 1,326 | 2,113 |
| 3 | 2,113 | 2,899 |
| 4 | 2,899 | 3,685 |
| 5 | 3,685 | 4,471 |
| 6 | 4,471 | 5,258 |
| 7 | 5,258 | 6,044 |
| 8 | 6,044 | 6,83 |

The population is divided into intervals until the beginning of the next one is equal to or greater than the maximum value in the population of analyzed data. Taking this into account, in our case we got 8 intervals.

After ranking the values, we will determine how the values are distributed over the previously established intervals. Taking this into account, we will construct an interval variation series of the distribution of asset volumes of the business entities under consideration based on a number of mandatory characteristics (Table 3).

Table 3. Interval variation series of distribution of volumes of enterprise assets as of 09.01.2023, million rubles.

| Intervals | Frequency | Cumulative frequency | Relative frequency (frequency / n) | Relative cumulative frequency (cumulative frequency / n) | Frequency / interval duration |
|---|---|---|---|---|---|
| 0,147 – 1,326 | 10 | 10 | 0,1 | 0,1 | 0,13 |
| 1.326 - 2.113 | 10 | 20 | 0,1 | 0,2 | 0,13 |
| 2.113 - 2.899 | 32 | 52 | 0,32 | 0,52 | 0,41 |
| 2.899 - 3.685 | 26 | 78 | 0,26 | 0,78 | 0,33 |
| 3.685 - 4.471 | 9 | 87 | 0,09 | 0,87 | 0,11 |
| 4.471 - 5.258 | 6 | 93 | 0,06 | 0,93 | 0,076 |
| 5.258 - 6.044 | 4 | 97 | 0,04 | 0,97 | 0,051 |
| 6.044 - 6.83 | 3 | 100 | 0,03 | 1 | 0,038 |
| **TOTAL** | 100 | | 1 | | |

Based on the obtained values, for a more visual presentation of the results, we will construct a graph that will display the distribution of enterprises by intervals (Figure 1).

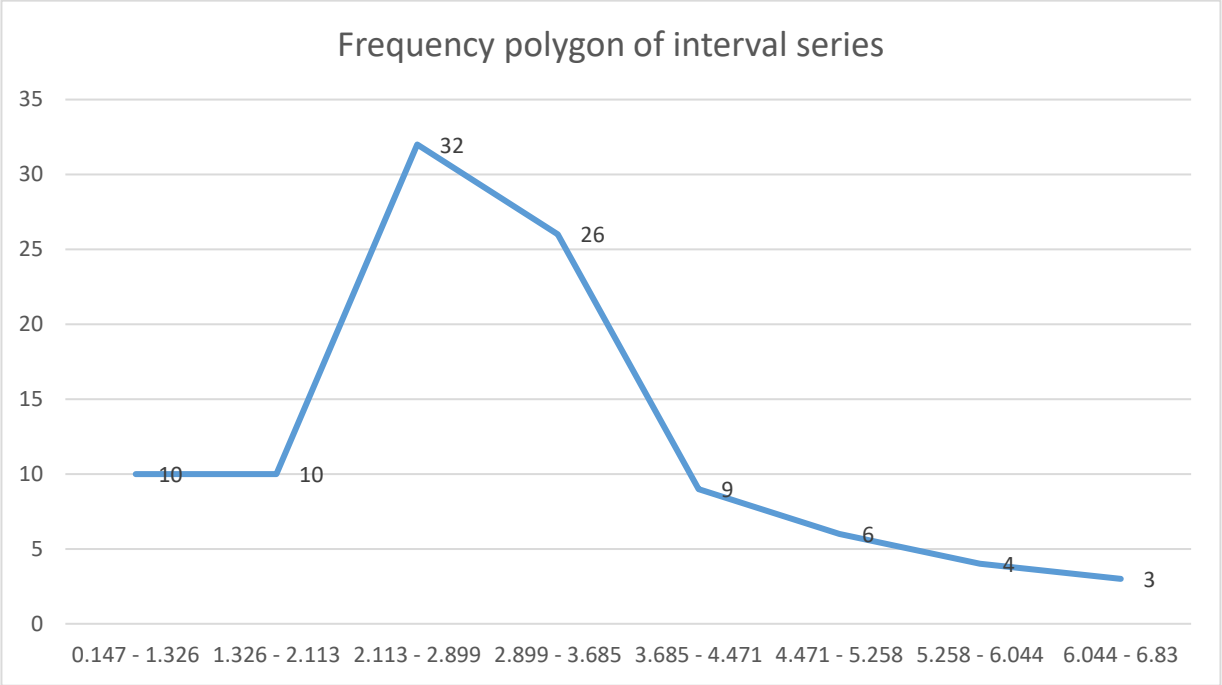



*Figure 1.* Final frequency distribution of the considered sample of values

Based on the results obtained, it can be noted that the largest number of enterprises under consideration belong to the average range of values, 32 enterprises in the range from 2.113 million rubles to 2.899 million rubles. and 26 enterprises in the range from 2.899 million rubles. up to 3.685 million rubles. The interval of maximum values includes three companies in the sample; the minimum interval includes 10 enterprises of the sample under consideration. 48 enterprises have

assets greater than the average value of 2.93 million rubles, which indicates a stable situation in the sample for the parameter under consideration.

## References

[1] Мыльников Л.А. Статистические методы интеллектуального анализа данных. – БХВ-Петербург, 2021. – 119 с.

[2] Data Science and Big Data Analytics. A Step by Step Guide to learn Data Science from Scratch with Python Machine Learning and Big Data / Andrew Park. – Published by Andrew Park, 2021. – 124 p.

[3] Nicholson W.L. Exploring Data Analysis. – Nobel Press, 2012. – 421 p.

# ОСОБЕННОСТИ ГРУППИРОВКИ И КЛАССИФИКАЦИИ НЕКОТОРЫХ ТИПОВ ФИНАНСОВЫХ ДАННЫХ

***Д.М. Рагель***
*к.э.н., доцент кафедры экономики*
*Белорусского государственного*
*университета информатики и*
*радиоэлектроники*

**Аннотация.** В статье изложен подход к анализу характера распределения различных типов финансовых данных. Предлагается подход, при котором можно проанализировать совокупность однотипных данных на основании построения ряда распределения и анализа значений в различных интервалах ряда. Характер распределения значений позволит сделать выводы о нормальности деятельности изучаемых контрагентов и, с учетом этого, провести дальнейший анализ данных.

**Ключевые слова:** анализ данных, финансовые данные, большие данные, массив данных, прогнозирование, аналитика, ряд распределения, правило Стерджеса, распределение данных, финансовые характеристики, прогнозная аналитика, финансовый анализ.

УДК 031.023:031.25

# PROCESSING OBJECT IN A VIDEO STREAM USING BIG DATA

***S.S. Beknazarova***
*Professor of the Department of Television and Media Technologies of M.Khorazmiy TUIT, Doctor of Technical Sciences, Professor*
saida.beknazarova@gmail.com

***С.С. Бекназарова***
*Окончила Ташкентский университет информационных технологий. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, обработкой медиаресурсов: видео, аудио, организацией учебного и научно-исследовательского процессов в техническом университете.*

**Аннотация.** The article is devoted to the study of methods for processing objects in a video stream using big data technologies. With the rapid growth of video data volumes and the need for operational analysis, traditional processing methods are becoming insufficient to effectively work with such information arrays. Approaches involving the use of distributed computing, machine learning algorithms, and real-time analytics to extract meaningful information from a video stream are considered. Special attention is paid to the issues of data collection, storage and processing, as well as the problems that arise when integrating big data technologies into video surveillance systems. The article discusses both technical and practical aspects, including the possibilities of using these technologies in various fields such as security, smart cities, and behavior monitoring. The results of the study may be useful for video surveillance system developers, data analysts, and researchers in the field of artificial intelligence.

**Ключевые слова:** big data, advanced analytics, video analytics.

**Introduction**. Modern video surveillance and video data processing technologies are rapidly developing, providing new opportunities for analyzing large amounts of information. One of such promising areas is the processing of objects in a video stream using big data analysis methods. Video streams are a series of images that can contain important information about the movement, behavior, and interaction of objects in various fields, from security and monitoring to medicine and commerce.

Video stream processing requires high-performance computing and efficient algorithms to extract and interpret data in real time. At the same time, as the amount of data generated by various video cameras increases, traditional processing methods become insufficient for adequate processing and analysis of such data. In this regard, the use of big Data technologies such as distributed computing, machine learning and real-time analytics opens up new horizons in the field of video stream processing.

The current stage of development of information and computing systems is characterized by the widespread introduction of multimedia technologies, and therefore there is a need to develop methods and algorithms for compressing digital video images and video streams based on processing groups of images with information redundancy. The ever-increasing computational complexity of video sequence processing algorithms and high data storage costs require significant

computing resources. Video compression not only makes it possible to use digital video in a transmission medium that does not support the original (uncompressed) video images, but also increases the efficiency of using high-speed communication channels for transmitting high-resolution video streams, including simultaneous transmission of multiple streams of high-quality video data[1-3].

Many other prominent scientists and specialists have also made a significant contribution to the theory of video data compression.

Existing video data processing systems use lossy compression methods due to the transmission of reference (I-frames) compressed by spatial coordinates, as well as reference (P- and B-frames) compressed by spatial-temporal coordinates, the number of which determines the degree of compression of the video sequence. At the same time, the peculiarities of the perception of this type of data by a person are taken into account. To date, the problem of eliminating the redundancy of reference frames has been solved to a certain extent, whereas this problem is relevant for reference frames [4-6].

The basis for encoding reference frames is the construction of motion vectors of individual pixels or groups of them. Also, the efficiency of algorithms for eliminating time redundancy is improved by increasing the efficiency of encoding algorithms by transformation and interpolation. This dissertation work is devoted to these problems.

The aim of the article is to improve existing and create new methods for eliminating information redundancy of video sequences, characterized by reduced computational costs with high compression ratio and quality of the restored video sequence.

**Methodology**

**Definition of requirements**

Based on the formulated goal and the conducted analytical review, it is possible to determine the requirements for the object search system being developed in the video stream. These requirements can be divided into two groups: basic (functional) and additional (structural). The main requirements can be attributed to the search technology and algorithms of the system, and additional requirements can be attributed to the structure of the system [7].

The main requirements for the object search system in the video stream are formulated based on the results of an analytical review and the needs of practical use:

1 The algorithm of the system should be based on the search for key points of the object.

2 The computational complexity of the search method should be minimally achievable for use in solving real-time problems.

3 The algorithm of the system should not contain a step of preliminary training.

4 The system should allow you to control the way the result is processed: save the result as text information, as a set of images with a selected found object, or display a set of frames with selected objects on the screen.

Based on the requirements of the video stream processing speed and ease of implementation, the structural requirements can be reduced to the following: the system structure model should be with a minimum number of static links. It follows from this that the desired system structure should be static and networked.

**Functional model**

According to the formulated requirements, the algorithm of the system should be based on the search for key points. It follows from the review of the technical literature that the methods of searching for key points have the following features that must be taken into account when using:

1 The key point may not be part of the image of only one object, it may also occur in images of other objects, since the key point is a local area in the image of a relatively small size and may have "similar" duplicates in the image due to low information content.

2 The key points may not be located on the entire area of the object image, but only in some areas, i.e. it is impossible to judge the location of the entire object image by the found key points,

therefore determining the location of the object position in the image based only on the positions of the key points is incorrect.

The presence of local features does not guarantee reliable identification due to the possibility of repetition of features on other objects and the possibility of distortion. In this regard, in order to increase the reliability of identification, it is necessary to expand the vector of local features with one of the global ones that allows identifying an object. The color histogram should be distinguished from the global features, since the color histogram is calculated quickly and does not depend on the rotation and scale of the image [8].

The model of representation of an object image in the form of a set of integral and local features of an object image is proposed:

$$F(x,y) \rightarrow \begin{cases} K^{ref} = \{K_0^{ref}, \dots, K_{m-1}^{ref}\} \\ H^{ref} = \{H_0^{ref}, \dots, H_{n-1}^{ref}\} \end{cases}$$

where $F(x,y)$ – rectangular matrix of pixel intensity values, $K^{ref}$ – vector of key points of the image of an object of dimension $m, H^{ref}$ – color histogram of an object image consisting of $n$ elements.

To improve performance, it is proposed to perform the following actions before starting the analysis of the video stream:

1 Create a set of images with projective distortions of the sample.

2 Find key points on each created image.

3 Calculate the descriptors of each key point.

The described steps for calculating key point descriptors on pre-projectively distorted sample images allow using algorithms to search for and describe key points that are unstable to projective transformations [9].

Thus, the general scheme of the search algorithm can be presented in the form of two stages [1]:

1 Quick search of candidate areas with an object image using key points.

2 Identification of an object in an image based on a global feature – a color histogram.

Based on this scheme, the formulated requirements for the system and the proposed method for calculating descriptors on projectively distorted images of the sample, it is possible to determine the set of functions that the system should have, to carry out functional design. The functional model is described based on the IDEF0 notation [10].

**Realization of the concept**

Top-level diagram

When considering the system from the point of view of the top - level function, the following provisions are highlighted (see Figure 1):

– the basic function of the system is to find an object in the video stream;

– the input streams of the system are the image of the object and the video stream;

– the mechanism by which the system functions is the program;

– the output data is a set of images of an object in a video stream.

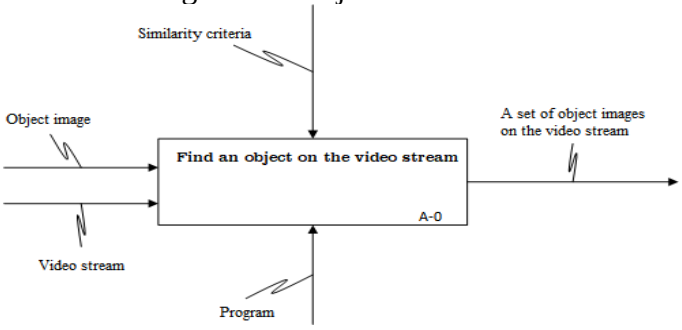



*Figure 1.* Top-level context diagram

First level diagram

As a result of the primary decomposition of the top-level diagram A-0, the diagram A0 is obtained (see Figure 2).

The decomposition of the top - level function of the system reveals the following main functions:

the function of calculating the vector of descriptors from a multidimensional matrix of projectively distorted images of an object (A1);

— the function of calculating the histogram of the object image (A2);

— the function of extracting a frame from a video stream (A3);

— function for calculating the vector of key points of a video stream frame (A4);

— the function of finding the areas of applicants of the object on the frame (A5);

— the function of identifying an object in the found areas of applicants based on the calculation of color histograms (A6).

Diagram A0 highlights the main information flows:

1 A stream containing an image of the desired object.

2 Video stream. This stream includes a set of frames on which the object is searched.

The identification function for the selected areas of the frame contains the area with the closest match of the histogram of the image of the desired object.

Second-level charts

The function of calculating the vector of descriptors of the multidimensional matrix of images of object A1 is decomposed into stages (see Figure 3), the mechanism of which is the program:

1 Converting the image to grayscale (A11). The control is carried out by the formula for converting the image into a halftone.

2 Generation of scale changes (A12) according to the control – scale parameters. At this stage, a vector of scaled images of the object is created.

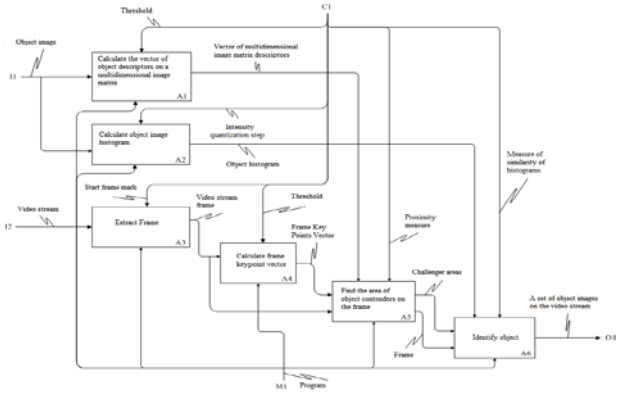



*Figure 2.* Context diagram of the first level

3 Generation of camera tilt changes (A13) according to the tilt parameters. This stage creates an image matrix from a vector of scaled images by changing the camera tilt [10].

4 Generating a change in the rotation of the object according to the rotation parameters (A14). This stage creates a three-dimensional array of images from a matrix of scaled images by changing the angle of rotation of the object image. The output of the function is a three-dimensional matrix of images of the object.

5 Finding the coordinates of the key points of the object image (A15). The key point is described by the coordinates of the image regions and the descriptor. At this step, the coordinates of the unique key points are found. The output of the function is a coordinate vector with the radius of the circles of the key points.

6 Calculation of key point descriptors (A16) on a three-dimensional matrix of object images. The key point descriptor is a vector of parameters calculated on the image in the area of

the key point. The key point is described by the coordinates of the center and the radius of the circle. The control is carried out by the coordinate vector of the center and the radius of the circles of the key points, and the threshold value. The output is a vector of object image descriptors.

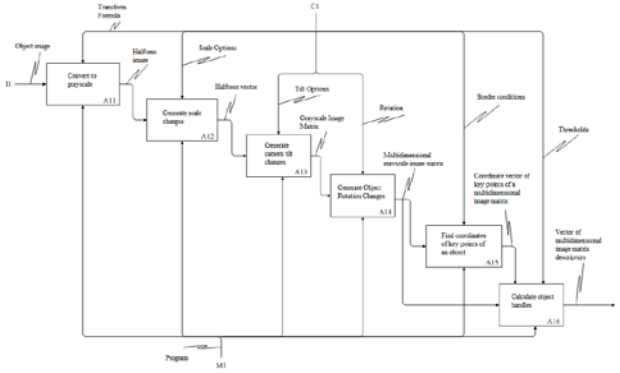



*Figure 3.* Diagram of the function for calculating the vector of object image descriptors (A1)

**Discussion of results.** Just as for the (4), ordering the values produced $C$ th feature The described steps A11, A12, A13, A14, A15 and A16 are the steps of the method for calculating descriptors on projectively distorted images of the sample.

The result of the decomposition of the frame extraction function from the A3 video stream is shown in Figure 4. During the execution of the function, the video stream goes through the following stages:

1 Demultiplication. This step is necessary to extract the encoded video signal from the container (A 31).

2 Decoding. The stage is necessary to represent the encoded video signal into a set of images that are convenient for processing in the object search function in the video stream (A32) [12].

3 Frame selection. At this stage, according to the frame parameters, a frame is extracted from the decoded video stream.

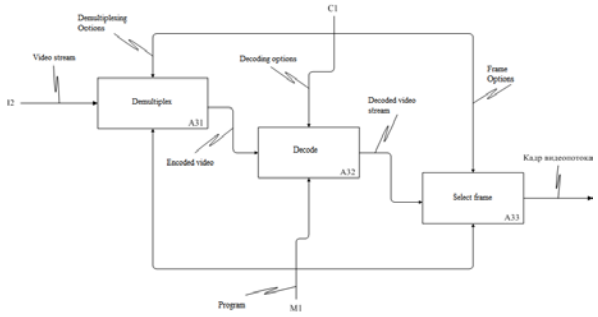



*Figure 4.* Diagram of the image extraction function from the video stream (A3)

The function of calculating the vector of key points of the A4 frame is decomposed into the following stages (see Figure 5):

1 Conversion to halftone (A41);

2 Finding the coordinates of the key points of the frame (A42);

3 Calculation of descriptors in the found frame coordinates (A43).

The coordinate vector of the key points is needed to calculate the descriptors (A43). The output of the function is a vector of key points with information about the frame number, coordinates and descriptor for each key point.

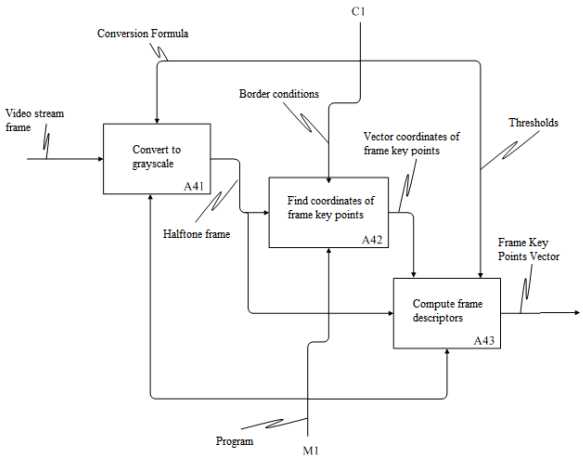



*Figure 5.* Diagram of the function for calculating the vector of key points of the frame (A4)

The function of finding the candidate areas of the object on the frame (A5) is decomposed into components (see Figure 6):

1 Finding common key points (A51). The input receives a vector of key points of the frame. The control is carried out by the vector of object descriptors. Using the proximity measure, the key points are determined, the descriptors of which are considered the same as the descriptors of the object image, i.e. the images of the areas indicated by the key points are considered "similar". If a non-empty set is obtained after the intersection, common key points and information about the frame number are output [11].

2 Frame skipping (A52). According to the information about the frame number, a decision is made to skip the input frame of the video stream for further processing.

3 Getting the applicant areas (A53). The input vector of the coordinates of the key points of the frame is combined according to the size of the window in the areas that are fed to the output of the function.

The central functions of the system are the functions of finding the key points of the object image (A15) and frame (A42), calculating the descriptors of the key points of the object image (A16) and frame (A43), finding common key points by descriptors (A51) and identifying the frame for the presence of an object image (A6). The quality and speed of object detection in the video stream depends on the result of performing these functions. Further, algorithms for finding key points together with the calculation of descriptors, algorithms for finding the intersection of descriptors and algorithms for identifying the image of an object in the frame area by calculating the histogram are considered in more detail.

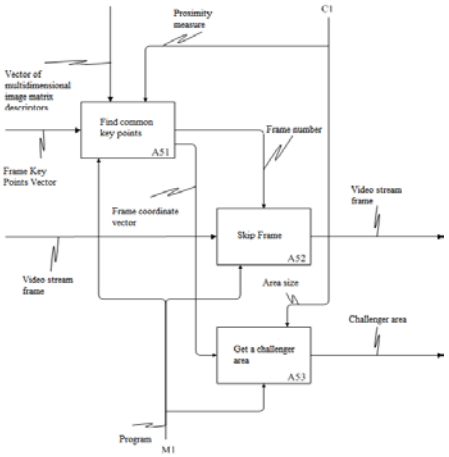



*Figure 6.* Diagram of the function of finding the areas of the object's applicants on the frame (A5)

**The practical value of the results**

1 The proposed method, as shown by the results of the study of its effectiveness, provides good quality of the restored video sequence with a higher value of the compression ratio compared to existing methods.

2 The developed fast algorithms of multidimensional Hartley transformations and PREP allowed to reduce the amount of computational costs.

3 Software and algorithmic tools have been developed that implement the proposed codec model, provide encoding and decoding of video files with the possibility of changing encoding parameters, as well as illustrating the process of codec functioning in the form of intermediate results of its operation in combination with a user-friendly interface.

Security systems based on the use of video technologies are widely used in various spheres of human activity. The development and use of production technology management systems, security

Modern video stream processing methods using big data technologies open up new possibilities for analyzing and interpreting information extracted from video images. However, the use of such technologies is associated with a number of both technical and practical challenges that must be considered for the successful implementation of effective video processing systems.

One of the main advantages of using big data is the ability to work with huge amounts of video data that can come in real time from multiple cameras. In traditional video surveillance systems, processing such volumes of information would become impossible without the use of distributed computing, such as cloud computing and the use of cluster systems. This allows you to efficiently distribute the load across multiple nodes, reducing processing time and ensuring system scalability.

In addition, the use of machine learning and deep learning techniques significantly improves the accuracy and speed of extracting meaningful information from a video stream. For example, algorithms for object recognition, motion analysis, and behavior prediction can work in real time, automatically classifying and tracking objects, which significantly improves the efficiency of video surveillance. This is especially true in areas such as security and monitoring of public places, where high accuracy and speed of data processing are required.

However, despite the obvious advantages, the use of big data technologies in video stream processing faces several challenges. One of them is the high demands on computing power and data storage. Real-time video processing requires significant resources, which can be difficult if there are limitations on computing power or a limited budget. For such systems to work effectively, it is also necessary to ensure a high level of algorithm optimization, which becomes even more difficult when it comes to dynamic and diverse video streams.

Another challenge is ensuring confidentiality and data security. When using video surveillance, it is necessary to take into account the risks of data leaks and ensure the protection of personal information. Therefore, in addition to technical solutions, it is necessary to pay attention to the legal and ethical aspects related to the use of video surveillance and big data analysis. Data storage and processing issues should be regulated accordingly in order to avoid violating user privacy.

The issue of data accuracy and quality is also important. Video streams can be subject to various interferences, such as changes in illumination, noise, object blocking, or distortion, making it difficult to accurately extract information. This poses the challenge for researchers to develop more resilient and adaptive algorithms that can work with incorrect or incomplete data.

In addition, it is important to note that the use of big data technologies in video stream processing opens up broad prospects for various industries. For example, within the framework of smart cities, such systems can be used to monitor traffic, analyze human flows, as well as to prevent offenses and optimize the operation of urban infrastructures. In the medical field, such technologies can be used to analyze video from medical devices or in the field of telemedicine to monitor the condition of patients.

In general, despite the existing challenges, the integration of big data technologies into video stream processing has enormous potential and opens up new opportunities for creating efficient and intelligent real-time data monitoring and analysis systems.

**Conclusion.** Processing objects in a video stream using big data technologies is a promising and highly efficient field that continues to evolve taking into account modern requirements for speed and accuracy of analysis. As a result of the integration of machine learning methods, distributed computing and real-time analytics, it becomes possible to efficiently extract meaningful information from huge amounts of video data, which significantly expands the possibilities of using such systems in various fields such as security, monitoring, smart cities and medicine.

However, despite the clear advantages, the technology faces a number of challenges. High demands on computing power, difficulties in ensuring data confidentiality, as well as problems with the quality of the video stream require the development of new approaches and algorithms to increase the stability and accuracy of processing. These challenges highlight the need for further research and improvement of existing technologies.

Given the rapid growth of video data volumes and the need for its operational processing, the future of processing video streams using big data looks promising. This opens up new horizons for the development of intelligent systems capable of not only efficiently analyzing data, but also predicting events, increasing security and improving the quality of life in various fields.

In conclusion, it can be argued that the use of big data technologies for video stream processing has significant potential, which requires further efforts to optimize and adapt to specific tasks and conditions.

## Reference

[1] Davenport, T. H. (2014). Big data at work: Dispelling the myths, uncovering the opportunities. Harvard Business Review Press.

[2] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 36(4), 1165-1188.

[3] Madaan, R., & Belyaeva, Z. (2016). Predictive analytics in education using big data. Procedia Computer Science, 98, 461-466.

[4] Kalyuga, S., & Nastashkin, M. (2018). Implementing big data analytics in education sector to improve student outcomes. In Proceedings of the International Scientific Conference "Digitalization of Modern Society" (pp. 241-245).

[5] Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on compliance, emotion, and learning. In Proceedings of the first ACM conference on Learning @ scale conference (pp. 189-194).

[6] Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. American educational research journal, 38(4), 915-945.

[7] Soundararajan, A., Alfayly, A., Ravindran, P., & Shah, V. (2016). Predicting academic performance of students using big data analytics. In 2016 IEEE international congress on big data (pp. 313-320). IEEE.

[8] West, E., & Tomlinson, C. A. (2016). Learning personalization in the digital age. Educational Leadership, 73(6), 34-39.

[9] Beknazarova, S., Abdullayev, S., Abdullayeva, O., Abdullayev, Z. Machine Learning Method for Predicting Human Movements/AIP Conference Proceedings 3244(1)

[10] Beknazarova, S., Abdullayeva, O., Abdullayev, S., Abdullayev, Z. Images processing: Face recognition by neural networks/E3S Web of Conferences 587.

## Авторский вклад

**Бекназарова Саида Сафибуллаевна** – исследования по внедрению Big Data and Advanced Analytics в образовательный процесс.

# ОБРАБОТКА ОБЪЕКТА В ВИДЕОПОТОКЕ С ИСПОЛЬЗОВАНИЕМ БОЛЬШИХ ДАННЫХ

***С.С. Бекназарова***
*Профессор кафедры телевизионные и медиатехнологии ТУИТ им. М.Хоразмий, доктор технических наук, профессор*

**Аннотация.** Статья посвящена изучению методов обработки объектов в видеопотоке с применением технологий больших данных. В условиях стремительного роста объемов видеоданных и необходимости их оперативного анализа, традиционные способы обработки становятся недостаточными для эффективной работы с такими массивами информации. Рассматриваются подходы, включающие использование распределенных вычислений, алгоритмов машинного обучения и аналитики в реальном времени для извлечения значимой информации из видеопотока. Особое внимание уделяется вопросам сбора, хранения и обработки данных, а также проблемам, возникающим при интеграции технологий больших данных в системы видеонаблюдения. В статье рассматриваются как технические, так и практические аспекты, включая возможности применения данных технологий в различных областях, таких как безопасность, умные города и мониторинг поведения. Результаты исследования могут быть полезны для разработчиков систем видеонаблюдения, специалистов по аналитике данных и исследователей в области искусственного интеллекта.

**Ключевые слова:** большие данные, продвинутая аналитика, видеоаналитика.

УДК 004.65:331.4-049.5

# BIG DATA В ПОВЫШЕНИИ ПРОМЫШЛЕННОЙ БЕЗОПАСНОСТИ

***Н. В. Щербина***
*Старший преподаватель кафедры*
*инженерной психологии и эргономики*
*УО «Белорусский государственный университет*
*информатики и радиоэлектроники»,*
*магистр технических наук*
shcherbina@bsuir.by

***Н. В. Щербина***

*Окончила Белорусский национальный технический университет. Область научных интересов связана с исследованием проблем саморегуляции функционального состояния человека, аспектов деятельности человека-оператора в системах контроля и управления, организацией учебного и научно-исследовательского процессов в техническом университете.*

**Аннотация**: Показаны возможности использования больших данных в промышленной безопасности для улучшения предотвращения рисков, принятия обоснованных решений и защиты работников. Описаны особенности использования цифрового средства индивидуальной защиты работника Drixit Tag разработка компании Drixit Technologies.

**Ключевые слова:** промышленная безопасность, управление промышленной безопасностью, BigData, прогнозирование опасных ситуаций, интернет вещей.

**Введение.** Большие данные открывают огромные возможности для сферы промышленной безопасности, позволяя проводить упреждающие и профилактические улучшения на основе «данных». В сфере промышленной безопасности точная информация жизненно важна для предотвращения рисков и защиты работников.

**Основная часть.** Большие данные относятся к огромному набору данных, которые постоянно генерируются различными источниками на производстве. Такими источниками могут выступать:

1 *Датчики и системы мониторинга*: устройства в промышленных условиях собирают данные о температуре, давлении, влажности, уровнях химических веществ и т.д., предоставляя подробную картину условий труда.

2 *Записи об инцидентах и несчастных случаях*: отчеты о прошлых инцидентах и несчастных случаях являются ценным источником информации для выявления закономерностей и предотвращения будущих аварий, инцидентов, несчастных случаев.

3 *Системы управления безопасностью*: платформы и программное обеспечение, используемые для управления безопасностью на промышленных объектах, регистрируют данные, связанные с политиками, процедурами, проверками и аудитами.

4 *Внешние данные*: помимо внутренних данных организаций (производств), внешние источники, такие как отраслевая статистика и специализированные базы данных, предоставляют ценную информацию для оценки рисков и улучшения мер безопасности.

Большие данные в промышленной безопасности характеризуются *объемом*, *скоростью* и *разнообразием*.

1 *Объем*: данные генерируются в больших количествах и ускоренными темпами. Управление и обработка такого объема данных требуют масштабируемых решений для хранения и аналитики, таких как облачные системы хранения и распределенные вычислительные технологии.

2 *Скорость*: данные в области промышленной безопасности генерируются и должны анализироваться в режиме реального времени или близком к реальному времени для получения соответствующей информации и принятия своевременных мер. Например, цифровые средства индивидуальной защиты (СИЗ) могут обнаруживать изменения в поведении работников и предупреждать об опасных ситуациях в режиме реального времени.

3 *Разнообразие*: данные могут быть структурированными (например, базы данных) или неструктурированными (например, текст, изображения или видео). Кроме того, они поступают из разных источников и могут иметь разные форматы, что требует гибких инструментов анализа и обработки. Обработка этого разнообразия данных требует гибких методов обработки и анализа, таких как использование алгоритмов обработки естественного языка (NLP) для извлечения информации из письменных отчетов или использование алгоритмов компьютерного зрения для анализа изображений или видео.

Так, например, компания Drixit Technologies (Аргентина), специализирующаяся на цифровизации и автоматизации существующих процессов в промышленных компаниях, предлагает использовать цифровое СИЗ (Drixit Tag) вместе с платформой Drixit. Такое решение позволяет повысить защиту промышленных работников благодаря использованию технологий и оцифровке процессов безопасности.

Устройство Drixit Tag представляет собой аппаратное обеспечение предложенного решения. Платформа Drixit – программное обеспечение, которое получает информацию от Drixit Tag в режиме реального времени и используется руководителями по безопасности в компании.

Drixit Tag – это прочное портативное устройство с компактным и эргономичным дизайном и долговечной внутренней батареей (рисунок 1), разработанное для работы в самых сложных условиях промышленной среды. Оно водонепроницаемо, пыленепроницаемо и сертифицировано для взрывоопасных сред.

Цифровое СИЗ Drixit Tag включают в себя технологии: датчики IoT, RFID-карту, LoRa, UWB, GNSS (GPS, ГЛОНАСС, Galileo) и USB-подключение. Кроме того, в пределах своего небольшого размера примерно 9×6 см, цифровой СИЗ включает в себя два акселерометра, барометр, магнитометр и гироскоп.

Отправка информации с Drixit Tag в режиме реального времени позволяет платформе Drixit:

– обнаруживать работников, находящихся в неподвижном состоянии в течение определенного периода времени (задается пользователем), например, операторы, потерявшие сознание, перенесшие головокружение, тепловой удар, сердечный приступ, сердечно-сосудистые заболевания или любые другие несчастные случаи;

– обнаруживать и сообщать о падениях работника с высоты 1, 2 метра и выше;

– использовать авторизованные сертификаты:

– обнаруживать вход, пребывание и выход работников из опасных или несанкционированных зон. Это особенно актуально для предприятий, на которых имеются зоны с высокими и/или низкими температурами, замкнутые пространства, зоны с горючими газами или любые другие условия, которые угрожают здоровью или неприкосновенности работника;

*Рисунок 1*. Устройство Drixit Tag [3]

– предотвращать столкновение. Это решение позволяет определять близость пешехода (держателя метки) и транспортного средства, оснащенного Drixit Spotter, и отправлять оповещения в режиме реального времени на оба устройства для предотвращения столкновения;

– сообщать об инциденте, аварии через кнопку тревоги и получать немедленную помощь;

– обнаруживать работников в режиме реального времени в случае чрезвычайной ситуации. При запуске модуля «Эвакуация» в чрезвычайной ситуации платформа может определить местонахождение пропавших работников и статус точек встречи, что облегчает спасательные работы.

Платформа Drixit позволяет настроить собственные протоколы, выбрав, какие события, связанные с безопасностью, будут вызывать настаиваемые оповещения, и кто их будет получить. Оповещения могут включать мгновенные уведомления, SMS-сообщения, сообщения Microsoft Teams или электронные письма. Платформа Drixit обрабатывает и представляет всю информацию об операции в реальном времени и за прошлые периоды, это позволяет знать, что произошло, почему и когда, а также заблаговременно улучшать протоколы. К основным особенностям платформы Drixit относятся следующих функции: 1) эвакуация – данная функция позволяет управлять процессами эвакуации в режиме реального времени, 2) «живая» карта – позволяет визуализировать все процессы и устройства в режиме реального времени, 3) историческая карта – позволяет визуализировать в различных форматах (тепловая карта, спагетти и карта путешествия) исторические данные любого устройства, 4) аналитика – собирает и предоставляет обработанную информацию для упреждающего улучшения процессов безопасности, 5) сертификаты – управляет разрешениями и сертификатами сотрудников, 6) смены и посещаемость – управляет сменами, посещаемостью и отработанными часами.

**Заключение.** Применение больших данных в промышленной безопасности открывает огромные возможности.

1 *Прогнозная аналитика*: используя передовые алгоритмы, можно определить закономерности и тенденции, указывающие на неизбежные рисковые ситуации, что позволяет принимать превентивные меры до возникновения аварий. Например, анализируя данные прошлых инцидентов и аварий, можно выявить общие факторы и разработать прогнозные модели для прогнозирования аналогичных ситуаций. Это позволяет реализовать соответствующие превентивные меры и снизить вероятность аварий.

2 *Мониторинг в реальном времени*: анализ собранных данных в реальном времени может предупреждать об аномалиях или отклонениях от нормальных параметров, позволяя быстро реагировать на опасные ситуации. Например, посредством мониторинга в реальном времени датчиков температуры на химическом заводе можно обнаружить внезапное повышение температуры, которое может указывать на риск взрыва.

3 *Оптимизация процессов*: анализ данных может выявить области для улучшения производственных процессов, что приведет к повышению эффективности и безопасности на рабочем месте.

4 *Обучение и повышение осведомленности*: данные можно использовать для разработки индивидуальных программ обучения и повышения осведомленности, нацеленных на конкретные опасности, выявленные с помощью анализа данных.

Таким образом, использование больших данных в промышленной безопасности позволяет извлечь выгоду из огромного массива полученных данных для улучшения предотвращения рисков, принятия обоснованных решений и защиты работников.

При правильном применении аналитических методов можно выявлять закономерности, прогнозировать опасные ситуации и оптимизировать процессы для обеспечения безопасной рабочей среды.

## Список литературы

[1] Drixit Technologies [Электронный ресурс]. – Режим доступа: https://drixit.com/products/platform. – Дата доступа: 20.12.2024.

[2] Drixit Technologies [Электронный ресурс]. – Режим доступа: https://drixit.com/meet-the-drixit-tag. – Дата доступа: 20.12.2024.

[3] Drixit Technologies [Электронный ресурс]. – Режим доступа: https://drixit.com/products/tag. – Дата доступа: 20.12.2024.

# BIG DATA IN INCREASING INDUSTRIAL SAFETY

***N.V. Shcherbina***
*Senior Lecturer, Department of*
*Engineering Psychology and Ergonomics,*
*Belarusian State University of*
*Informatics and Radioelectronics,*
*Master of Technical Sciences*

**Abstract.** The possibilities of using big data in industrial safety to improve risk prevention, informed decision-making and worker protection are shown. The features of using digital personal protective equipment for workers, Drixit Tag, developed by Drixit Technologies, are described.

**Keywords:** industrial safety, industrial safety management, BigData, forecasting of dangerous situations, Internet of things.

УДК: 517.977.5:631.672.4

# ПРИМЕНЕНИЕ БОЛЬШИХ ДАННЫХ ПРИ МОДЕЛИРОВАНИИ ВОДОХОЗЯЙСТВЕННЫХ ОБЪЕКТОВ

***A.J. Seytov***
*Associate Professor, Department of Computational Mathematics and Information Systems, DSc of Technical sciences*
saybek868@gmail.com

***O.N. Abduraxmonov***
*Senior teacher, Department of Computational Mathematics and Information Systems*
abdalim83@mail.ru

***A.J. Seytov***
*Окончил Национальный университет Узбекистана. Направление научного интереса связано с внедрением современных информационных технологий и организацией научно-исследовательских процессов в ирригации и водопользовании.*

***O.N. Abduraxmonov***
*Окончил Национальный университет Узбекистана. Направление научного интереса связано с внедрением современных информационных технологий и организацией научно-исследовательских процессов в ирригации и водопользовании.*

**Аннотация.** Особое внимание уделено использованию методов искусственного интеллекта и машинного обучения для анализа данных и оптимизации работы водохозяйственных систем. Выделяются преимущества применения больших данных, включая повышение точности моделей, улучшение управления водными ресурсами и устойчивое развитие, а также обсуждаются вызовы и ограничения, связанные с внедрением современных технологий.

Охвачены ключевые источники данных, такие как гидрометеорологические показатели, данные спутниковых наблюдений, результаты мониторинга устройств IoT и исторические записи.

**Ключевые слова:** IoT-устройств, водохозяйственных объектах, насосные станции, водных ресурсов, ирригационные системы.

**Введение.** Современное управление водохозяйственными объектами сталкивается с множеством вызовов, включая изменение климата, рост населения, увеличение потребностей в водных ресурсах и необходимость их рационального использования. Традиционные методы моделирования и управления часто оказываются недостаточно эффективными из-за ограниченности данных, сложности процессов и увеличения масштабов задач.

Технологии больших данных (Big Data) открывают новые возможности для решения этих проблем, предоставляя инструменты для сбора, анализа и использования огромных массивов информации. Они позволяют существенно повысить точность моделей, ускорить прогнозирование и разработать эффективные стратегии управления водными ресурсами.

Водохозяйственные объекты, такие как реки, водохранилища, каналы и ирригационные системы, являются ключевыми элементами инфраструктуры, от которых зависят сельское хозяйство, промышленность и обеспечение населения водой. Однако их моделирование требует учета множества факторов: гидрологических, климатических,

экологических и технических. Технологии Big Data способны объединить и обработать эти данные, обеспечивая комплексный подход к анализу.

**Преимущества применения больших данных в управлении водными объектами.** Применение больших данных в управлении водными объектами открывает широкие возможности для улучшения мониторинга, прогнозирования и управления водными ресурсами[1,3]. Вот основные преимущества использования Big Data в этой области: Повышение точности прогнозов, оптимизация использования водных ресурсов, улучшение управления водохозяйственными объектами, прогнозирование рисков и предотвращение катастроф.

Обработка больших объемов гидрологических, климатических и метеорологических данных позволяет точнее прогнозировать изменения уровня воды, частоту и интенсивность наводнений и засух. С помощью исторических данных и современных моделей машинного обучения можно прогнозировать долгосрочные изменения, что важно для устойчивого управления водными ресурсами.

На основе анализа больших данных можно более точно определять потребности различных пользователей (сельское хозяйство, промышленность, коммунальное водоснабжение), что способствует эффективному распределению водных ресурсов. Анализ информации о водопользовании и климате позволяет предсказать потребности в воде в разные сезоны и при различных сценариях изменения климата.

Использования IoT-датчиков и спутниковых данных позволяет в реальном времени отслеживать состояние водохранилищ, плотин, насосных станций и других инфраструктурных объектов, что способствует своевременному реагированию на проблемы. Применение алгоритмов для автоматического контроля и управления водными объектами (например, для открытия и закрытия клапанов) помогает снизить человеческий фактор и повысить эффективность работы.

Big Data помогает Анализ данных о скорости таяния снега, уровнях рек и осадках позволяет заранее предсказывать, позволяя заранее принять меры по водообеспечению и сокращению потерь.

Большие данные включают информацию, поступающую из различных источников:

*Гидрометеорологические данные:* данные о температуре, осадках, скорости ветра, уровне и расходе воды в реках. Гидрометеорологические данные представляют собой информацию, связанную с состоянием атмосферы, водных объектов и климатических условий. Эти данные используются для мониторинга, анализа и прогнозирования природных явлений, а также для управления водными ресурсами.

*Данные датчиков:* показания IoT-устройств, установленных на водохозяйственных объектах (плотины, насосные станции, каналы и др.). Данные датчиков представляют собой информацию, собранную с помощью специализированных устройств, которые автоматически фиксируют различные параметры окружающей среды или объектов. В контексте водных ресурсов и гидросистем такие датчики играют ключевую роль в мониторинге и управлении.

*Космические данные:* спутниковые снимки для анализа изменений водных объектов, контроля состояния водохранилищ и выявления утечек. Космические данные – это информация, полученная с помощью спутников дистанционного зондирования Земли (ДЗЗ), которая используется для наблюдения, анализа и управления природными ресурсами и экологическими процессами. В контексте водных ресурсов такие данные играют важную роль в мониторинге состояния водоёмов, оценки водных запасов и прогнозирования природных явлений.

*Исторические данные:* архивы по водопользованию, изменениям климата и гидрологических режимов. Исторические данные – это информация, собранная и зафиксированная за определённый период в прошлом, которая используется для анализа, прогнозирования и принятия решений в настоящем и будущем. В контексте водных

ресурсов и управления водохозяйственными объектами такие данные играют важнейшую роль в понимании динамики процессов и разработке эффективных стратегий.

*Эффективное распределение водных ресурсов:* данные Big Data помогают точно оценить состояние водных ресурсов в каналах и эффективно их распределять.

Предварительный расчет необходимого количества воды для ирригации. Предотвращение утечки избыточной воды через каналы.

*Использование Big Data дает много возможностей:* исключение ошибок, связанных с вводом данных вручную, снижение зависимости от субъективных решений, автоматизация контроля и управления численными значениями модели.

**Моделирование движения воды -** это сложный процесс, который требует учёта множества факторов. При этом расчёты выполняются на основе большого объёма данных, что приводит к генерации больших данных.

Канал или водный объект делится на множество сегментов (например, каждые 100–500 метров), чтобы рассчитать параметры воды (скорость течения, глубину, расход) на каждом участке. Расчёты проводятся для каждого временного шага (например, каждую минуту, час или сутки). Чем меньше шаг времени, тем более детализированное моделирование, но тем больше данных генерируется[2].

В моделировании участвуют тысячи сегментов, временных шагов и параметров. Это приводит к созданию миллионов строк данных. Для их обработки и анализа требуется использовать технологии больших данных (Big Data), такие как распределённые вычисления (Hadoop, Spark) или облачные платформы.

Для наглядного представления процесса получения больших данных при расчете скорости потока и уровня воды в канале или водоеме при моделировании движения воды приведем примеры в виде таблиц и графиков.

В этом примере рассмотрим канал, разбитый на 5 сегментов, и приведём значения скорости воды (u) и уровня воды (h).

Таблица 1. Значений скорости потока и уровня воды в разных сегментах канала

| Сегмент | Длина (м) | Скорость потока u (м/с) | Уровень воды h (м) |
|---|---|---|---|
| 1 | 100 | 1.2 | 3.0 |
| 2 | 100 | 1.5 | 3.2 |
| 3 | 100 | 1.8 | 3.1 |
| 4 | 100 | 1.3 | 2.9 |
| 5 | 100 | 1.1 | 2.8 |

Видно, что скорость воды увеличивается в среднем участке и уменьшается в конце. Уровень воды меняется в зависимости от рельефа дна и градиента потока.

**Графики скорости потока и уровня воды по сегментам.** Графики показывают, как изменяется уровень воды и скорость в зависимости от координаты вдоль канала. Можно видеть, что на определённых участках есть увеличение или уменьшение параметров, что важно для гидродинамического анализа.

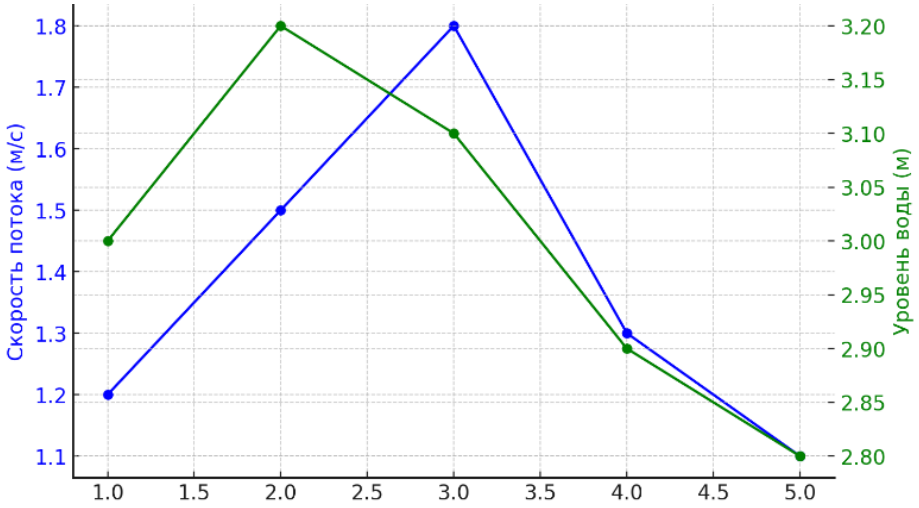



*Рисунок 1.* Изменение скорости потока и уровня воды по сегментам

Синий график (скорость потока) показывает увеличение скорости на участке 2-3 и последующее снижение. Зелёный график (уровень воды) демонстрирует небольшие колебания, снижаясь к концу канала.

Это характерно для гидродинамических процессов, когда на уклонах скорость растёт, а на пологих участках уменьшается.

**Заключение.** Использование больших данных в управлении водными объектами открывает новые горизонты для повышения эффективности водообеспечения, защиты от природных катастроф и устойчивого использования водных ресурсов. Оно также способствует улучшению экологического состояния водоемов и рациональному использованию водных ресурсов, что особенно важно в условиях изменений климата и роста населения. Применение технологий Big Data в водохозяйственном моделировании открывает большие перспективы для улучшения управления водными ресурсами и минимизации воздействия на окружающую среду.

## Список литературы

[1] D.T. Muxamedieva, O.N.Abduraxmonov, M.H.Primova. Big data и анализ данных. T.:Fidokor Yosh Avlod, 2023. -246 c.

[2] Rakhimov, S., Seytov, A., Rakhimova, N., Xonimqulov, B. Mathematical models of optimal distribution of water in main canals. 2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT), INSPEC Accession Number: 20413548, IEEE Access, Tashkent, Uzbekistan, DOI:10.1109/AICT50176.2020.9368798 (AICT) pp. 1-4,(№ 5, Scopus, IF=3,557)

[3] Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. Data Analysis / Kindle Edition, 2013.

## Авторский вклад

**Айбек Сейтов** – Руководил исследованием по применение больших данных при моделировании водохозяйственных объектов.

**Олим Абдурахмонов** – Исследователь и анализатор формирования больших данных при моделировании объектов водного хозяйства.

UDC 336.774.3

# EVALUATION OF INDICATORS CLUSTERING METHODS IN INTERRELATED COUNTERPARTIES FINANCIAL ASSESSMENT

***D.M. Rahel***
*PhD in Economics, Associate Professor of the Department of Economics of the Belarusian State University of Informatics and Radioelectronics*
ragel@bsuir.by

***D.M. Rahel***
*In 2000 he graduated from the Faculty of Economics of the Belarusian State University of Informatics and Radioelectronics with a degree in Economic Informatics. In 2016 he graduated from the postgraduate course of the Academy of Management under the President of the Republic of Belarus. In 2018 he defended his PhD thesis. Research interests: data mining in marketing, process modeling, data analysis, statistical forecasting, macroeconomics.*

**Annotation.** The article discusses approaches to clustering and classification of financial data that can be used in scoring indicators of various counterparties. The author considers two types of clustering of financial data using the Euclidean distance and the Mahalanobis distance and draws conclusions about their advantages and disadvantages.

**Keywords:** data analysis, financial data, big data, data array, scoring, analytics, data distribution, financial characteristics, financial analysis, clustering, Euclidean distance, Mahalanobis distance.

When evaluating counterparties, there is often a need to classify and search for a dependent in the data sets they provide. There are quite a large number of methods aimed at classifying and distributing large volumes of data. In this case, we will compare the reliability of the results that data clustering provides based on calculating the Euclidean distance between the data in the set under consideration and the results that can be obtained using the agglomerative clustering algorithm, which is calculated using the Mahalanobis distance formula.

**Initial data set.** The data set used contained data on 50 counterparties, which we need to classify and divide into groups depending on the values of the indicators. Thus, we have a table with 50 records, each of which contains the parameter X1 – the amount of revenue from sales and X2 – the amount of net profit. Based on these indicators, we need to classify the legal entities in question, which belong to the same market, and their business activities have the same scale according to preliminary estimates.

**Segmentation of indicators based on Euclidean distance.** The peculiarity of the algorithm is the search for the minimum distance between objects and on this basis the formation of similar groups of objects. The following characteristics were used to configure the algorithm: 2 clusters (based on empirical assessment), iterations are stopped at the moment of obtaining equivalent values of distances between groups.

*Listing 1.* Code for calculating Euclidean distance using nearest neighbor method

```
# Calculating Euclidean distance
matrix = pdist(X, metric='euclidean')
# Nearest Neighbor Method
link_matrix = sch.linkage(matrix, method='single')
# Number of clusters, in this case, 2 clusters
```

```
c = fcluster(link_matrix, t=2, criterion='maxclust')
```

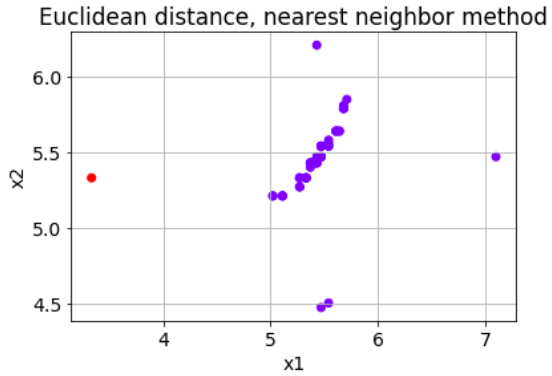



*Figure 1*. Results of data clustering based on Euclidean distance using the nearest neighbor method

Figure 1 shows the clusters formed based on the algorithm under consideration; as a result of the implementation, no clear cluster groups were formed and it is impossible to draw conclusions about the essential features and interrelationships of the counterparties considered in the analysis.

When implementing the algorithm for calculating the Euclidean distance based on the search for a distant neighbor, we see a similar picture of the final clustering of counterparties (Figure 2).

*Listing 2*. Program code for calculating the Euclidean distance using the distant neighbor method

```
# Calculating Euclidean distance
matrix = pdist(X, metric='euclidean')
# Distant Neighbor Method
link_matrix = sch.linkage(matrix, method='complete')
# Number of clusters, in this case, 2 clusters
c = fcluster(link_matrix, t=2, criterion='maxclust')
```

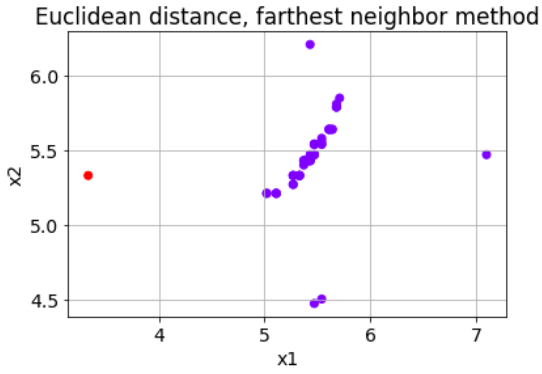



*Figure 2*. Results of data clustering based on Euclidean distance using the farthest neighbor method

**Data segmentation based on Mahalanobis distance.** When using the agglomerative clustering algorithm, which assumes the correlation between the data under consideration and, due to this, is invariant to the scale of the data volume under consideration, we obtained a picture that is more definite (Figure 3).

*Listing 3*. Program code for calculating the Mahalanobis distance

```
# Calculate the covariance matrix
cov_matrix = np.cov(X.T)
# Calculate the inverse covariance matrix
icov_matrix = np.linalg.inv(cov_matrix)
# Function to calculate Mahalanobis distance
```

```
def mah_dist(x, y, icov_matrix):
    d = x - y
    return np.sqrt(np.dot(np.dot(d, icov_matrix), diff.T))
# Calculate the Mahalanobis distance matrix between all pairs of points
dist_matrix = cdist(X, X, metric=lambda u, v: mah_dist(u, v, icov_matrix))
# Hierarchical Agglomerative Clustering Using Nearest Neighbor Method
c = linkage(dist_matrix, method='single')
```

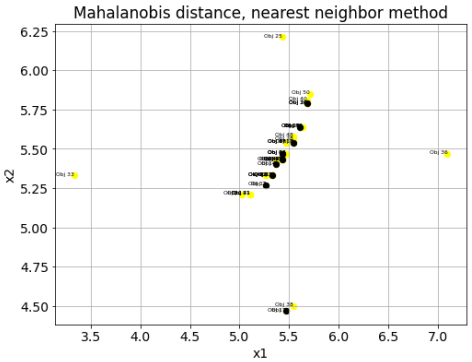



*Figure 3.* Results of data clustering based on Mahalanobis distance using nearest neighbor method

In this case, we see a clearer division of groups, similar companies were in the same cluster, which allows us to make initial conclusions about their connectivity and specify further analytical procedures. From this point of view, this type of algorithm is more effective in implementing scoring procedures.

**Conclusion.** During the study of the features of assessing the connectivity of counterparties, we considered two types of algorithms - based on the calculation of the Euclidean distance and agglomerative algorithms with the calculation of the Mahalanobis distance. The second type of algorithms, which takes into account the correlation estimates of indicators, showed its higher efficiency in implementing scoring procedures for assessing counterparties.

## References

[1] Cainiao Smart Logistics [Электронный ресурс]. Режим доступа: https://www.cainiao.com/en/index.html. – Дата доступа: 12.01.2025

[2] Data Science and Big Data Analytics. A Step by Step Guide to learn Data Science from Scratch with Python Machine Learning and Big Data / Andrew Park. – Published by Andrew Park, 2021. – 124 p.

[3] Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. Applied Logistic Regression // John Wiley & Sons. - 2013. - Vol. 398, No. 56. – P. 38–60.

[4] Nicholson W.L. Exploring Data Analysis. – Nobel Press, 2012. – 421 p.

# ОЦЕНКА МЕТОДОВ КЛАСТЕРИЗАЦИИ ПОКАЗАТЕЛЕЙ ДЛЯ ФИНАНСОВОЙ ОЦЕНКИ ВЗАИМОСВЯЗАННЫХ КОНТРАГЕНТОВ

***Д.М. Рагель***
*к.э.н., доцент кафедры экономики Белорусского государственного университета информатики и радиоэлектроники*

**Аннотация.** В статье изложен подход к анализу характера распределения различных типов финансовых данных. Предлагается подход, при котором можно проанализировать совокупность однотипных данных на основании построения ряда распределения и анализа значений в различных интервалах ряда. Характер распределения значений позволит сделать выводы о нормальности деятельности изучаемых контрагентов и, с учетом этого, провести дальнейший анализ данных.

**Ключевые слова:** анализ данных, финансовые данные, большие данные, массив данных, прогнозирование, аналитика, распределение данных, финансовые характеристики, финансовый анализ, кластеризация, Евклидово расстояние, расстояние Махаланобиса.

УДК 378.016: 339.138

# ОПЫТ ПРИМЕНЕНИЯ В ОБРАЗОВАНИИ ЧАТ-БОТОВ, ОСНОВАННЫХ НА БОЛЬШОЙ ЯЗЫКОВОЙ МОДЕЛИ

***Н.В. Павлов***
*Доцент института промышленного менеджмента, экономики и торговли СПбПУ, Dr. Sc., доцент*
pavlov_nv@spbstu.ru

***Н.В. Павлов***

*Окончил Ленинградский электротехнический институт (1980). Кандидат технических наук (1991). Доцент СПбПУ (с 1994), Doctor of Science (2016, Åbo Academi University, Finland). Исследования и преподавание маркетинговых исследований, информационных систем, расширенного интеллекта.*

**Аннотация.** Рассмотрена проблема проникновения чат-ботов, основанных на большой языковой модели, в высшее образование в области маркетинга. Показана необходимость обучения студентов работе с этим инструментом в условиях его быстрого развития и совершенствования.

Предложен комплекс практических заданий в разделе учебной дисциплины «Информационная поддержка маркетинга», посвященный этой тематике. С их помощью студенты осваивают основные приемы работы с чат-ботами, раскрывают преимущества и недостатки данного инструмента.

**Ключевые слова:** чат-боты, информационная поддержка, обучение, маркетинг, практические задания.

**Введение.** Бум применения студентами нейросетевых чат-ботов, основанных на большой языковой модели, начался с публикации о том, что дипломная работа была написана с помощью чат-бота за 23 часа и успешно защищена [1, 2]. В итоге студент получил документ об образовании, но возникла проблема отношения к подобной практике [3, 4].

Этот прецедент вдохновил студентов. Среди них появилось множество последователей. В основном такие попытки не были успешными. Появились рекомендации по написанию выпускных работ с помощью нейросетей. Источник [5] предлагает для этого несколько нейросетей, описывает сложности процесса, включая их дообучение. В [6] указывается на необходимость осторожности в использовании чат-ботов при работе. Становится понятным, что необходимость сформировать детальное задание и дообучить чат-бот требует специальных компетенций и не слишком упрощает процесс написания выпускной работы.

Началось и противодействие практике генерации выпускных работ с помощью чат-ботов. Вводились нормы процента сгенерированного текста, системы антиплагиата стали распознавать такие тексты. Появились рекомендации, как сделать сгенерированный текст

нераспознаваемым [7]. Исследование НИУ ВШЭ 2024 года [8] посвящено вопросам борьбы со сгенерированными текстами.

В СПбПУ нет запрета на использование сгенерированных текстов, хотя наличие такого текста и отмечается в справке антиплагиата. Оценка результата – в ведении руководителя и комиссии. В итоге речь идет о том, чтобы студент сам отвечал за смысл написанного.

Ряд источников предлагают авторам выделять в тексте сгенерированные фрагменты и указывать, что они сгенерированы.

Однако очевидно, что чат-боты могут принести пользу в обучении студентов [9, 10]. Поэтому возникает задача правильно обучить студентов работе с ними. Тем более, что этот инструмент имеет большие перспективы также и в процессе трудовой деятельности.

Данная работа посвящена анализу опыта обучения студентов СПбПУ использованию чат-ботов в учебном процессе при подготовке будущих маркетологов.

**Структура учебного раздела.** Раздел, посвященный чат-ботам, входит в курс «Информационная поддержка маркетинга» на его начальном этапе. Этот курс успешно опробован в институте промышленного менеджмента, экономики и торговли.

Основная идея раздела – подготовить студентов к тому, что чат-боты быстро изменяются и совершенствуются. Важно научиться анализировать их функционал и находить возможности его применения.

Задание 1 состояло в том, чтобы найти различные чат-боты и поработать с ними, выполняя различные практические задачи, после чего оценить возможности этих ботов. Упор делался на русскоязычные чат-боты. В качестве задач предлагались вспомогательные действия для процесса обучения.

1 Вывести определения важных понятий, например, маркетинга, информации, информационной системы. После этого нужно было оценить качество сгенерированных определений исходя из знаний, полученных в процессе обучения. Различные чат-боты выдавали различные определения. При этом определения понятий по одной тематике не давали целостной картины.

2 Выдать ссылки на найденные определения. Требовалась проверка их правильности: реальности источника и определения в нем. Довольно часто встречались ссылки на несуществующие источники, указывались Интернет и Википедия.

Все это подводило к выводу о том, что чат-бот пригоден лишь как начальный шаг в изучении предмета, получения данных для дальнейшей проверки.

3 Собрать литературу по определенной теме. Здесь были варианты ответов: отказ, ссылка на «просторы Интернет», название и автор книги, но без точной библиографической ссылки. В лучшем случае полезный результат состоял в том, какую книгу искать в поисковой системе.

4 Привести оглавление книги по специальности, краткое содержание книги или научной статьи. Часто выдавались несуществующее содержание и оглавление.

5 Спровоцировать чат-бота на выдачу заведомо неверных результатов: что Земля плоская, что собаки пишут книги (чат-бот представил несколько таких книг) и т.п.

В итоге стало понятно, что доверять чат-боту полностью нельзя.

Задание 2: проанализировать книгу [11], составленную ИИ, и находящуюся в свободном доступе. Предлагалось найти в тексте нарушения следующего типа:

- неконкретность;
- отсутствие смысла;
- излишняя многословность;
- отсутствие логики.

Возможно, результат этого задания был несколько предвзятым: требовалось найти именно недостатки. Но цель задания как раз и состояла в том, чтобы разрушить розовые мечты о том, что чат-бот сам создаст хорошую выпускную работу.

Итог этой части курса – получить представление об ограниченности возможностей чат-ботов в роли заменителя студента как автора.

Задание 3. Далее следовала часть обучения использованию чат-бота в работе маркетолога.

Предлагалось разработать рекламное обращение, содержащее текст с описанием товара и призывом его купить, и изображение.

В этой работе важная роль принадлежит правильному составлению промпта – задания для чат-бота. Промпты вводились в основном в печатном виде, хотя некоторые студенты, работающие на смартфонах и персональных компьютерах, вводили задания голосом. Результаты были часто довольно оригинальными, а если в промпте указывался юмористический стиль, то и забавными.

Полученные несколько версий обращения анализировались и, иногда, тексты улучшались вручную, а изображения генерировались заново, с уточнениями в промпте.

Это задание предпочтительно проводить в бригадах из двух человек, так как при этом лучше решаются подобные креативные задачи.

В результате был отработан навык использования ИИ как помощника и генератора идей для дальнейшего анализа.

Версии рекламных обращений требовалось дать для оценки потенциальным потребителям, в данном случае – студентам этой же группы. От чат-бота требовалось составить для этого подходящую анкету. Сгенерированная анкета анализировалась и при необходимости корректировалась вручную (студенты уже прошли курс маркетинговых исследований, предполагалось, что они знали, как разрабатывать анкеты).

С помощью чат-ботов процесс составления анкеты значительно ускорился.

Результаты анкетирования также вводились в чат-бот, чтобы он исправил рекламное обращение. Но это часто не давало заметных улучшений.

Оказалось, что лучше пытаться указать на желаемое улучшение промптами.

В результате был освоен процесс работы с поддержкой ИИ, прочувствованы его сильные и слабые стороны.

Для полноты картины, сгенерированные рекламные обращения были использованы при выполнении задания уже другого раздела по разработке электронного магазина в системе Битрикс24.

Впоследствии чат-бот также применялся

– в задании по построению наивного байесовского классификатора для генерации коротких сообщений от друзей и спама;

– для анализа настроений отзывов покупателей (это сложная проблема, результат получается далеко не идеальный);

– генерации исходных вариантов ответов на реальные отзывы покупателей.

**Комментарии**. Введение заданий, связанных с чат-ботами, в дисциплину «Информационная поддержка маркетинга» было нацелено на знакомство с этим инструментом, а затем – на освоение приемов использования его в работе маркетологов, где для него имеется много применений. Это способствует появлению навыков совместной работы человека и чат-бота.

**Заключение.** Таким образом, ход выполнения заданий по разделу чат-ботов показал ограниченность их возможностей и научил использовать их помощь в обучении и работе. Но главный вывод состоит в том, что все результаты работы нейросетей надо проверять вручную.

Можно сказать, что практический опыт подтверждает рекомендации, высказанные в [12], главные из которых:

– необходимость развития ключевых навыков будущего;

– общение с искусственным интеллектом как с армией обезьянок: их изучение; выдача правильных заданий и недопустимость ставить их на руководство.

В качестве дальнейшего развития темы предполагается разработать задание анализа потока отзывов и их классификации на положительные и отрицательные, помощь в решении вычислительных задач в Excel. Это значительно облегчит работу преподавателя.

## Список литературы

[1] Абрамов А. Нейросеть за один вечер написала диплом за российского студента. Преподаватели в шоке – как теперь проверять знания?. – URL https://www.msk.kp.ru/daily/27460/4714947/.

[2] Жадан А. Как я написал диплом с помощью ChatGPT и оказался в центре спора о нейросетях в образовании. – URL https://journal.tinkoff.ru/neuro-diploma/.

[3] Воробьев Д. «Найдут лазейку»: Аналитик о запрете чат-бота по написанию дипломов. – URLhttps://nsn.fm/nauka-i-tehnologii/naidut-lazeiku-analitik-ne-verit-v-zapret-chat-bota-po-napisaniu-diplomov.

[4] Кеффер Л. Глава Минобрнауки Фальков призвал не наказывать выпускника, написавшего диплом с помощью ChatGPT. – URL https://www.kommersant.ru/doc/5812075.

[5] Нейросети для написания дипломной работы. // Giga chat. – URL https://giga.chat/help/articles/ai-for-diploma.

[6] Гресева Инна. Как написать диплом или курсовую с помощью нейросети. – URL https://softline.ru/about/blog/kak-napisat-diplom-ili-kursovuyu-s-pomoshyu-nejroseti.

[7] Мустакримов В. GPT студентам. Издательские решения. – 2023. – 325 с.

[8] Опрос о применении искусственного интеллекта учащимися вузов. // ICT Moscow. – URL https://ict.moscow/projects/ai/research/opros-o-primenenii-iskusstvennogo-intellekta-uchashchimisia-vuzov/.

[9] Nartey E.K. Generative AI in Higher Education: Guiding Principles for Teaching and Learning. – CRC Press. – 2025. – 215 p.

[10] Khan S. Brave New Words : How Ai Will Revolutionize Education (And Why That`s A Good Thing). – Viking. – 2024. - 272 p.

[11] 4 Chat GPT. Умнее с ИИ: Как использовать ChatGPT для достижения успеха и создания богатства. – 2023. – 24 с. – URL https://moreknig.org/drugoe/nauchnaya-literatura-prochee/369684-umnee-s-ii-kak-ispolzovat-chatgpt-dlya-dostizheniya-uspeha-i-sozdaniya-bogatstva.html.

[12] Руз К. Устойчивы к будущему. 9 правил для людей в эпоху машин. – Маннб Иванов и Фарбер. – 2021. – 239 с.

# EXPERIENCE OF USING CHATBOTS BASED ON A LARGE LANGUAGE MODEL IN EDUCATION

***N.V. Pavlov***
*Associate Professor of the Institute of Industrial Management, Economics and Trade, SpbPU, Dr Sc (Technology), Associate Professor*

**Abstract.** The problem of penetration of large language model based chatbots into higher education in the field of marketing is considered. The necessity of training students to work with this tool under conditions of rapid development and improvement of chatbots is shown. A number of practical tasks are offered in the respective section of the academic discipline "Information support of marketing". With their help, students master the basic techniques of working with chatbots, reveal the advantages and disadvantages of this tool.

**Keywords:** chatbots, information support, training, marketing, practical tasks.

УДК 519.624.2

# МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ПРОПУСКА ОШИБКИ ПРИ НАБЛЮДЕНИИ ВЕКТОРОВ ПЕРЕХОДОВ

***И.П. Кобяк***
*Белорусский государственный университет*
*Информатики и радиоэлектроники,*
*Республика Беларусь,*
*доцент кафедры ЭВМ,*
*канд. техн. наук, доцент*
IPKobyak2012@mail.ru

**И.П. Кобяк**

*Работает в Белорусском государственном университете с 1982 г. Занимаемые должности инженер, ассистент, доцент кафедры ЭВМ. Защитил кандидатскую диссертацию в 1993 г. Область научных интересов: методы идентификации сообщений, проектирование спецпроцессоров.*

**Аннотация.** В работе исследован метод идентификации последовательностей случайных событий с помощью оценок наблюдения векторов переходов заданного вида. Определено математическое ожидание плотности распределения вероятностей ошибки, чего требует задача сравнения статистических показателей предлагаемого и других алгоритмов свертки. Основой для расчетов послужила перечисляющая производящая функция, позволяющая представить произведение комбинаторных моментов, характеризующих вероятность пропуска ошибки, в виде суммы произведений статистик на соответствующие моменты. Полученные в работе соотношения характеризуют метод наблюдения векторов переходов как наиболее эффективный алгоритм синтеза оценок для идентификации сообщений со случайной или псевдослучайной природой.

**Ключевые слова:** вектор перехода, сложные события, монообъекты и постобъекты, идентификация последовательностей, вероятность пропуска ошибки, сигнатурный анализ, вектор состояния.

**Введение.** Современные методы передачи данных при реализации сетевых технологий, как правило, базируются на применении шумоподобного кодирования для сокрытия детерминизма в передаваемых сообщениях. При этом основное время работы системы отводится для циркуляции шума по каналам связи и только в определенное время (рассчитанное компьютерной системой) в сеть передается шифротекст пользователя. Такой принцип организации обмена данными обладает рядом преимуществ перед системами, основанными на методах вычисления шифровеличин над конечными полями. При этом обнаружение каналов с криптосвязью затруднено, если само шифрованное сообщение обладает близкими к эталонным вероятностными характеристиками. Кроме того, если передаваемые данные распределяются по «плавающей» базе неравномерно, а время передачи не определено, то пересылаемый текст, что особенно важно, приобретает дополнительную алгоритмическую защиту.

Решение задач в области идентификации случайных процессов с помощью точечных оценок конечной выборки чаще всего основываются на анализе свертки по модулю два или регистрации и исследовании числа заданных элементарных событий. Однако у всех

известных методов свертки сумма вероятностей под интегральной кривой равна единице, что говорит, в среднем, об их эквивалентности.

Одним из современных алгоритмов обнаружения детерминизма в $r$ - разрядных случайных процессах следует считать принцип наблюдения сложных событий на базе двух и более векторов состояний (ВС) [1,2]. Данный подход приводит к уменьшению перестановок сложных событий на $n$ местах размещения и, соответственно, к уменьшению числа коллизий оценки. При этом вероятность пропуска ошибки оказывается существенно меньшей, чем при наблюдении элементарных векторов. В целом, вопросы, связанные с анализом процессов на основе методов наблюдения выборочных вероятностей сложных объектов таких, например, как векторы переходов заданного вида (ВПЗВ), рассмотрены в печати недостаточно, хотя производящие функции (ПФ) и соответствующих вероятностные моменты опубликованы в работах [3,4].

Таким образом, с учетом концепции использования для систем связи только шумоподобных методов выполним расчет эталонного значения математического ожидания при наблюдении ВПЗВ, образованных двумя ВС со сдвигом друг относительно друга на интервал времени $\tau = 1.$

**Математическое ожидание распределения вероятностей пропуска ошибки при наблюдении ВПЗВ с параметром** $j = 1$**.** Анализ подпоследовательностей различной длины в выборке событий размерностью $n$ позволяет ряд свойств, полученных в [5,6], использовать при выводе требуемых соотношений в представляемой работе. Так в указанных статьях показано, что соотношением для энумератора или перечисляющей производящей функции для частных значений $n$ и $j = 1$, с учетом кусочно-линейного представления ПФ, является равенство:

$$P_{ifc} = Mo + \sum_{g} \pi(g) \left( k_{1,1} p e^t x_1^1 + \sum_{i=2}^{n-4} k_{1,i} p e^t x_1^i \frac{1}{2^{i+1}} \beta_{1,i} + k_{1,n-2} p e^t x_1^{n-2} \frac{1}{2^{n-1}} \beta_{1,n-2} \right)^{n-g}, \qquad (1)$$

где статистики $k_{1,i}$ - это случайные значения, характеризующие число монообъектов длиной $2+i$ векторов состояний, $x_{j=1}^{i}$ - модификатор ПФ, $p$ - вероятность наблюдения ВПЗВ, в которой все объекты с $\mu$ битами перехода учитываются соотношением:

$$p = \frac{3^{r-\mu}}{m^2}, \qquad m = 2^r.$$

Чтобы получить равенство для математического ожидания параметра (1), необходимо выполнить дифференцирование данного соотношения по времени. При этом получаем:

$$P'_{ifc} = \sum_{g} \pi(g)(n-g) \left( k_{1,1} p e^t x_1^1 + \sum_{i=2}^{n-4} k_{1,i} p e^t x_1^i \frac{1}{2^{i+1}} \beta_{1,i} + k_{1,n-2} p e^t x_1^{n-2} \frac{1}{2^{n-1}} \beta_{1,n-2} \right)^{n-g-1} \times$$
$$\times \left[ k_{1,1} p \left( \frac{\partial}{\partial t} e^t \right) x_1^1 + \sum_{i=2}^{n-4} k_{1,i} p \left( \frac{\partial}{\partial t} e^t \right) x_1^i \frac{1}{2^{i+1}} \beta_{1,i} + k_{1,n-2} p \left( \frac{\partial}{\partial t} e^t \right) x_1^{n-2} \frac{1}{2^{n-1}} \beta_{1,n-2} \right].$$

При $t = 0,\ x_1^i = 1$ данное соотношение достаточно просто приводится к виду:

$$P'_{ifc} = \sum_{g} \pi(g)(n-g) \left( k_{1,1} p + \sum_{i=2}^{n-4} k_{1,i} p \frac{1}{2^{i+1}} \beta_{1,i} + k_{1,n-2} p \frac{1}{2^{n-1}} \beta_{1,n-2} \right)^{n-g-1} \times \qquad (2)$$
$$\times \left( k_{1,1} p + \sum_{i=2}^{n-4} k_{1,i} p \frac{1}{2^{i+1}} \beta_{1,i} + k_{1,n-2} p \frac{1}{2^{n-1}} \beta_{1,n-2} \right).$$

Для формирования суммы моментов в (2) будем использовать концепцию единственности представления вероятностных объектов, то есть со статистиками $k_{1,i} = 1$, а,

как следствие, и при $\pi(g)=1$, определяя дополнительную статистику $n_1 >> 1$, $n \geq n_1$ с целью упрощения в дальнейшем формул суммирования дробей. Тогда из соотношения (2) имеем:

$$P'_{ifc} = n\left( p + \sum_{i=2}^{n_1-4} p\frac{1}{2^{i+1}}\beta_{1,i} + p\frac{1}{2^{n_1-1}}\beta_{1,n_1-2} \right)^{n-1} \left( p + \sum_{i=2}^{n_1-4} p\frac{1}{2^{i+1}}\beta_{1,i} + p\frac{1}{2^{n_1-1}}\beta_{1,n_1-2} \right).$$

Для максимального значения вероятности наблюдения ВПЗВ, что соответствует значению параметра $p=\frac{3}{16}$ при $r=2$, с учетом результатов работы [7] можем записать:

$$P'_{ifc} \approx n\left( p\frac{33}{8} \right)^n.$$

Таким образом, для минимальной суммы моментов энумератора (1), что соответствует максимуму вероятности $p$, математическое ожидание плотности распределения вероятностей ошибки при регистрации ВПЗВ с $j=1$ и $i=\overline{\text{var}}$ имеет численное значение:

$$P'_{ifc} \approx n(0{,}7734)^n.$$

Итак, полученное значение говорит о том, что с увеличением длины выборки математическое ожидание в данной задаче стремится к бесконечно малой величине.

**Математическое ожидание распределения вероятностей пропуска ошибки при наблюдении ВПЗВ с параметрами** $j=1,2$. Итак, усложним вид функции (1) полагая, что параметр $j$ может принимать два значения $j=1$ и $j=2$. Тогда:

$$P_{ifc} = Mo + \sum_g \pi(g)\left[ \frac{1}{\xi+1}\left( k_{1,1}pe^t x_1^1 + \sum_{i=2}^{n-4} k_{1,i}pe^t x_1^i \frac{1}{2^{i+1}}\beta_{1,i} + k_{1,n-2}pe^t x_1^{n-2}\frac{1}{2^{n-1}}\beta_{1,n-2} \right) + \right.$$
$$\left. + \frac{\xi}{\xi+1}\left( k_{2,1}p^2e^{2t}x_2^1 + \sum_{i=2}^{n-6} k_{2,i}p^2e^{2t}x_2^i\frac{1}{2^{i+1}}\beta_{2,i} + k_{2,n-4}p^2e^{2t}x_2^{n-4}\frac{1}{2^{n-3}}\beta_{2,n-4} \right) \right]^{n-g},$$

где значение $\xi$ это параметр квадратичной формы вероятности $p$ [8].

Первая производная ПФ указанного вида определяется аналогично случаю с $j=1$. При этом имеем:

$$P'_{ifc} = \sum_g \pi(g)(n-g)\left[ \frac{1}{\xi+1}\left( k_{1,1}pe^t x_1^1 + \sum_{i=2}^{n-4} k_{1,i}pe^t x_1^i \frac{1}{2^{i+1}}\beta_{1,i} + k_{1,n-2}pe^t x_1^{n-2}\frac{1}{2^{n-1}}\beta_{1,n-2} \right) + \right.$$
$$\left. + \frac{\xi}{\xi+1}\left( k_{2,1}p^2e^{2t}x_2^1 + \sum_{i=2}^{n-6} k_{2,i}p^2e^{2t}x_2^i\frac{1}{2^{i+1}}\beta_{2,i} + k_{2,n-4}p^2e^{2t}x_2^{n-4}\frac{1}{2^{n-3}}\beta_{2,n-4} \right) \right]^{n-g-1} \times$$
$$\times\left[ \frac{1}{\xi+1}\left( k_{1,1}p\left(\frac{\partial}{\partial t}e^t\right)x_1^1 + \sum_{i=2}^{n-4} k_{1,i}p\left(\frac{\partial}{\partial t}e^t\right)x_1^i\frac{1}{2^{i+1}}\beta_{1,i} + k_{1,n-2}p\left(\frac{\partial}{\partial t}e^t\right)x_1^{n-2}\frac{1}{2^{n-1}}\beta_{1,n-2} \right) + \right.$$
$$\left. + \frac{\xi}{\xi+1}\left( k_{2,1}p^2\left(\frac{\partial}{\partial t}e^{2t}\right)x_2^1 + \sum_{i=2}^{n-6} k_{2,i}p^2\left(\frac{\partial}{\partial t}e^{2t}\right)x_2^i\frac{1}{2^{i+1}}\beta_{2,i} + k_{2,n-4}p^2\left(\frac{\partial}{\partial t}e^{2t}\right)x_2^{n-4}\frac{1}{2^{n-3}}\beta_{2,n-4} \right) \right].$$

В соответствии с преобразованием (2) при $t=0,\ x_1^i = x_2^i = 1$ сформированная ПФ может быть приведена к виду:

$$P'_{ifc} = \sum_g \pi(g)(n-g)\left[ \frac{1}{\xi+1}\left( k_{1,1}p + \sum_{i=2}^{n-4} k_{1,i}p\frac{1}{2^{i+1}}\beta_{1,i} + k_{1,n-2}p\frac{1}{2^{n-1}}\beta_{1,n-2} \right) + \right.$$

$$+\frac{\xi}{\xi+1}\left(k_{2,1}p^2+\sum_{i=2}^{n-6}k_{2,2}p^2\frac{1}{2^{i+1}}\beta_{2,i}+k_{2,n-4}p^2\frac{1}{2^{n-3}}\beta_{2,n-4}\right)\Bigg]^{n-g-1}\times \tag{3}$$

$$\times\left[\frac{1}{\xi+1}\left(k_{1,1}p+\sum_{i=2}^{n-4}k_{1,i}p\frac{1}{2^{i+1}}\beta_{1,i}+k_{1,n-2}p\frac{1}{2^{n-1}}\beta_{1,n-2}\right)+\right.$$

$$\left.+\frac{\xi}{\xi+1}\left(2k_{2,1}p^2+2\sum_{i=2}^{n-6}k_{2,i}p^2\frac{1}{2^{i+1}}\beta_{2,i}+2k_{2,n-4}p^2\frac{1}{2^{n-3}}\beta_{2,n-4}\right)\right].$$

Для формирования суммы моментов в соответствии с функцией (3) будем выполнять (аналогично предыдущему варианту) преобразования при $k_{1,i}=1$ и $k_{2,i}=1$. Кроме того для данной функции введем обозначения:

$$\frac{1}{\xi+1}=\xi_1, \qquad \frac{\xi}{\xi+1}=\xi_2.$$

Тогда, полагая $\pi(g)=1$ и определяя статистику $n_1>>1$, при $n\ge n_1$ получаем:

$$P'_{ifc}=n\left[\xi_1\left(p+\sum_{i=2}^{n_1-4}p\frac{1}{2^{i+1}}\beta_{1,i}+p\frac{1}{2^{n_1-1}}\beta_{1,n_1-2}\right)+\xi_2\left(p^2+\sum_{i=2}^{n_1-6}p^2\frac{1}{2^{i+1}}\beta_{2,i}+p^2\frac{1}{2^{n_1-3}}\beta_{2,n_1-4}\right)\right]^{n-1}\times$$

$$\times\left[\xi_1\left(p+\sum_{i=2}^{n_1-4}p\frac{1}{2^{i+1}}\beta_{1,i}+p\frac{1}{2^{n_1-1}}\beta_{1,n_1-2}\right)+\xi_2\left(2p^2+2\sum_{i=2}^{n_1-6}p^2\frac{1}{2^{i+1}}\beta_{2,i}+2p^2\frac{1}{2^{n_1-3}}\beta_{2,n_1-4}\right)\right].$$

Для максимального значения вероятности наблюдения ВПЗВ опять же с учетом результатов работ [7,8] при $\xi_1=1-p,\ \xi_2=p$ можем записать:

$$P'_{ifc}\approx n\left[(1-p)p\frac{33}{8}+p^3\frac{33}{8}\right]^{n-1}\left[(1-p)p\frac{33}{8}+2p^3\frac{33}{8}\right]=0{,}683n(0{,}656)^{n-1}.$$

Таким образом, полученное значение практически не отличается от результата с параметром $j=1$, так как монообъекты с $j=2$ встречаются крайне редко.

**Общий случай математического ожидания распределения вероятностей пропуска ошибки при наблюдении ВПЗВ для $\forall j$.**

Получим общее соотношение для энумератора (1) при всех $j=1\div 0{,}5(n-1)$. При этом имеем:

$$P_{ifc}=Mo+\sum_g\pi(g)\left[\sum_{j=1}^{0,5(n-5)}\xi_jk_{j,1}p^je^{jt}x_j^1+\sum_{j=1}^{0,5(n-5)}\xi_j\sum_{i=2}^{n-2j-2}k_{j,i}p^je^{jt}x_j^i\frac{1}{2^{i+1}}\beta_{j,i}+\right. \tag{4}$$

$$\left.+\sum_{j=1}^{0,5(n-3)}\xi_jk_{j,n-2j}p^je^{jt}x_j^{n-2j}\frac{1}{2^{n-2j+1}}\beta_{j,n-2j}+\xi_{0,5(n-1)}k_{\frac{n-1}{2},1}p^{\frac{n-1}{2}}e^{\frac{n-1}{2}t}x_{\frac{n-1}{2}}^1\right]^{n-g}.$$

Дифференцирование равенства (4) по времени приводит к полиному вида:

$$P'_{ifc}=\sum_g\pi(g)(n-g)\left[\sum_{j=1}^{0,5(n-5)}\xi_jk_{j,1}p^je^{jt}x_j^1+\sum_{j=1}^{0,5(n-5)}\xi_j\sum_{i=2}^{n-2j-2}k_{j,i}p^je^{jt}x_j^i\frac{1}{2^{i+1}}\beta_{j,i}+\right.$$

$$\left.+\sum_{j=1}^{0,5(n-3)}\xi_jk_{j,n-2j}p^je^{jt}x_j^{n-2j}\frac{1}{2^{n-2j+1}}\beta_{j,n-2j}+\xi_{0,5(n-1)}k_{\frac{n-1}{2},1}p^{\frac{n-1}{2}}e^{\frac{n-1}{2}t}x_{\frac{n-1}{2}}^1\right]^{n-g-1}\times \tag{5}$$

$$\times\left[\sum_{j=1}^{0,5(n-5)}\xi_jk_{j,1}p^j\left(\frac{\partial}{\partial t}e^{jt}\right)x_j^1+\sum_{j=1}^{0,5(n-5)}\xi_j\sum_{i=2}^{n-2j-2}k_{j,i}p^j\left(\frac{\partial}{\partial t}e^{jt}\right)x_j^i\frac{1}{2^{i+1}}\beta_{j,i}+\right.$$

$$+\sum_{j=1}^{0,5(n-3)}\xi_j k_{j,n-2j}p^j\left(\frac{\partial}{\partial t}e^{jt}\right)x_j^{n-2j}\frac{1}{2^{n-2j+1}}\beta_{j,n-2j}+\xi_{0,5(n-1)}k_{\frac{n-1}{2},1}p^{\frac{n-1}{2}}\left(\frac{\partial}{\partial t}e^{\frac{n-1}{2}t}\right)x_{\frac{n-1}{2}}^1\Bigg]^{n-g-1}.$$

Суммирование моментов в функции (5) выполним аналогично предыдущим случаям, то есть при всех $k_{j,i}=1$, $\pi(g)=1$. Тогда, опять же при $n_1>>1$ и $n\geq n_1$, $n_1$ - нечетно, имеем:

$$P'_{ifc}=n\Bigg[\sum_{j=1}^{0,5(n_1-5)}\xi_j p^j e^{jt}x_j^1+\sum_{j=1}^{0,5(n_1-5)}\xi_j\sum_{i=2}^{n_1-2j-2}p^j e^{jt}x_j^i\frac{1}{2^{i+1}}\beta_{j,i}+$$
$$+\sum_{j=1}^{0,5(n_1-3)}\xi_j p^j e^{jt}x_j^{n_1-2j}\frac{1}{2^{n_1-2j+1}}\beta_{j,n_1-2j}+\xi_{0,5(n_1-1)}p^{\frac{n_1-1}{2}}e^{\frac{n_1-1}{2}t}x_{\frac{n_1-1}{2}}^1\Bigg]^{n-1}\times \quad (6)$$
$$\times\Bigg[\sum_{j=1}^{0,5(n_1-5)}\xi_j p^j\left(je^{jt}\right)x_j^1+\sum_{j=1}^{0,5(n_1-5)}\xi_j\sum_{i=2}^{n_1-2j-2}p^j\left(je^{jt}\right)x_j^i\frac{1}{2^{i+1}}\beta_{j,i}+$$
$$+\sum_{j=1}^{0,5(n_1-3)}\xi_j p^j\left(je^{jt}\right)x_j^{n_1-2j}\frac{1}{2^{n_1-2j+1}}\beta_{j,n_1-2j}+\xi_{0,5(n_1-1)}p^{\frac{n_1-1}{2}}\frac{n_1-1}{2}e^{\frac{n_1-1}{2}t}x_{\frac{n_1-1}{2}}^1\Bigg].$$

При $t=0$ и $x_j^i=1$ из (6) соответственно следует:

$$P'_{ifc}=n\Bigg[\sum_{j=1}^{0,5(n_1-5)}\xi_j p^j+\sum_{j=1}^{0,5(n_1-5)}\xi_j\sum_{i=2}^{n_1-2j-2}p^j\frac{1}{2^{i+1}}\beta_{j,i}+\sum_{j=1}^{0,5(n_1-3)}\xi_j p^j\frac{1}{2^{n_1-2j+1}}\beta_{j,n_1-2j}+\xi_{0,5(n_1-1)}p^{\frac{n_1-1}{2}}\Bigg]^{n-1}\times$$
$$\times\Bigg[\sum_{j=1}^{0,5(n_1-5)}\xi_j jp^j+\sum_{j=1}^{0,5(n_1-5)}\xi_j\sum_{i=2}^{n_1-2j-2}jp^j\frac{1}{2^{i+1}}\beta_{j,i}+\sum_{j=1}^{0,5(n_1-3)}\xi_j jp^j\frac{1}{2^{n_1-2j+1}}\beta_{j,n_1-2j}+\xi_{0,5(n_1-1)}\frac{n_1-1}{2}p^{\frac{n_1-1}{2}}\Bigg]. \quad (7)$$

Практическое моделирование данной задачи показало, что ВПЗВ с параметром $j=3$ встречаются достаточно редко, а параметр с числом векторов $j=4$ например при $r=8$ отсутствует в течение весьма длительного времени наблюдения. Таким образом, для верхней границы вероятности $P$ в соответствии с результатами [8] можем записать полиномиальные множители:

$$\xi_1=\frac{1}{1{,}146}\left(1-\frac{3}{16}\right),\quad \xi_2=\frac{1}{1{,}146}\cdot\frac{3}{16},\quad \xi_3=\frac{0{,}146}{1{,}146}.$$

Подставляя данные значения в многочлен (7) получаем:

$$P'_{ifc}\approx n\left[\frac{33}{8}\left(\xi_1 p+\xi_2 p^2+\xi_3 p^3\right)\right]^{n-1}\left[\frac{33}{8}\left(\xi_1 p+2\xi_2 p^2+3\xi_3 p^3\right)\right]= \quad (8)$$
$$=0{,}606n\left(0{,}576\right)^{n-1}.$$

Итак, математическое ожидание плотности распределения вероятностей ошибки при наблюдении ВПЗВ для максимального значения $p=\frac{3}{16}$ (то есть для нижней границы интеграла вероятностей $P_{ifc}$) определяется соотношением (8).

Рассмотрим математическое ожидание плотности распределения вероятностей ошибки для ВПЗВ, имеющих минимальное значении параметра $p=\frac{1}{m^2}$, что соответствует максимуму вероятности пропуска ошибки. При этом имеем:

$$P'_{ifc}=\frac{\partial}{\partial t}\left(1-\frac{\Delta}{2m}\right)^{n+1}+\frac{\partial}{\partial t}\frac{1}{2^n}\left(pe^t\right)^n\xrightarrow{t=0}n\left(\frac{1}{2}p\right)^{n-1}.$$

Принципиально, данное значение стремиться к нулю, хотя сама вероятность $p$ стремиться к аргументу моды распределения.

## Список литературы

[1] Варакин Л.Е. Системы связи с шумоподобными сигналами. М.: Радио и связь. 1985. 384 с.

[2] Яблонский С.В. Введение в дискретную математику. 2-е изд., перераб. и доп. М.: Наука. гл. ред. физ.-мат. лит. 384 с.

[3] Weathers G.D., Graf E.R. The Subsequence Weight Distributions of Summed Maximum Length Digital Sequences // IEEE Trans. on Commun. 1974. com-22. №8. P.997-1004.

[4] Кобяк И.П. Сравнительная оценка достоверности методов сигнатурного анализа и счета состояний // Электрон. моделирование. 1996. Т.18. №3. С. 58-62.

[5] Кобяк И.П. Теория внутрисхемного наблюдения СБИС с использованием автокорреляционных функций // АВТ. 2009. № 2. С.37-46.

[6] Кобяк И.П. Производящая функция для распределения статистик автокорреляционной функции // Электрон. моделирование. 2010. 32. №2. С. 61-76.

[7] Кобяк, И. П. Производящая функция для вероятности пропуска ошибки при наблюдении векторов переходов / И. П. Кобяк // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей IX Международной научно-практической конференции, Минск, 17–18 мая 2023 г. : в 2 ч. Ч. 2 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. Минск, 2023. С. 16-23.

[8] Кобяк, И.П. Интеграл вероятностей ошибки при идентификации выборки числом ВП / И. П. Кобяк // Информационные технологии и системы 2023 (ИТС 2023) = Information Technologies and Systems 2023 (ITS 2023) : материалы Международной научной конференции, Минск, 22 ноября 2023 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. Минск : БГУИР, 2023. С. 145–146.

# MATHEMATICAL EXPECTATION OF THE PROBABILITY DISTRIBUTION OF MISSING AN ERROR WHEN OBSERVING TRANSITION VECTORS

***I.P. Kobiack***

*PhD, Associate Professor,*
*Chair of* ECM, BSUIR.
*Belarusian State University of*
*Informatics and Radioelectronics,*
*Republic of Belarus*

**Abstract.** The paper investigates a method for identifying sequences of random events using estimates of the observation of transition vectors of a given type. The mathematical expectation of the density of the error probability distribution is determined, which is required by the task of comparing the statistical indicators of the proposed and other convolution algorithms. The basis for the calculations was an enumerating generating function, which allows us to represent the product of combinatorial moments characterizing the probability of missing an error in the form of the sum of the products of statistics for the corresponding moments. The relations obtained in the work characterize the method of observing transition vectors as the most effective algorithm for synthesizing estimates for identifying messages with a random or pseudorandom nature.

**Keywords:** transition vector, complex events, mono-objects and post-objects, sequence identification, probability of missing an error, signature analysis, state vector

УДК 004.7:628.14

# ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ МОНИТОРИНГА ИНЖЕНЕРНЫХ СЕТЕЙ: ПРИМЕНЕНИЕ ИНТЕРНЕТА ВЕЩЕЙ И АНАЛИТИКИ БОЛЬШИХ ДАННЫХ ДЛЯ ПРЕДОТВРАЩЕНИЯ ПРОТЕЧЕК В МНОГОКВАРТИРНЫХ ДОМАХ

***И.В. Тимошкевич***
*Аспирант ИКТ, БГУИР,*
*научный сотрудник*
*Института жилищно-*
*коммунального хозяйства*
*НАН Беларуси*
tim6iv@gmail.com

***И.В. Тимошкевич***

*Окончил: высшее образование первого уровня Белорусский национально технический университет, высшее образование второго уровня Университет НАН Беларуси по специальности экономика. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, создание информационных моделей многоквартирных жилых домов, повышение энергоэффективности многоквартирных зданий с применение технологий Интернета вещей.*

**Аннотация.** В статье рассматривается применение технологий Интернета вещей (IoT) и аналитики больших данных для мониторинга и предотвращения протечек в водопроводных и канализационных системах многоквартирных домов. Анализируются основные проблемы, связанные с утечками воды, и предлагаются интеллектуальные решения на основе датчиков и алгоритмов машинного обучения. Описаны принципы работы таких систем, их преимущества, включая снижение экономических потерь, повышение безопасности и экологическую устойчивость. Также обсуждаются вызовы, связанные с внедрением данных технологий, и перспективы их развития в контексте умных городов.

**Ключевые слова:** Интернет вещей, большие данные, мониторинг трубопроводов, предотвращение протечек, многоквартирные дома, интеллектуальные системы, умные города, инженерные сети.

**Введение.** Современные многоквартирные дома представляют собой сложные инженерные системы, где водопроводные и канализационные сети играют ключевую роль в обеспечении комфорта жильцов. Однако протечки в этих системах могут привести к серьезным последствиям: от повреждения имущества до аварийных ситуаций, требующих дорогостоящего ремонта. Внедрение интеллектуальных систем мониторинга, основанных на технологиях Интернета вещей и анализе больших данных, становится ключевым инструментом для предотвращения этих проблем. Традиционные методы мониторинга и предотвращения протечек зачастую недостаточно эффективны, так как основаны на ручном контроле и реагировании на уже возникшие проблемы. В этой статье мы рассмотрим, как технологии больших данных (Big Data) и расширенной аналитики, интегрированные с Интернетом вещей, могут революционизировать подход к мониторингу и предотвращению протечек в многоквартирных домах.

**Проблема протечек в многоквартирных домах.**

Протечки водопроводных и канализационных труб — это не только вопрос комфорта, но и серьезная экономическая и экологическая проблема. По данным исследований, до 20% воды в городских системах теряется из-за утечек. В многоквартирных домах это может привести к:

- повреждению стен, потолков и полов;
- образованию плесени и грибка;
- увеличению счетов за воду;
- риску аварийных ситуаций, таких как затопление подъездов или подвалов.

Традиционные методы обнаружения протечек, такие как визуальный осмотр или использование простых датчиков, часто недостаточно оперативны и точны. Это создает необходимость внедрения более интеллектуальных и автоматизированных решений [1].

**Технологии IoT в мониторинге трубопроводов.** Интернет вещей предлагает новый подход к мониторингу состояния трубопроводов. С помощью сети подключенных устройств, таких как датчики давления, температуры, влажности и вибрации, можно в реальном времени отслеживать состояние водопроводных и канализационных систем. Эти устройства собирают данные и передают их в централизованную систему для анализа. Технологии IoT обеспечивают беспроводное соединение между различными устройствами, что позволяет создавать распределённые сети датчиков [2]. Эти сети могут:

1 Автоматически собирать данные о состоянии трубопроводов без необходимости ручного контроля.

2 Передавать информацию в режиме реального времени, что позволяет быстро реагировать на изменения и предотвращать аварийные ситуации.

3 Интегрироваться с другими системами управления зданием (BMS), что обеспечивает комплексный подход к мониторингу всех инженерных сетей.

Преимущества IoT в мониторинге трубопроводов:

1 Реальное время: Данные поступают непрерывно, что позволяет оперативно реагировать на изменения.

2 Точность: Датчики могут обнаруживать даже незначительные изменения в давлении или влажности, указывающие на потенциальные протечки.

3 Масштабируемость: Система может быть легко расширена для охвата всего многоквартирного дома или даже целого микрорайона.

Умные счетчики обеспечивают точные данные, что исключает ошибки при начислении платежей и повышает доверие к системе расчетов.

**Роль больших данных и расширенной аналитики.** Сбор данных с IoT-устройств — это только первый шаг. Для эффективного использования этой информации необходимы технологии больших данных и расширенной аналитики. Аналитика больших данных играет важную роль в повышении эффективности ИСМ. Сбор больших объемов данных с датчиков позволяет:

1 Датчики генерируют огромное количество данных в секунду. Большие данные позволяют обрабатывать и анализировать эту информацию в реальном времени.

2 Выявлять закономерности: анализ исторических данных помогает выявить закономерности, связанные с утечками, можно выявлять аномалии, которые указывают на возможные протечки. Например, резкое падение давления в трубе может быть признаком разрыва.

3 Оптимизировать управление ресурсами: на основе анализа можно улучшить планирование технического обслуживания и минимизировать затраты.

4 Прогнозировать риски: Расширенная аналитика позволяет не только обнаруживать уже возникшие проблемы, но и прогнозировать потенциальные утечки на основе исторических данных и текущих показателей.

5 Создавать отчеты: Регулярная генерация отчетов о состоянии инженерных систем помогает управляющим компаниям принимать более обоснованные решения.

**Принципы и примеры работы интеллектуальных систем мониторинга.** Интеллектуальные системы мониторинга (ИСМ) представляют собой комплекс программно-аппаратных решений, которые осуществляют сбор, обработку и анализ данных о состоянии инженерных сетей зданий в режиме реального времени. Основные компоненты таких систем включают [2]:

1 Датчики: устанавливаются на трубопроводах и в потенциально уязвимых местах для обнаружения утечек. Они фиксируют изменения в таких параметрах, как влажность, давление и температура.

2 Облачные платформы: собранные данные передаются на облачные серверы для хранения и анализа. Это позволяет обрабатывать большие объемы информации и применять алгоритмы машинного обучения для прогнозирования возможных утечек.

3 Интерфейсы для пользователей: специальные приложения и панели управления позволяют управляющим компаниям и жильцам отслеживать состояние систем и получать уведомления о потенциальных проблемах.

Рассмотрим, как может работать интеллектуальная система мониторинга на практике:

1 Сбор данных: Датчики, установленные на трубах, собирают информацию о давлении, температуре, влажности и вибрации.

2 Передача данных: Данные передаются в облачную платформу через IoT-шлюзы.

3 Анализ данных: Система анализирует данные с помощью алгоритмов машинного обучения, выявляя аномалии.

4 Оповещение: При обнаружении потенциальной протечки система отправляет уведомление управляющей компании или жильцам.

5 Автоматическое реагирование: В некоторых случаях система может автоматически перекрыть воду в проблемном участке, чтобы минимизировать ущерб [3].

**Преимущества интеллектуальных систем мониторинга.** Внедрение интеллектуальных систем мониторинга на базе IoT и больших данных имеет множество преимуществ:

1 Снижение затрат: Раннее обнаружение протечек позволяет избежать дорогостоящего ремонта и снизить потери воды.

2 Повышение безопасности: Система минимизирует риск аварийных ситуаций, таких как затопления.

3 Экологическая устойчивость: Снижение потерь воды способствует более рациональному использованию ресурсов.

4 Удобство для жильцов: Автоматизация процессов мониторинга и реагирования избавляет жильцов от необходимости самостоятельно контролировать состояние труб.

**Вызовы и перспективы.** Несмотря на очевидные преимущества, внедрение интеллектуальных систем мониторинга сталкивается с рядом вызовов:

1 Высокая стоимость оборудования: Установка IoT-датчиков и создание инфраструктуры для сбора и анализа данных требует значительных инвестиций.

2 Защита данных: Необходимо обеспечить безопасность передачи и хранения данных, чтобы предотвратить утечки конфиденциальной информации.

3 Интеграция с существующими системами: В старых домах может быть сложно интегрировать новые технологии с устаревшими инженерными системами.

Однако с развитием технологий и снижением стоимости оборудования эти вызовы постепенно преодолеваются. В будущем можно ожидать массового внедрения интеллектуальных систем мониторинга в многоквартирных домах, что сделает их более безопасными, экономичными и экологически устойчивыми.

**Примеры успешного применения.** Система мониторинга SODIS Building FM: эта система объединяет данные с различных датчиков и предоставляет информацию о

состоянии инженерных систем в режиме реального времени. Она позволяет автоматически уведомлять управляющие компании о возникновении аварийных ситуаций [4,5].

Проект «Умный дом»: В рамках этого проекта используются IoT-датчики для мониторинга состояния водопроводных труб. При обнаружении утечки система автоматически перекрывает подачу воды и уведомляет жильцов через мобильное приложение [6,7].

Аналитические платформы: такие платформы, как AWS IoT Analytics, позволяют обрабатывать данные с множества устройств, обеспечивая глубокую аналитику и визуализацию состояния инженерных систем [8,9].

Внедрение IoT-систем позволяет оптимизировать использование ресурсов за счёт автоматического регулирования подачи воды в зависимости от потребностей жильцов. Это также способствует экономии воды и снижению затрат на коммунальные услуги.

Устройства IoT могут быть интегрированы с системами «умного дома», позволяя жильцам управлять инженерными системами через мобильные приложения. Например, жильцы могут получать уведомления о состоянии водопроводных систем и управлять ими удаленно.

Как работают IIoT-системы на стеке Big Data: архитектура и принципы действия. Современные промышленные системы Интернета вещей (IIoT) объединяют физические устройства, облачные вычисления и аналитику данных, создавая интеллектуальные решения для оптимизации процессов. Рассмотрим несколько примеров применения IIoT в сочетании с Big Data и Machine Learning (ML), которые демонстрируют их потенциал в реальных условиях.

Пример: Мониторинг качества воды в умных городах. Проблема загрязнения воды в городских канализационных системах.

Решение: Датчики: Сенсоры pH, мутности и содержания химикатов (BLE или LoRaWAN для передачи). Брокер сообщений: RabbitMQ управляет очередями данных с тысяч устройств, обеспечивая доставку без потерь. Облако: Данные хранятся в InfluxDB (для временных рядов) и анализируются в Grafana. ML: Нейросети (PyTorch) обнаруживают аномалии в составе воды, предупреждая о выбросах токсинов.

Результат: Сокращение времени реагирования на загрязнения с часов до минут. Ключевые технологии для масштабирования IIoT-решений. Распределенные брокеры сообщений:

– Apache Kafka — для обработки высоконагруженных потоков данных (например, с тысяч датчиков на конвейере);

– RabbitMQ — для гарантированной доставки сообщений в системах с жесткими требованиями к надежности;

– Edge Computing.

Локальная обработка данных на устройствах (NVIDIA Jetson, Raspberry Pi) снижает задержки и нагрузку на облако.

Гибридные облака:

Использование AWS Greengrass или Azure Edge для синхронизации edge-устройств и облачных аналитических моделей.

Вызовы интеграции IIoT и Big Data

Латентность: Реальные системы требуют обработки данных за миллисекунды (решение — Apache Flink для потоковой аналитики).

Безопасность: Шифрование данных на уровне MQTT (TLS/SSL) и аутентификация устройств через X.509-сертификаты.

Интероперабельность: Интеграция legacy-оборудования с современными протоколами (например, шлюзы OPC UA для систем на базе CAN bus).

На рисунке 1 представлена схематичная архитектура IIoT-системы [10].

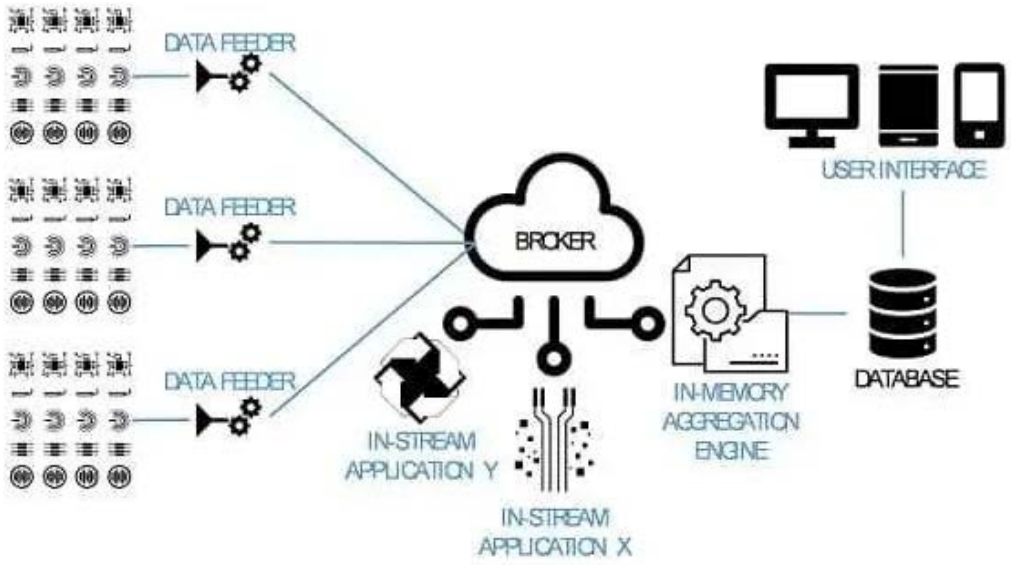



*Рисунок 1.* Схематичная архитектура *IIoT*-системы

Системы мониторинга протечек на основе технологий Big Data и Интернета вещей (IoT) позволяют значительно повысить эффективность управления инженерными сетями многоквартирных домов. Аналогично тому, как мобильные операторы анализируют телеметрию для обеспечения стабильной связи, интеллектуальные системы мониторинга водопроводных и канализационных труб собирают и обрабатывают огромные объемы данных. С помощью анализа этих данных можно выявлять утечки, прогнозировать аварии и оптимизировать затраты на обслуживание. Автоматизированные алгоритмы машинного обучения позволяют не только оперативно обнаруживать проблемные участки, но и предлагать экономически эффективные решения, например, локальный ремонт труб вместо полной замены системы. Такой подход способствует снижению потерь воды, уменьшению расходов на ремонт и повышению безопасности жилых объектов.

**Заключение.** Интеллектуальные системы мониторинга инженерных сетей и предотвращения протечек водопроводных и канализационных труб на базе технологий Интернета вещей и больших данных представляют собой важный шаг в развитии умных городов. Эти системы не только повышают комфорт жильцов, но и способствуют более рациональному использованию ресурсов, снижению затрат и повышению безопасности, что делает их незаменимыми в современном строительстве и эксплуатации зданий. Внедрение таких решений — это не только технологический прогресс, но и вклад в устойчивое развитие городской инфраструктуры.

IIoT-системы, объединенные с Big Data и Machine Learning, становятся основой для цифровой трансформации промышленности. Они позволяют не только реагировать на проблемы, но и прогнозировать их, оптимизируя ресурсы и снижая затраты. Успешная реализация таких проектов требует:

Правильного выбора протоколов (MQTT для датчиков, Kafka для потоков).

Использования облачных и edge-вычислений для баланса между скоростью и мощностью. Этот пример демонстрируют, как технологии IIoT переходят из области экспериментов в критически важные инструменты для промышленности 4.0.

## Список литературы

[1] Тимошкевич И.В. «Метод априорного ранжирования факторов при исследовании технологического процесса водоснабжения». // Сборник материалов XII Международного водно-химического форума Минск 2024. С. 159-164.

[2] Тимошкевич И.В. Основные принципы взаимодействия совместной работы BIM И IOT / Межвузовская научно-техническая конференция студентов, аспирантов и молодых специалистов имени Е.В. Арменского. // Материалы конференции. - М. ~: МИЭМ НИУ ВШЭ, 2024. –С. 155-159.

[3] И.В. Тимошкевич, К.А. Полтавцев Технология интернета вещей в жизни и в ЖКХ. / Технологии передачи и обработки информации = Technologies of Information Transmission and Processing : материалы Международного научно-технического семинара (Минск, апрель 2024 г.). – Минск : БГУИР, 2024. – 100-105 с.

[4] Система мониторинга инженерных систем (СМИС) / СОДИС Лаб [Электронный ресурс]. – Режим доступа: https://www.sodislab.com/ru/smis – Дата доступа: 15.12.2024.

[5] СМИС. Система мониторинга и управления инженерными системами / сайт СМИС эксперт [Электронный ресурс]. – Режим доступа: https://smis-expert.com/smis-sistema-monitoringa-inzhenernykh-sistem/ – Дата доступа: 03.01.2024.

[6] Интеллектуальные системы мониторинга и управления технически сложных объектов / Центр инженерного предпринимательства и инноватики [Электронный ресурс]. – Режим доступа: http://cipi.samgtu.ru/node/40 – Дата доступа: 09.01.2024.

[7] Система мониторинга инженерных систем / ЦЕНТР СМИС "БАЗИС" [Электронный ресурс]. – Режим доступа: https://basis-smis.ru/smis – Дата доступа: 13.01.2024.

[8] Мониторинг инженерных систем ЦОДа: что, зачем и как / Информационный портал [Электронный ресурс]. – Режим доступа: https://www.iksmedia.ru/articles/5937262-Monitoring-inzhenernyx-sistem-CZODa.html – Дата доступа: 18.01.2024.

[9] Система мониторинга инженерными системами — здания и сооружения [Электронный ресурс]. – Режим доступа: https://www.akruks.net/article/ustrojstvo_inzhiniringovyh_sis/p499-monitoring-inzhenernyh-sistem/ – Дата доступа: 25.01.2024.

[10] Industrial Internet of Things: как работает умное железо с Big Data / Школа больших данных сооружения [Электронный ресурс]. – Режим доступа: https://bigdataschool.ru/blog/news/цифровая-трансформация/industrial-iot-big-data-machine-learning-cases.html – Дата доступа: 03.02.2024.

## Авторский вклад

**И.В. Тимошкевич –** Полный анализ вопроса и возможностей в создании интеллектуальные системы мониторинга инженерных сетей: применение интернета вещей и аналитики больших данных для предотвращения протечек в многоквартирных домах

# INTELLIGENT ENGINEERING NETWORK MONITORING SYSTEMS: THE USE OF THE INTERNET OF THINGS AND BIG DATA ANALYTICS TO PREVENT LEAKS IN APARTMENT BUILDINGS

***I.V. Timoshkevich***
*Postgraduate student of ICT, BSUIR,*
*Researcher Institute of Housing and*
*Communal Services of the National*
*Academy of Sciences of Belarus*

**Abstract.** The article discusses the use of Internet of Things (IoT) technologies and big data analytics to monitor and prevent leaks in water and sewer systems of apartment buildings. The main problems associated with water leaks are analyzed and intelligent solutions based on sensors and machine learning algorithms are proposed. The principles of operation of such systems and their advantages, including reduction of economic losses, improvement of safety and environmental sustainability, are described. The challenges associated with the introduction of these technologies and the prospects for their development in the context of smart cities are also discussed.

**Keywords:** Internet of Things, big data, pipeline monitoring, leak prevention, apartment buildings, intelligent systems, smart cities, engineering networks.

УДК 519.714.5

# ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ НОВЫХ АЛГОРИТМОВ ВЫДЕЛЕНИЯ ПОДСИСТЕМ ИЗ СОВМЕСТНЫХ МНОГОУРОВНЕВЫХ ПРЕДСТАВЛЕНИЙ СИСТЕМ БУЛЕВЫХ ФУНКЦИЙ

***П.Н. Бибило***
*Заведующий лабораторией Объединенного института проблем информатики НАН Беларуси, доктор технических наук, профессор*
bibilo@newman.bas-net.by

***Н.А. Кириенко***
*Доцент кафедры экономической информатики БГУИР, кандидат технических наук, доцент*
kirienkonaminsk@mail.ru

***В.И. Романов***
*Ведущий научный сотрудник Объединенного института проблем информатики НАН Беларуси, кандидат технических наук, доцент*
rom@newman.bas-net.by

***П.Н. Бибило***

*С 1994 г. является заведующим лабораторией логического проектирования Объединенного института проблем информатики НАН Беларуси, с которым неразрывно связана вся его трудовая деятельность. Основные научные результаты Бибило П.Н. относятся к теории логического проектирования цифровых устройств, основные направления – разработка методов синтеза цифровых устройств с использованием современной элементной базы сверхбольших интегральных схем (СБИС), автоматизация процессов проектирования (синтеза, моделирования и функциональной верификации) заказных цифровых СБИС, разработка систем автоматизированного проектирования (САПР).*

***Н.А. Кириенко***

*Окончила Минский радиотехнический институт. Область научных интересов связана с автоматизацией процессов логического проектирования дискретных устройств, преобразованием и оптимизацией функциональных описаний логических схем.*

***В.И. Романов***

*Область научных интересов – разработка инструментария для решения задач логико-комбинаторного характера, разработка программного обеспечения САПР дискретных устройств и применения в них методов искусственного интеллекта.*

**Аннотация.** Рассматриваются два вида уравнений для совместных многоуровневых представлений систем полностью определенных булевых функций – BDDI-представления, полученные по разложениям Шеннона, и Bool-представления – уравнения, соответствующие булевым сетям. Вершинам графа булевой сети соответствуют логические операции «конъюнкция» либо «дизъюнкция» над литералами булевых переменных. Предлагаются два новых алгоритма выделения подсистем. Для полученных подсистем проводится схемная реализация без дополнительной оптимизации и с дополнительной оптимизацией, которая предполагает переход к матричному представлению каждой подсистемы в виде системы ДНФ и последующую минимизацию в виде BDDI-представления либо Bool-представления. Показано, что такой подход позволяет чаще получать схемы заказных КМОП СБИС меньшей площади и большего быстродействия по сравнению со схемами, для которых дополнительная минимизация подсистем не проводилась. Проведено сравнение новых экспериментальных результатов с полученными ранее.

**Ключевые слова**: система булевых функций, дизъюнктивная нормальная форма (ДНФ), Binary Decision Diagram (BDD), булева сеть, разложение Шеннона, мера связанности функций, синтез логической схемы, VHDL, заказные СБИС.

**Введение.** Актуальной задачей процесса автоматизированного проектирования цифровых устройств является технологически независимая оптимизация функциональных описаний блоков комбинационной логики заказных СБИС. Математическими моделями функционирования схем комбинационной логики являются системы булевых функций. Одним из направлений логической оптимизации многоуровневых представлений систем булевых функций, предваряющей синтез логических схем, являются методы, основанные на выделении подсистем функций, которые имеют одинаковые части в областях определения функций выделенных подсистем. Такие подсистемы называются связанными. Связанность функций по характеристическим множествам (областям единичных значений функций) приводит к появлению большого числа одинаковых структурных частей (конъюнкций, алгебраических выражений, подфункций и др.) в оптимизированных формах представления функций, по которым строятся в дальнейшем комбинационные логические схемы. Чем сильнее связаны функции выделенной подсистемы, тем скорее можно ожидать, что в представлениях функций данной подсистемы будет больше одинаковых подвыражений, и синтезированные логические схемы будут иметь меньшую сложность.

Алгоритм (программа) *Splitter1* выделения связанных подсистем по BDD-представлениям (BDD - Binary Decision Diagram) для заданной меры связанности логических уравнений разложений Шеннона был экспериментально исследован в [1]. В [2] была предложена программа *AutoSplit* выделения подсистем, позволяющая перебирать значения меры связанности, и находить подсистемы, характеризуемые большой долей (мерой) – числом одинаковых уравнений разложения Шеннона в описаниях выделяемых связанных подсистем. В [3] программы *Splitter1* и *AutoSplit* были исследованы для различных видов исходных логических уравнений и показана их эффективность для двух видов уравнений - BDDI-представлений, полученных по разложениям Шеннона, и Bool-представлений – уравнениям, соответствующим булевым сетям. Эти представления будут описаны далее и являются базовыми для проведения экспериментов. В данной работе предлагаются новые алгоритмы *Splitter3* и *Splitter4* выделения связанных подсистем, проводится экспериментальное исследование новых алгоритмов и сравнение с лучшими результатами, полученными в [3], для того же множества практических примеров систем булевых функций, что были рассмотрены ранее в [2, 3].

**BDDI- и Bool-представления системы булевых функций.** BDDI-представления основываются на разложениях Шеннона. *Разложением Шеннона* булевой функции $f(\boldsymbol{x})$ по переменной $x_i$ называется представление

$$f(\boldsymbol{x}) = \overline{x}_i f_0 \vee x_i f_1 .$$

Функции $f_0=f(x_1,\ldots, x_{i-1}, 0, x_{i+1},\ldots, x_n)$, $f_1=f(x_1,\ldots, x_{i-1},1,x_{i+1},\ldots, x_n)$ в правой части называются кофакторами (*cofactors*) разложения по переменной $x_i$. Каждый из кофакторов $f(x_1,\ldots,x_{i-1}, 0, x_{i+1},\ldots,x_n)$, $f(x_1,\ldots,x_{i-1}, 1, x_{i+1},\ldots,x_n)$ может быть разложен по одной из переменных из множества $\{x_1,\ldots,x_{i-1}, x_{i+1}, \ldots, x_n\}$. Процесс разложения кофакторов заканчивается, когда все $n$ переменных будут использованы для разложения.

Под *BDDI-представлением* (BDDI – *Binary Decision Diagram with Inverse cofactors*) понимается ориентированный бесконтурный граф, задающий последовательные разложения Шеннона булевой функции $f(\boldsymbol{x})=f(x_1,\ldots,x_n)$, $\boldsymbol{x}=(x_1,\ldots,x_n)$, либо системы $F(\boldsymbol{x})=\{f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x})\}$ булевых функций по всем переменным $x_1, x_2, \ldots, x_n$ при заданном порядке (перестановке) переменных, по которым проводятся разложения, и при условии нахождения пар взаимно инверсных кофакторов [4]. BDDI-представлению соответствует совокупность взаимосвязанных формул разложения Шеннона. Минимизация сложности BDDI [4] заключается в нахождении последовательности (*перестановки*) переменных разложений Шеннона, при которой число пар взаимно инверсных кофакторов является наименьшим.

*Bool-представление* системы булевых функций соответствует булевой сети (ориентированному бесконтурному графу), функциями вершин которой могут быть логические операции «конъюнкция» либо «дизъюнкция» над литералами булевых переменных. Литерал – это булева переменная либо ее отрицание. Логическая минимизация булевых сетей на основе разложения Шеннона заключается в поиске такой перестановки переменных разложения, при которой число литералов в булевой сети является наименьшим.

**Понятие связанности формульных представлений булевых функций.** Для многоуровневого формульного представления системы $F=\{f^1(\boldsymbol{x}),\ldots,f^m(\boldsymbol{x})\}$ булевых функций обозначим через $R(F)$ множество внутренних формул системы $F$, т. е. формул, задающих промежуточные (не выходные) булевы переменные для всех функций системы: $R(F)=R(f^1,\ldots,f^m)=\bigcup_{i=1}^{m} R(f^i)$.

Под мерой $q_{ur}=M_{ur}(F^m\cup\{f^{m+1}\})$ связанности системы функций $F^m\cup f^{m+1}$ ***по уравнениям*** будем понимать величину

$$q_{ur}=M_{ur}(F^m\cup\{f^{m+1}\})=\frac{\left|R(f^1,\ldots,f^m)\cap R(f^{m+1})\right|}{\max\left(\left|R(f^1,\ldots,f^m)\right|,\left|R(f^{m+1})\right|\right)}, \quad (1)$$

где через $|R(F)|$ обозначена мощность множества $R(F)$.

Через $S(F)$ обозначим число литералов в задании всех формул системы $F$, т. е. формул, задающих как промежуточные, так и выходные булевы переменные для всех функций системы $F$.

Для определения меры связанности системы функций $F^m\cup f^{m+1}$ ***по литералам*** предлагается использовать одно из следующих значений:

$$q_{lit}=M_{lit}(F^m\cup\{f^{m+1}\})=\frac{S(f^1,\ldots,f^m,f^{m+1})}{S(f^1,\ldots,f^m)+S(f^{m+1})}, \quad (2)$$

или

$$q_{lits}=M_{lits}(F^m\cup\{f^{m+1}\})=\frac{S(f^1,\ldots,f^m,f^{m+1})}{S(f^1)+S(f^2)+\ldots+S(f^m)+S(f^{m+1})}. \quad (3)$$

Алгоритмы выделения связанных подсистем основаны на построении подсистем связанных функций, для которых мера связанности удовлетворяет некоторому ограничению $q_{limit}$. Ранее в [2, 3] был исследован алгоритм выделения подсистем, удовлетворяющих критерию $q_{ur}>q_{limit}$ - это алгоритм *Splitter1*, использующий меру связанности (1). Выделение подсистем, удовлетворяющих критерию $q_{lit} < q_{limit}$ (алгоритм *Splitter2* [5], использует меру связанности (2).

В настоящей работе предлагаются новые алгоритмы выделения подсистем, использующие новые способы задания меры связанности формульных представлений функций.

**Новые алгоритмы выделения подсистем связанных функций из многоуровневых представлений систем булевых функций.** Разработан и исследован алгоритм *Splitter3* выделения подсистем, использующий меру связанности (3), и алгоритм *Splitter4*, использующий два ограничения при формировании связанной подсистемы.

**Алгоритм *Splitter3*** выделения подсистем для заданного ограничения $q_{limit}$ меры связанности состоит в последовательном формировании (на каждой итерации $i$) по текущей системе функций очередной подсистемы функций $P^i$ (подсистема $P^i$ характеризуется мерой связанности $q_{lit}$, не большей $q_{limit}$). На первой итерации $i$=1 текущую систему функций образуют функции исходной системы, а остаточная подсистема $P^{ost}$ является пустой. Перед выполнением итераций алгоритма *Splitter3* требуется составить список $W$, в котором имени каждой функции сопоставлено число литералов в уравнении, задающем функцию.

На каждой итерации требуется выполнить шаги 1 – 3.

*Шаг 1.* Рассматривать по очереди функции $f^i$ из списка $W$ и найти такую пару $L=\{f^i, f^j\}$, $i, j=1,...,m$, $i \neq j$, которая (пара) имеет наименьшее (но не большее $q_{limit}$) значение меры связанности $q_{lits}$ по формуле (3), в этом случае при подсчете меры связанности (3) полагается $F^m=\{f^i(\boldsymbol{x})\}$. Если таких пар функций несколько, то выбирается первая из них. Если указанной пары функций нет, то переход на шаг 4. Если для какой-то функции $f^i$ все остальные функции образуют с функцией $f^i$ такие пары $L$, которые имеют меру связанности, меньшую $q_{limit}$, то функция $f^i$ исключается из текущей системы и помещается в *остаточную* подсистему $P^{ost}$. Если система функций содержит $m$ функций и для всех $C_m^2 = m(m-1)/2$ пар выяснено, что их связанность $q_{lits}$ не превышает $q_{limit}$, то из системы нельзя выделить ни одной подсистемы с требуемой связанностью $q_{limit}$.

*Шаг 2.* Составить из функций найденной на шаге 1 пары $L$ формируемую подсистему $P^i$, исключив функции выбранной пары $L$ из текущей системы, и добавлять в формируемую подсистему поочередно те функции $f^r$, которые находятся с помощью следующей эвристики: из множества функций текущей системы выбирается та функция $f^r$, которая обеспечивает наименьшее (но не большее $q_{limit}$) возможное значение меры связанности $q_{lits}$ (3) для подсистемы $P^i \cup \{f^r\}$. Если таких функций несколько, то выбирается и добавляется в формируемую подсистему $P^i$ первая из них.

*Шаг 3.* Если нет ни одной функции $f^r$ такой, что подсистема $P^i \cup \{f^r\}$ имеет меру связанности $q_{lits}$, не большую $q_{limit}$, то закончить формирование подсистемы $P^i$ и объявить не входящие в нее функции текущей системой. Переход на шаг 1 для формирования подсистемы на итерации $i$+1.

*Шаг 4.* Закончить формирование подсистем, когда все функции текущей системы будут включены в формируемые подсистемы. Формирование подсистем заканчивается (алгоритм прекращает работу) также и тогда, когда в текущей подсистеме нельзя найти ни одной пары функций, характеризуемых мерой связанности, не большей $q_{limit}$, либо когда в текущей системе имеется только одна функция – эта функция добавляется в остаточную подсистему $P^{ost}$.

Конец алгоритма.

**Алгоритм *Splitter4*** выделения подсистем для заданного ограничения $q_{limit}$ меры связанности состоит в последовательном формировании (на каждой итерации $i$) по текущей (остаточной) системе функций очередной подсистемы $P^i$ функций.

Подсистема $P^i$ характеризуется тем, что каждая из ее функций обладает мерой связанности $q_{ur}$ (1), не меньшей $q_{limit}$, с каждой из $p$ функций подсистемы. Значение $p$ является параметром алгоритма.

На первой итерации ($i$=1) текущую систему функций образуют функции исходной системы. На каждой итерации требуется выполнить шаги 1 – 3.

*Шаг 1.* Рассмотреть $C_m^2 = m(m-1)/2$ неупорядоченную пару функций $\{f^i, f^j\}$, $i, j=1,...,m$, $i \neq j$, текущей системы и найти такую пару $L$ функций, которая имеет максимальное значение меры связанности $q_{ur}$, в этом случае при подсчете меры связанности (1) полагается $F^m=\{f^i(\boldsymbol{x})\}$. Если таких пар функций несколько, то выбирается первая из них. Если указанной пары функций нет, то переход на шаг 4.

*Шаг 2.* Составить из функций найденной на первом шаге пары $L$ формируемую подсистему $P^i$, исключив выбранную пару функций $L$ из текущей системы, и добавлять в

формируемую подсистему поочередно те функции $f^r$, которые находятся с помощью следующей эвристики: из множества функций текущей системы выбирается та функция $f^r$, которая обеспечивает значение меры связанности $q_{ur}$ (1) с каждой функцией подсистемы $P^i$, не меньшее заданного значения $q_{limit}$. Количество функций подсистемы $P^i$, с которыми связана подключаемая функция $f^r$ должно быть не меньше $p$.

Отличие от алгоритма *Splitter1* заключается в следующем. В алгоритме *Splitter1* за один проход цикла выбиралась функция с максимальным значением меры связанности, не меньшим заданного значения $q_{limit}$, и включалась в подсистему. В алгоритме *Splitter4* за один проход цикла выбирается функция, которая обеспечивают значение меры связанности (1) не меньшее заданного значения $q_{limit}$ с каждой из $p$ функций подсистемы $P^i$.

*Шаг 3.* Если нет ни одной функции $f^r$ такой, что обеспечивает значение меры связанности $q_{ur}$ (1) не меньшее заданного значения $q_{limit}$ с каждой из $n$ функций подсистемы $P^i$, то закончить формирование подсистемы $P^i$ и объявить не входящие в нее функции текущей подсистемой. Переход на шаг 1 для формирования подсистемы на итерации $i$+1.

*Шаг 4.* Закончить формирование подсистем, когда все функции текущей системы будут включены в формируемые подсистемы. Формирование подсистем заканчивается (алгоритм прекращает работу) также в следующих случаях:

- в текущей системе нельзя найти ни одной функции, характеризуемой мерой связанности, не меньшей $q_{limit}$ с каждой из $p$ функций подсистемы $P^i$ ;

- в текущей системе имеется только одна функция - текущая подсистема объявляется *остаточной.*

Конец алгоритма.

При экспериментальном исследовании параметры $q_{limit}$ и $p$ задавались в процентах.

**Результаты экспериментального исследования.** Входными данными были логические уравнения, задающие те же ДНФ-представления 39 примеров исходных систем булевых функций, что были испытаны в работах [2, 3, 5], это были примеры из библиотеки примеров [6]. Состав проведенных десяти экспериментов и используемых в них алгоритмов (и реализующих эти алгоритмы программ) дан в табл. 1. В экспериментах использовались система FLC-2 [7] и следующие *программы*:

*BDD_Builder* – программа минимизации совместных BDDI-представлений системы булевых функций, исходными для программы *BDD_Builder* являются исходные матричные представления (формат SDF системы FLC-2);

*BoolNetOpt2* - программа минимизации совместных Bool-представлений системы булевых функций, исходными для программы BoolNetOpt2 являются логические уравнения, поэтому исходные матричные представления (формат SDF системы FLC-2) переводились в формат LOG логических уравнений;

*Splitter3, Splitter4* – программы выделения подсистем функций при заданном значении $q_{limit}$ меры связанности и заданной доли $p$ функций (для программы *Splitter4*) из совместных BDDI- либо из совместных Bool-представлений систем функций;

*AutoSplit* – программа нахождения лучшего значения $q$ меры связанности (и доли $p$ функций для программы *Splitter4*) по критерию минимальности общего числа литералов в BDDI- либо в Bool-представлениях систем функций подсистем функций, полученных программами выделения подсистем.

Для полученных подсистем в экспериментах 3, 5, 8, 10 проводилась дополнительная оптимизация, которая предполагает переход к матричному представлению каждой подсистемы в виде системы ДНФ и последующую совместную минимизацию BDDI-представления (эксперименты 3 и 5) либо Bool-представления (эксперименты 8 и 10).

Таблица 1. Состав экспериментов

| Уравнения для системы | Алгоритм | Оптимизация выделенных подсистем | |
|---|---|---|---|
| BDDI (базовые решения) | - | Эксперимент 1. Подсистемы не выделялись | |
| | *Splitter3* | Эксперимент 2. Нет оптимизации | Эксперимент 3. Совместная BDDI |
| | *Splitter4* | Эксперимент 4. Нет оптимизации | Эксперимент 5. Совместная BDDI |
| Bool (базовые решения) | - | Эксперимент 6. Подсистемы не выделялись | |
| | *Splitter3* | Эксперимент 7. Нет оптимизации | Эксперимент 8. Совместная Bool |
| | *Splitter4* | Эксперимент 9. Нет оптимизации | Эксперимент 10. Совместная Bool |

Результаты экспериментальных исследований представлены в таблицах 2 – 5.

Совместные BDDI-представления систем булевых функций, полученные в эксперименте 1, являлись исходными данными для экспериментов 2 – 5, аналогично, совместные Bool-представления, полученные в эксперименте 6, являлись исходными данными для экспериментов 7 – 10. Значение параметра *Area* площади синтезированной схемы в библиотеке КМОП элементов [4] задается в условных единицах, значение *Delay* – время задержки схемы в наносекундах (нс). Параметры примеров систем функций заданы в табл. 6: $n$ – число аргументов; $m$ – число функций системы ДНФ, заданной на $k$ общих элементарных конъюнкциях. Если $k = 2^n$, то исходной является матричная форма таблицы истинности системы функций.

*В табл. 2 и 3 одним символом * и полужирным шрифтом выделено лучшее решение (в рамках этих двух таблиц) для каждого примера – меньшее значения параметра площади либо задержки схемы. Аналогично для пары таблиц 4, 5. Символом # для каждого примера отмечено лучшее решение по всем четырем табл. 2 – 5.*

Для алгоритмов Splitter3(табл. 2 и 4), Splitter4 (табл. 3 и 5) указано значение (в процентах) параметра меры связанности $q$, для Splitter4 указано еще и значение параметра $p$ (также в процентах). Текст «Нет подс.» означает, что рассматриваемые алгоритмы не смогли выделить ни одной подсистемы из совместного представления системы функций.

Сравнение лучших решений, полученных в данной работе с помощью алгоритмов Splitter3, Splitter4 (помеченных в табл. 2 - 3 символом #), с лучшими решениями, полученными в работе [3], представлено в табл. 6. Например, для системы функций – примера Add6 в табл. 2 указано значение площади **#*14 826** и соответствующее этой схеме значение ***5,76** задержки, эти параметры выбраны и занесены в табл. 6 для сравнения лучших результатов. Аналогично выбирались и лучшие результаты, полученные в статье [3], – выбиралась схема меньшей площади и соответствующее ей значение задержки. Остается заметить, что лучшее значение #*3,81 для примера Add6 было получено в эксперименте 7. Это означает, что для данного примера схема наименьшей площади не обладает наименьшей задержкой. В табл. 6 лучшие решения помечены символом * и выделены полужирным шрифтом.

**Заключение.** Рассмотрены два вида уравнений для совместных многоуровневых представлений систем полностью определенных булевых функций. Предлагаются два новых алгоритма выделения подсистем для их совместной многоуровневой минимизации. Показано, что предлагаемые алгоритмы являются конкурентоспособными с известным алгоритмом и позволяют в бóльшей половине случаев получать лучшие по площади схемы комбинационных блоков заказных КМОП СБИС. Показано, что дополнительная оптимизация выделенных подсистем часто позволяет улучшать решения по площади и задержке схем.

Таблица 2. Результаты экспериментов 1 – 3 для **BDDI-представлений** систем булевых функций

| № | Пример | Эксперимент 1 (базовые решения) | | Splitter3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Эксперимент 2 | | | Эксперимент 3 | | |
| | | *Area* | *Delay* | *q*, % | *Area* | *Delay* | *q*, % | *Area* | *Delay* |
| 1 | Add6 | 16 534 | 6.82 | 70 | 16 534 | 6,82 | 70 | 19 340 | 7,02 |
| 2 | B12 | 17 064 | 3.04 | 95 | 16 411 | 3,34 | 95 | 15 736 | **#*2,11** |
| 3 | B9 | 32 247 | 5.32 | 90 | 32 035 | 5,68 | 90 | ***27 364** | ***4,37** |
| 4 | Dist | 62 808 | 6.41 | 80 | **#*62 262** | ***5,12** | 80 | 68 327 | 5,41 |
| 5 | Gary | **#*91 395** | ***7.32** | - | Нет подс. | - | - | Нет подс. | - |
| 6 | Ibm | **#*66 776** | ***5.62** | - | Нет подс. | - | - | Нет подс. | - |
| 7 | In0 | 93 058 | ***7.45** | - | Нет подс. | - | - | Нет подс. | - |
| 8 | In1 | ***166 758** | ***7.54** | 65 | 168 616 | 7,58 | 65 | 181 389 | 8,50 |
| 9 | In2 | **#*70 291** | 7.32 | - | Нет подс. | - | - | Нет подс. | - |
| 10 | Intb | 235 883 | 9.51 | 85 | 238 450 | 10,13 | 85 | 287 627 | ***8,80** |
| 11 | Jbp | 98 308 | **#*5.22** | - | Нет подс. | - | 65 | **#*94 821** | 5,56 |
| 12 | M181 | 16 595 | 3.60 | 95 | 16 048 | 3,94 | 95 | 16 416 | ***3,04** |
| 13 | M3 | 51 632 | 5.13 | 75 | 52 865 | 4,92 | 75 | 59 650 | ***4,86** |
| 14 | M4 | **#*78 181** | 5.87 | 55 | 78237 | 5,91 | 55 | 80 408 | **#*4,93** |
| 15 | Max1024 | 148 846 | 7.66 | 85 | ***148 629** | ***7,33** | 85 | 226 559 | 9,50 |
| 16 | Max512 | ***84 682** | 5.91 | 85 | 85 279 | 6,06 | 85 | 86 300 | 7,56 |
| 17 | Mlp4 | ***71 095** | 6.09 | 90 | ***71 095** | 6,07 | 90 | 76 357 | ***5,33** |
| 18 | Mp2d | 18 805 | 4.87 | 55 | 18 369 | 4,87 | 55 | **#*16 835** | 5,05 |
| 19 | Mult_7 | 2 492 681 | 14.28 | 90 | 2 509 945 | ***13,61** | 90 | 3 159 932 | 13,66 |
| 20 | Mult_8 | 7 069 051 | ***15.99** | - | Нет подс. | - | - | Нет подс. | - |
| 21 | P82 | 21 355 | 3.49 | 80 | 21 355 | 3,49 | 80 | ***19 402** | ***3,44** |
| 22 | Psevdo1 | ***766 686** | 9.01 | 90 | 774 750 | 9,23 | 90 | 840 024 | 8,93 |
| 23 | Radd | 11 802 | 3.69 | 65 | 11 802 | 3,69 | 65 | 10 117 | **#*3,34** |
| 24 | Root | 26 075 | 4.78 | - | Нет подс. | - | - | Нет подс. | |
| 25 | Sist_4 | 112 722 | 7.21 | 95 | 112 337 | 8,20 | 95 | **#*102 371** | **#*6,71** |
| 26 | Soar | 163 957 | 6.14 | - | Нет подс. | - | 55 | **#*138 217** | 6,33 |
| 27 | Sqn | 22 878 | 3.99 | - | Нет подс. | - | - | Нет подс. | - |
| 28 | Tial | 279 458 | 9.08 | 90 | 278 626 | 9,17 | 90 | 286 405 | **#*7,75** |
| 29 | Tms | **#*39 004** | 4.65 | 70 | **#*39 004** | 4,65 | 70 | 44 216 | **#*4,42** |
| 30 | Ttt2_matr | 76 429 | ***5.01** | 75 | 79 152 | 5,76 | 75 | **#*50 136** | 7,39 |
| 31 | Verg_1 | 465 032 | 14.41 | 55 | 479 897 | 14,30 | 55 | **#*386 387** | **#*13,63** |
| 32 | Verg_2 | 536 037 | 14.15 | 60 | 539 201 | 13,95 | 60 | **#*430 503** | 15,15 |
| 33 | Vtx1 | 31 600 | **#*4.23** | - | Нет подс. | - | - | Нет подс. | - |
| 34 | X1_matr | 107 912 | 6.45 | 90 | | | 90 | ***89 503** | ***5,72** |
| 35 | X3_matr | 211 633 | 5.73 | 95 | **#*209 161** | 6,78 | 95 | 236 514 | **#*4,97** |
| 36 | X4_matr | ***120 031** | 6.17 | 75 | 124 138 | 6,86 | 75 | 181 417 | 8,37 |
| 37 | X9dn | 34 719 | 6.58 | - | Нет подс. | - | - | Нет подс. | - |
| 38 | Z4 | ***6 640** | ***4.25** | - | Нет подс. | - | - | Нет подс. | - |
| 39 | Z5xp1 | 22 582 | 5.24 | 85 | 22 582 | 5,06 | 85 | 25 919 | **#*3,64** |
| **Лучших решений *** | | **11** | **9** | **-** | **5** | **3** | **-** | **10** | **16** |
| **Лучших решений #** | | **5** | **2** | **-** | **3** | **0** | **-** | **7** | **9** |

Таблица 3. Результаты экспериментов 1, 4, 5 для **BDDI-представлений** систем булевых функций

| № | Пример | Эксперимент 1 (базовые решения) | | Splitter4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Эксперимент 4 | | | Эксперимент 5 | | |
| | | *Area* | *Delay* | *q, p %* | *Area* | *Delay* | *q, p %* | *Area* | *Delay* |
| 1 | Add6 | 16 534 | 6.82 | 5,5 | 16 534 | 6,82 | 5,5 | **#*14 826** | ***5,76** |
| 2 | B12 | 17 064 | 3.04 | 5,5 | 16 155 | 3,40 | 5,5 | ***15 507** | 3,40 |
| 3 | B9 | 32 247 | 5.32 | 5,5 | 32 035 | 5,68 | 5,5 | 31 722 | 6,27 |
| 4 | Dist | 62 808 | 6.41 | 5,5 | 62 903 | 5,95 | 5,5 | 62 808 | 6,41 |
| 5 | Gary | **#*91 395** | ***7.32** | 5,5 | 91 947 | 7,58 | 5,5 | 93 058 | 8,12 |
| 6 | Ibm | **#*66 776** | ***5.62** | - | Нет подс. | - | - | Нет подс. | - |
| 7 | In0 | 93 058 | ***7.45** | 5,5 | ***92 031** | 7,60 | 5,5 | 93 058 | 8,12 |
| 8 | In1 | ***166 758** | ***7.54** | 5,20 | 168 639 | **#*7,54** | 5,20 | 170 436 | 7,87 |
| 9 | In2 | **#*70 291** | 7.32 | 5,5 | 74 683 | 7,24 | 5,5 | 70 620 | ***6,84** |
| 10 | Intb | 235 883 | 9.51 | 10,5 | 239 962 | 10,13 | 10,5 | **#*234 600** | 9,17 |
| 11 | Jbp | 98 308 | **#*5.22** | 55,5 | 100 596 | 5,89 | 55,5 | 100 596 | 5,89 |
| 12 | M181 | 16 595 | 3.60 | 5,5 | 16 193 | 3,40 | 5,5 | ***15 875** | 3,22 |
| 13 | M3 | 51 632 | 5.13 | 5,5 | 51 157 | 5,28 | 5,5 | **#*50 650** | 5,34 |
| 14 | M4 | **#*78 181** | 5.87 | 60,5 | 78 237 | 5,91 | 60,5 | 78 237 | 5,91 |
| 15 | Max1024 | 148 846 | 7.66 | 10,55 | 150 370 | 7,80 | 10,55 | 149 656 | 7,39 |
| 16 | Max512 | ***84 682** | 5.91 | 10,55 | 84 905 | 6,09 | 10,55 | 85 904 | ***5,74** |
| 17 | Mlp4 | ***71 095** | 6.09 | 5,5 | ***71 095** | 6,05 | 5,5 | 72 696 | 6,56 |
| 18 | Mp2d | 18 805 | 4.87 | 5,5 | 18 358 | 4,45 | 5,5 | 18 447 | ***3,94** |
| 19 | Mult_7 | 2 492 681 | 14.28 | 5,5 | 2 513 282 | 14,14 | 5,5 | ***2 486 186** | 14,23 |
| 20 | Mult_8 | 7 069 051 | ***15.99** | 5,5 | 6 998 369 | 16,30 | 5,5 | ***6 010 503** | 16,53 |
| 21 | P82 | 21 355 | 3.49 | 5,5 | 21 355 | 3,49 | 5,5 | 21 355 | 3,49 |
| 22 | Psevdo1 | ***766 686** | 9.01 | 10,70 | 768 740 | 9,29 | 10,70 | 771 675 | **#*8,38** |
| 23 | Radd | 11 802 | 3.69 | 5,5 | 11 802 | 3,69 | 5,5 | *9 480 | 3,85 |
| 24 | Root | 26 075 | 4.78 | 45,5 | **#*24 585** | 5,21 | 45,5 | 27 046 | ***4,36** |
| 25 | Sist_4 | 112 722 | 7.21 | 5,5 | 110 891 | 7,83 | 5,5 | 115 188 | 7,42 |
| 26 | Soar | 163 957 | 6.14 | 5,35 | 162 690 | **#*5,23** | 5,35 | 164 080 | 6,47 |
| 27 | Sqn | 22 878 | 3.99 | 15,55 | 22 526 | 3,74 | 15,55 | **#*20 975** | **#*3,55** |
| 28 | Tial | 279 458 | 9.08 | 5,5 | **#*272 884** | 8,55 | 5,5 | 294 724 | 8,73 |
| 29 | Tms | **#*39 004** | 4.65 | 5,35 | **#*39 004** | 4,65 | 5,35 | 42 068 | 4,58 |
| 30 | Ttt2_matr | 76 429 | ***5.01** | 20,5 | 83 627 | 5,87 | 20,5 | 76 429 | **#*5,01** |
| 31 | Verg_1 | 465 032 | 14.41 | 5,5 | 466 985 | 14,09 | 5,5 | 456 974 | 14,24 |
| 32 | Verg_2 | 536 037 | 14.15 | 5,5 | 536 936 | **#*13,9** | 5,5 | 564 043 | 14,04 |
| 33 | Vtx1 | 31 600 | **#*4.23** | 5,5 | 31 600 | **#*4,23** | 5,5 | ***28 285** | 4,66 |
| 34 | X1_matr | 107 912 | 6.45 | 5,5 | 101 539 | 8,22 | 5,5 | 120 232 | 8,34 |
| 35 | X3_matr | 211 633 | 5.73 | 70,5 | 209 680 | 6,06 | 70,5 | 211 912 | 5,76 |
| 36 | X4_matr | ***120 031** | 6.17 | 5,5 | 124 663 | 6,87 | 5,5 | 120 383 | **#*6,09** |
| 37 | X9dn | 34 719 | 6.58 | 5,5 | ***34 696** | 6,55 | 5,5 | 37 224 | ***6,20** |
| 38 | Z4 | ***6 640** | ***4.25** | 5,5 | ***6 640** | ***4,25** | 5,5 | ***6 640** | ***4,25** |
| 39 | Z5xp1 | 22 582 | 5.24 | 5,5 | 22 582 | 5,24 | 5,5 | **#*22 287** | 4,46 |
| **Лучших решений *** | | **11** | **9** | **-** | **7** | **5** | **-** | **12** | **11** |
| **Лучших решений #** | | **5** | **2** | **-** | **3** | **4** | **-** | **5** | **4** |

Таблица 4. Результаты экспериментов 6 – 8 для **Bool-представлений** систем булевых функций

| № | Пример | Эксперимент 6 (базовые решения) | | Splitter3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Эксперимент 7 | | | Эксперимент 8 | | |
| | | *Area* | *Delay* | *q*, % | *Area* | *Delay* | *q*, % | *Area* | *Delay* |
| 1 | Add6 | ***18 314** | 6,18 | 85 | 19 664 | **#*3,81** | 85 | 23 001 | 5,75 |
| 2 | B12 | 15 936 | 3,55 | 90 | **#*15 396** | 3,18 | 90 | 19 949 | 2,99 |
| 3 | B9 | **#*22 454** | 4,69 | 90 | | 4,69 | 90 | 28 106 | **#*3,65** |
| 4 | Dist | 66 134 | 5,33 | 85 | ***64 282** | 6,09 | 85 | 64 628 | 5,32 |
| 5 | Gary | 97 795 | 7,94 | 80 | 96 545 | 7,16 | 80 | ***92 249** | 6,80 |
| 6 | Ibm | **#*66 776** | 5,62 | 80 | 89 235 | 6,75 | - | Нет подс | - |
| 7 | In0 | 95.792 | **#*6,20** | - | Нет подс. | - | - | Нет подс | - |
| 8 | In1 | 162 478 | 8,52 | 65 | 164 247 | 9,13 | 65 | 176 663 | ***8,32** |
| 9 | In2 | 93 716 | 6,25 | 80 | 93 727 | **#*6,15** | 80 | 115 752 | 7,65 |
| 10 | Intb | 253 299 | 9,30 | 95 | 270 418 | **#*8,62** | 95 | ***237 870** | 8,99 |
| 11 | Jbp | 98 308 | **#*5,22** | 75 | 99 486 | 6,93 | 75 | 100 652 | 6,29 |
| 12 | M181 | 16 204 | 3,02 | 90 | 16 204 | 3,02 | 90 | 15 853 | 3,77 |
| 13 | M3 | 65 414 | 5,14 | 85 | 65 710 | 5,14 | 85 | 71 240 | 6,20 |
| 14 | M4 | ***93 153** | ***5,24** | 65 | 93 867 | 5,83 | 65 | 96 043 | 5,59 |
| 15 | Max1024 | 129 478 | 7,38 | 90 | 128 457 | 7,58 | 90 | **#*121 789** | 7,61 |
| 16 | Max512 | 72 300 | 6,26 | 85 | 72 746 | 6,63 | 85 | 77 032 | **#*5,62** |
| 17 | Mlp4 | 68 773 | ***4,95** | 85 | 68 773 | **#*4,95** | 85 | 68 121 | 5,41 |
| 18 | Mp2d | ***17 203** | 3,81 | 55 | ***17 203** | 3,81 | 55 | 17 253 | **#*3,67** |
| 19 | Mult_7 | 2 298 056 | 13,34 | 95 | **#*2 155 063** | 13,31 | 95 | 2 537 499 | 13,50 |
| 20 | Mult_8 | **#*5 938 638** | **#*15,43** | - | Нет подс. | - | - | Нет подс. | - |
| 21 | P82 | **#*19 056** | 3,23 | 75 | 19 167 | 3,26 | 75 | 19 926 | **#*2,80** |
| 22 | Psevdo1 | **#*735 796** | ***8,91** | | Нет подс. | - | - | Нет подс. | - |
| 23 | Radd | **#*7 399** | ***3,99** | 90 | 7 622 | 4,08 | 90 | **#*7 399** | ***3,99** |
| 24 | Root | 31 773 | 4,10 | - | Нет подс. | - | - | Нет подс. | - |
| 25 | Sist_4 | 128 351 | ***7,38** | 85 | ***127 403** | 7,70 | 85 | 133 881 | 8,44 |
| 26 | Soar | 151 787 | 6,52 | 55 | 152 691 | 6,52 | 55 | ***141 805** | ***5,69** |
| 27 | Sqn | 26 695 | 4,23 | - | Нет подс | - | - | Нет подс. | - |
| 28 | Tial | 351 607 | 9,79 | 95 | 356 858 | 9,92 | 95 | 344 040 | 9,70 |
| 29 | Tms | 50 471 | 6,88 | 65 | 51 436 | 7,44 | 65 | 53 010 | ***5,02** |
| 30 | Ttt2_matr | 52 653 | 4,05 | 90 | 52 725 | ***4,01** | 90 | ***50 823** | 5,83 |
| 31 | Verg_1 | 634 954 | 19,25 | 55 | 641 427 | 17,18 | 55 | 755 889 | **#*15,55** |
| 32 | Verg_2 | ***694 425** | 18,90 | 55 | 698 013 | 18,51 | 55 | 800 021 | ***15,68** |
| 33 | Vtx1 | 28 112 | ***4,31** | 70 | 28 112 | ***4,31** | - | Нет подс. | - |
| 34 | X1_matr | 107 912 | 6,45 | 90 | 83 215 | 5,35 | 90 | 103 515 | 7,56 |
| 35 | X3_matr | ***211 633** | ***5,73** | 80 | 214 908 | 7,08 | 80 | 258 851 | 7,19 |
| 36 | X4_matr | 120 031 | ***6,17** | 95 | **#*114 284** | 6,75 | 95 | 125 260 | 6,81 |
| 37 | X9dn | **#*20 741** | 5,48 | 55 | **#*20 741** | 5,48 | 55 | 41 214 | 6,13 |
| 38 | Z4 | 6 685 | 3,83 | - | Нет подс. | - | - | Нет подс. | - |
| 39 | Z5xp1 | 26 014 | ***3,76** | 85 | ***25 361** | 4,27 | 85 | 28 871 | 4,65 |
| **Лучших решений *** | | **12** | **12** | **-** | **8** | **6** | **-** | **6** | **10** |
| **Лучших решений #** | | **7** | **3** | **-** | **4** | **4** | **-** | **2** | **5** |

Таблица 5. Результаты экспериментов 6, 9, 10 для **Bool-представлений** систем булевых функций

| № | Пример | Эксперимент 6 (базовые решения) | | Splitter4 | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Эксперимент 9 | | | Эксперимент 10 | | |
| | | *Area* | *Delay* | *q, p %* | *Area* | *Delay* | *q, p %* | *Area* | *Delay* |
| 1 | Add6 | ***18 314** | 6,18 | 5,5 | ***18 314** | 6,18 | 5,5 | 18 766 | 5,21 |
| 2 | B12 | 15 936 | 3,55 | 5,5 | **#*15 396** | 3,18 | 5,5 | 15 652 | ***2,80** |
| 3 | B9 | **#*22 454** | 4,69 | 5,5 | 22 510 | 4,79 | 5,5 | 25 049 | 4,55 |
| 4 | Dist | 66 134 | 5,33 | 10,30 | 66 435 | 5,64 | 10,30 | 66 664 | **#*4,90** |
| 5 | Gary | 97 795 | 7,94 | 10,65 | 97 064 | 6,86 | 10,65 | 97 008 | **#*6,72** |
| 6 | Ibm | **#*66 776** | 5,62 | 25,5 | 89 235 | 6,75 | 25,5 | 66 899 | **#*5,50** |
| 7 | In0 | 95.792 | **#*6,20** | 5,25 | 96 238 | 6,36 | 5,25 | **#*91 674** | 6,42 |
| 8 | In1 | 162 478 | 8,52 | 5,55 | 162 478 | 8,52 | 5,55 | **#*161 491** | 9,48 |
| 9 | In2 | 93 716 | 6,25 | 5,5 | 93 727 | **#*6,15** | 5,5 | ***81 859** | 8,39 |
| 10 | Intb | 253 299 | 9,30 | 5,5 | 266 345 | 8,80 | 5,5 | 245 018 | 9,35 |
| 11 | Jbp | 98 308 | **#*5,22** | 60,5 | 100 579 | 7,05 | 60,5 | ***97 639** | 6,45 |
| 12 | M181 | 16 204 | 3,02 | 5,5 | 16 204 | 3,02 | 5,5 | **#*15 356** | **#*2,56** |
| 13 | M3 | 65 414 | 5,14 | 5,5 | 65 414 | 5,14 | 5,5 | ***59 606** | **#*4,67** |
| 14 | M4 | ***93 153** | ***5,24** | 45,5 | 93 867 | 5,83 | 45,5 | 99 402 | 6,62 |
| 15 | Max1024 | 129 478 | 7,38 | 5,5 | 128 457 | 7,58 | 5,5 | 124 021 | **#*7,28** |
| 16 | Max512 | 72 300 | 6,26 | 5,55 | 71 882 | 6,38 | 5,55 | **#*70 208** | 6,33 |
| 17 | Mlp4 | 68 773 | ***4,95** | 5,5 | 68 506 | 5,66 | 5,5 | **#*67 066** | 5,82 |
| 18 | Mp2d | ***17 203** | 3,81 | 5,5 | 17 661 | 3,81 | 5,5 | 17 599 | 4,15 |
| 19 | Mult_7 | 2 298 056 | 13,34 | - | Нет подс. | - | 5,5 | 2 581 224 | **#*12,99** |
| 20 | Mult_8 | **#*5 938 638** | **#*15,43** | - | Нет подс. | - | - | Нет подс. | - |
| 21 | P82 | **#*19 056** | 3,23 | 60,5 | 19 257 | 3,02 | 60,5 | 20 886 | 3,30 |
| 22 | Psevdo1 | **#*735 796** | ***8,91** | 10,80 | 740 918 | 9,31 | 10,80 | 748 814 | 9,83 |
| 23 | Radd | **#*7 399** | ***3,99** | 5,5 | 7 622 | 4,08 | 5,5 | **#*7 399** | ***3,99** |
| 24 | Root | 31 773 | 4,10 | 10,55 | 31 449 | 4,07 | 10,55 | ***26 209** | **#*3,77** |
| 25 | Sist_4 | 128 351 | ***7,38** | 5,5 | 128 351 | ***7,38** | 5,5 | 128 044 | 7,65 |
| 26 | Soar | 151 787 | 6,52 | 5,15 | 152 580 | 6,52 | 5,15 | 145 683 | 6,31 |
| 27 | Sqn | 26 695 | 4,23 | 15,5 | 27 175 | ***3,75** | 15,5 | ***22 164** | 4,44 |
| 28 | Tial | 351 607 | 9,79 | 5,5 | 354 179 | 9,79 | 5,5 | ***310 538** | ***9,03** |
| 29 | Tms | 50 471 | 6,88 | 5,5 | 50 957 | 8,52 | 5,5 | **#*47 921** | 6,21 |
| 30 | Ttt2_matr | 52 653 | 4,05 | 5,5 | 52 725 | ***4,01** | 5,5 | 51 775 | 5,22 |
| 31 | Verg_1 | 634 954 | 19,25 | 95,5 | 633 073 | 16,28 | 95,5 | ***580 711** | 17,65 |
| 32 | Verg_2 | ***694 425** | 18,90 | 5,10 | 703 113 | 19,88 | 5,10 | 707 248 | 16,32 |
| 33 | Vtx1 | 28 112 | ***4,31** | 5,5 | 28 112 | ***4,31** | 5,5 | **#*22 621** | 5,58 |
| 34 | X1_matr | 107 912 | 6,45 | 10,55 | **#*82 071** | **#*5,32** | 10,55 | 103 665 | 7,52 |
| 35 | X3_matr | ***211 633** | ***5,73** | 25,5 | 220 181 | 6,38 | 25,5 | 269 302 | 7,87 |
| 36 | X4_matr | 120 031 | ***6,17** | 15,5 | 115 344 | 6,37 | 15,5 | 121 030 | 6,37 |
| 37 | X9dn | **#*20 741** | 5,48 | 5,70 | **#*20 741** | 5,48 | 5,70 | 32 849 | **#*5,04** |
| 38 | Z4 | 6 685 | 3,83 | 10,70 | 6 685 | 3,83 | 10,70 | **#*6 411** | **#*3,74** |
| 39 | Z5xp1 | 26 014 | ***3,76** | 5,5 | 26 014 | ***3,76** | 5,5 | 26 639 | 4.05 |
| **Лучших решений *** | | **12** | **12** | **-** | **4** | **7** | **-** | **16** | **13** |
| **Лучших решений #** | | **7** | **3** | **-** | **3** | **2** | **-** | **9** | **10** |

Таблица 6. Сравнение лучших схемных реализаций

| № | Пример | *n* | *m* | *k* | Статья [3] Splitter1 | | Splitter3, Splitter4 | |
|---|---|---|---|---|---|---|---|---|
| | | | | | *Area* | *Delay* | *Area* | *Delay* |
| 1 | Add6 | 12 | 7 | 1 092 | 15 976 | ***6,48** | ***14 826** | 5,76 |
| 2 | B12 | 15 | 9 | 431 | 15 529 | ***2,73** | ***15 396** | 3,18 |
| 3 | B9 | 16 | 5 | 123 | 22 755 | ***3,61** | ***22 510** | 4,79 |
| 4 | Dist | 8 | 5 | 256 | 62 457 | *5,47 | ***62 262** | ***5,12** |
| 5 | Gary | 15 | 11 | 442 | ***91 473** | ***6,49** | 91 947 | 7,58 |
| 6 | Ibm | 48 | 17 | 173 | ***59 667** | ***5,32** | 66 899 | 5,50 |
| 7 | In0 | 15 | 11 | 138 | 93 007 | 7,6 | ***91 674** | ***6,42** |
| 8 | In1 | 16 | 17 | 110 | ***151 698** | 9,64 | 161 491 | ***9,48** |
| 9 | In2 | 19 | 10 | 137 | 73 037 | ***5,89** | *70 620 | 6,84 |
| 10 | Intb | 15 | 7 | 664 | ***233 283** | ***8,82** | 234 600 | 9,17 |
| 11 | Jbp | 36 | 57 | 166 | 95 440 | ***4,98** | ***94 821** | 5,56 |
| 12 | M181 | 15 | 9 | 430 | ***15 295** | ***2,31** | 15 356 | 2,56 |
| 13 | M3 | 8 | 16 | 128 | *51 632 | ***5,13** | ***50 650** | 5,34 |
| 14 | M4 | 8 | 16 | 256 | ***78 237** | **5,91** | ***78 237** | ***5,91** |
| 15 | Max1024 | 10 | 6 | 1 024 | 129 272 | ***7,09** | ***121 789** | 7,61 |
| 16 | Max512 | 9 | 6 | 512 | ***68 489** | ***5,84** | 70 208 | 6,33 |
| 17 | Mpl4 | 8 | 8 | 256 | ***66 179** | ***4,93** | 67 066 | 5,82 |
| 18 | Mp2d | 14 | 14 | 123 | 17 237 | ***4,34** | *16 835 | 5,05 |
| 19 | Mult_7 | 14 | 14 | 13 060 | ***2 062 212** | ***12,61** | 2 155 063 | 13,31 |
| 20 | Mult_8 | 16 | 16 | 52 810 | ***5 914739** | 17,13 | 6 010 503 | ***16,53** |
| 21 | P82 | 5 | 14 | 24 | ***18 922** | ***2,96** | 19 167 | 3,26 |
| 22 | Psevdo1 | 10 | 20 | 1 000 | ***729 507** | ***8,42** | 740 918 | 9,31 |
| 23 | Radd | 8 | 5 | 120 | ***7 399** | ***3,99** | ***7 399** | ***3,99** |
| 24 | Root | 8 | 5 | 256 | ***24 585** | ***5,21** | ***24 585** | ***5,21** |
| 25 | Sist_4 | 17 | 12 | 370 | 114 569 | *6,79 | ***102 371** | ***6,71** |
| 26 | Soar | 83 | 94 | 529 | 141 983 | 7,36 | ***138 217** | ***6,33** |
| 27 | Sqn | 7 | 3 | 96 | 21 505 | 3,81 | ***20 975** | ***3,55** |
| 28 | Tial | 14 | 8 | 640 | 294 724 | *8,73 | ***272 884** | ***8,55** |
| 29 | Tms | 8 | 16 | 30 | 40 751 | 6,23 | ***39 004** | ***4,65** |
| 30 | Ttt2_matr | 24 | 21 | 222 | ***44 941** | ***4,16** | 50 136 | 7,39 |
| 31 | Verg_1 | 17 | 61 | 2 004 | 467 956 | 15,83 | ***386 387** | ***13,63** |
| 32 | Verg_2 | 18 | 63 | 2 129 | 589 968 | ***14,06** | ***430 503** | 15,15 |
| 33 | Vtx1 | 27 | 6 | 110 | ***20 518** | ***5,44** | 22 621 | 5,58 |
| 34 | X1_matr | 51 | 35 | 324 | 101 818 | 5,76 | ***82 071** | ***5,32** |
| 35 | X3_matr | 135 | 99 | 915 | ***179 364** | ***5,19** | 209 161 | 6,78 |
| 36 | X4_matr | 94 | 71 | 371 | ***93 180** | 5,51 | 114 284 | 6,75 |
| 37 | X9dn | 27 | 7 | 120 | ***20 741** | ***5,48** | ***20 741** | ***5,48** |
| 38 | Z4 | 7 | 4 | 128 | 6 640 | 4,25 | ***6 411** | ***3,74** |
| 39 | Z5xp1 | 7 | 10 | 128 | ***20 267** | 4,64 | 22 287 | ***4,46** |
| Лучших решений * | | | | | **20** | **25** | **23** | **17** |

## Список литературы

[1] Бибило П.Н., Позняк А.М. Выделение подсистем связанных функций из многоуровневого представления системы булевых функций // Информатика. 2020. Т. 17. № 1. С. 63 – 77.

[2] Бибило П.Н., Кириенко Н.А., Романов В.И. Выделение из многоуровневого представления системы булевых функций подсистем для совместной логической минимизации // Программные продукты и системы. 2023. Т. 36. № 4. С. 509–522. doi: 10.15827/0236-235X.142.

[3] Бибило П.Н., Кириенко Н.А., Романов В.И. Экспериментальное исследование алгоритма выделения подсистем булевых функций для совместной многоуровневой оптимизации // Программная инженерия. 2024. Том 15, № 4. С. 176—189.

[4] Бибило П.Н. Бинарные диаграммы решений в логическом проектировании. М.: ЛЕНАНД, 2024. - 560 с.

[5] Бибило П.Н., Кириенко Н.А., Романов В.И. Алгоритмы выделения из многоуровневого представления системы булевых функций подсистем для совместной минимизации // Информатика. – 2024. – Т. 21, № 4. – С. 7–23.

[6] http://www1.cs.columbia.edu/~cs6861/sis/espresso-examples/ex

[7] Бибило П.Н., Романов В.И. Система логической оптимизации функционально-структурных описаний цифровых устройств на основе продукционно-фреймовой модели представления знаний // Проблемы разработки перспективных микро- и наноэлектронных систем. – 2020. Сб. трудов / под общ. ред. акад. РАН А.Л. Стемпковского. М.: ИППМРАН. 2020. N 4. С. 9–16.

## Авторский вклад

**Бибило Петр Николаевич** – руководство исследованием, выполнение экспериментов, подготовка текста доклада

**Кириенко Наталья Алексеевна** – разработка программных средств для выделения из многоуровневого представления системы булевых функций связанных подсистем, подготовка текста доклада

**Романов Владимир Ильич** – разработка программных средств для анализа результатов экспериментов и выбора лучших решений.

# EXPERIMENTAL RESEARCH OF NEW ALGORITHMS FOR EXTRACTING SUBSYSTEMS FROM JOINT MULTILEVEL REPRESENTATIONS OF SYSTEMS OF BOOLEAN FUNCTIONS

***P.N. Bibilo,***
*Head of Laboratory,*
*UIIP of NAS of Belarus,*
*Dr.Sc. (Engineering),*
*Professor*

***N.A. Kirienko,***
*Associate Professor, Department of economic informatics*
*PhD of Technical sciences,*
*Associate Professor*

***V.I. Romanov,***
*Leading Researcher,*
*UIIP of NAS of Belarus,*
*PhD of Technical sciences,*
*Associate Professor*

**Abstract.** Two types of equations for joint multilevel representations of systems of completely defined Boolean functions are considered: BDDI-representations obtained from Shannon expansions and Bool-representations, which are equations corresponding to Boolean networks. The vertices of the graph of a Boolean network correspond to the logical operations “conjunction” or “disjunction” over literals of Boolean variables. Two new algorithms for extracting subsystems are proposed. For the resulting subsystems, a circuit implementation is carried out without additional optimization and with additional optimization. Additional optimization involves moving to a matrix representation of each subsystem in the form of a DNF system and following minimization in the form of a BDDI-representation or a Bool-representation. It is shown that additional optimization of subsystems often makes it possible to obtain custom CMOS VLSI circuits of smaller area and higher performance compared to circuits for which additional minimization of subsystems was not carried out. The new experimental results are compared with those obtained previously.

**Key words:** Boolean function system, DNF, Binary Decision Diagram, Boolean network, Shannon expansion, measure of function connectivity, logic synthesis, VHDL, ASIC.

УДК 004.67

# ПРЕОДОЛЕНИЕ ПРОБЛЕМ BIG DATA С ПОМОЩЬЮ DATA LAKE

***Е.А. Алуев***
*Инженер-исследователь*
*АТЕК, бакалавр*
*технических наук,*
alooeff@gmail.com

***Е.А. Алуев***

*Выпускник Брестского государственного технического университета. Область научных интересов связана с облачными технологиями, машинным обучением и мульти-агентным моделированием.*

**Аннотация.** Произведен анализ актуальных проблем предприятий по хранению и обработке неструктурированных данных. Рассмотрены способы решения этих проблем при помощи использования озер данных, рассмотрены особенности их использования. Сделаны выводы о перспективах перехода на использование озер данных в некоторых актуальных областях.

**Ключевые слова:** Big Data, Data Lake, озеро данных, обработка информации. база данных.

**Введение.** Фирмы, независимо от вида деятельности, имеют много разных видов данных. Есть PDF-файлы со спецификациями для выпускаемой продукции, CSV-файлы с транзакциями продаж, клиентами и товарами, а также JSON-файлы с обзорами продуктов и журналами. Основная проблемы хранения и обработки данных предприятия - их разнородность.

Современный путь решения этой проблемы - вводить эти данные в озеро данных (Data Lake) и выводить из него, объединять их все вместе и получать необходимую информацию.

Что именно представляет собой озеро данных? Это централизованное хранилище, которое позволяет хранить структурированные и неструктурированные данные в любом масштабе. Это означает, что сколько бы данных ни было, возможно добавить их все в централизованное хранилище. В озере данных можно хранить данные как есть, без необходимости их предварительной структуризации и запускать различные типы аналитики. Основная идея Data Lake заключается в том, что нет необходимости заранее определять схему, то есть структуру этих данных, поэтому возможно обрабатывать необработанные данные, не зная, какие сведения нужно будет получить в будущем.

**Ключевые характеристики озера данных.** Масштабируемость. Озера данных могут хранить петабайты данных и масштабироваться по мере необходимости без существенного снижения производительности. Эта возможность имеет важное значение в современном мире, управляемом данными, где объем данных, которые мы генерируем и анализируем, продолжает расти. Далее следует экономическая эффективность. Озера данных часто используют объектное хранилище, которое стоит дешевле традиционного хранилища, что делает его экономически эффективным решением для хранения огромных объемов данных. Также гибкость. В отличие от традиционных баз данных, которые требуют форматирования и структурирования данных предопределенным образом, озера данных могут хранить данные в различных форматах, включая PDF, XML, JSON, изображения и многое другое. Эта гибкость позволяет хранить данные из различных источников. И, наконец, доступность.

Централизуя хранение данных, озера данных облегчают различным потребителям данных, таким как специалисты по данным и аналитики, быстрый доступ и анализ больших объемов данных с помощью инструментов и фреймворков, которые интегрируются в среду озера данных. Озера данных служат основой для расширенной аналитики, предоставляя специалистам по данным гибкость для доступа и анализа различных форматов данных. Эта среда идеально подходит для разработки и совершенствования моделей и алгоритмов машинного обучения. Аналогичным образом озера данных все чаще используются для приложений на основе ИИ. Возможность хранить и управлять большими наборами данных из различных данных позволяет группам данных обучать более эффективные модели ИИ. Озера данных часто используются для помощи в обработке и аналитике данных в реальном времени, позволяя группам быстро реагировать на условия бизнеса и рынка. Например, розничные компании анализируют взаимодействие с клиентами и транзакции, поскольку они предлагают персонализированные акции или динамическое ценообразование. В эпоху ИИ и расширенной аналитики традиционные системы управления данными часто не могут справиться с масштабом или обеспечить необходимую гибкость. Озера данных решают проблемы, предлагая более гибкий и масштабируемый подход к их хранению и анализу. Они адаптируются к меняющимся аналитическим требованиям по мере их возникновения.

**Необходимость создания озер данных.** С появлением Интернета, электронной коммерции и социальных сетей традиционные системы хранения и обработки данных были перегружены. Потребность в технологии, которая обеспечивала бы не только объем, но и разнообразие и скорость обработки данных, стала очевидной, но за этим стояли значительные проблемы. Традиционные базы данных и хранилища данных были структурированы для последовательной, предсказуемой загрузки структурированных данных, таких как имена клиентов и контактная информация, финансовые транзакции и информация о запасах и т. д. Однако новые типы данных, такие как электронные письма, видео, сообщения в социальных сетях, данные датчиков и т. д., были неструктурированными и не вписывались в строки и столбцы. Требовалось решение, которое было бы гибким и разнообразным, как и сами данные. На основании Google MapReduce и Google File System, был разработан Apache Hadoop для хранения и обработки огромных объемов данных в кластерах серверов массового потребления. Он обеспечил хранение их в необработанной форме в распределенных средах. Это было концептуальным началом того, что позже назовут озером данных. Сам термин «озеро данных» был придуман Джеймсом Диксоном, техническим директором Pentaho. Он использовал его для описания большого хранилища необработанных данных, хранящихся в своем исходном формате до тех пор, пока они не понадобятся. В отличие от хранилищ данных, озера данных были разработаны так, чтобы быть гибкими, масштабируемыми и способными обрабатывать сложность и характер данных реального мира. Эволюция озер данных также тесно связана с быстрым развитием облачных вычислений. Поскольку облачные сервисы, такие как Amazon Web Services, Microsoft Azure и Google Cloud, становились все более надежными, они предоставили идеальную экосистему для озер данных [1]. Эти платформы предлагают масштабируемые, безопасные и экономичные решения для хранения данных, позволяя компаниям и командам всех размеров создавать озера данных без необходимости в дорогостоящей локальной инфраструктуре. Концепция демократизации данных, делающая данные доступными для всех без посредников, является еще одним ключевым фактором в развитии озер данных. Позволяя хранить данные в центральном месте, доступном для различных заинтересованных сторон в компании, озера данных обеспечивают среду для совместной работы, в которой идеи и принятие решений доступны для всех отделов.

**Основные компоненты архитектуры** проще представлять в виде уровней.

**Уровень хранения** является основой озера данных. Он предназначен для хранения огромного количества необработанных данных в различных форматах, от структурированных до неструктурированных. Этот уровень должен быть

высокомасштабируемым, чтобы приспосабливаться к росту объема данных, не снижая производительности. Есть три основные характеристики, которыми должен обладать уровень хранения: он должен иметь возможность расширяться для хранения больших объемов данных по мере необходимости, он должен быть оптимизирован, чтобы поддерживать управляемые затраты, он должен иметь возможность поддерживать несколько форматов и структур данных.

**Уровень приема** отвечает за прием данных из различных источников, будь то потоковая передача в реальном времени или пакетами. Основные характеристики уровня приема: он должен гарантировать, что данные собираются точно и надежно без потерь, он должен быть способен обрабатывать большие объемы данных на высокой скорости, он также должен поддерживать различные методы приема для работы с различными источниками данных.

**Уровень обработки** — это то, где данные преобразуются, агрегируются или объединяются. Этот уровень позволяет применять бизнес-правила и преобразования данных для получения информации из необработанных данных, хранящихся в озере, в дальнейшем. Этот уровень можно рассматривать как часть уровня приема или отдельно.

Основные характеристики уровня обработки: он должен быть оптимизирован для быстрой обработки больших наборов данных, он должен поддерживать несколько потребностей обработки от пакетной обработки больших наборов данных до аналитики в реальном времени, он должен иметь возможность масштабирования для удовлетворения требований к обработке по мере увеличения объема или сложности данных.

**Уровень управления** гарантирует, что озеро данных остается организованной, безопасной и хорошо управляемой средой. Он включает такие аспекты, как управление метаданными, каталогизация данных, безопасность и соответствие требованиям. Основные характеристики надежного уровня управления: он должен предоставлять механизмы доступа к данным и авторизации для защиты конфиденциальных данных, он также должен позволять вам отслеживать происхождение данных и метаданные для прозрачности и проверяемости, и он должен гарантировать, что методы управления данными соответствуют нормативным требованиям.

**Уровень аналитики** — это то место, где данные становятся применимыми, предоставляя инструменты и интерфейсы для конечных пользователей для запроса, визуализации и анализа данных, хранящихся в озере. Этот уровень обеспечивает следующее: удобные для пользователя инструменты для доступа к данным и их анализа, совместимость с различными инструментами бизнес-аналитики и аналитики и обычно оптимизирован для быстрого и эффективного выполнения запросов.

**Типы хранилищ.** В озере данных выбранные типы хранения играют важную роль в том, как данные поддерживаются, доступны и используются. Понимание этих типов очень важно, поскольку это может помочь оптимизировать стратегию работы с данными для баланса стоимости, доступности и производительности. Основные типы хранения, используемые в озерах данных:

**Объектное хранилище** на сегодняшний день является самым важным типом хранения для озер данных, особенно в облачных средах. Здесь данные хранятся в виде отдельных единиц или объектов, каждый из которых имеет уникальный идентификатор, что позволяет извлекать их из распределенного пула ресурсов хранения. Этот тип хранилища отличается высокой масштабируемостью и экономичностью, что делает его идеальным для хранения огромных объемов неструктурированных данных, таких как фотографии, видео и большие наборы данных. Примерами являются Amazon S3, Google Cloud Storage и Microsoft Azure Blob Storage.

**Файловое хранилище** организует данные в иерархии файлов и каталогов, что позволяет вам получать доступ к данным и управлять ими способом, который знаком большинству пользователей компьютеров. Хотя он не так масштабируем, как объектное

хранилище, файловые системы хранения, такие как сетевое хранилище или NAS, часто используются для небольших или устаревших приложений, которым требуются традиционные интерфейсы файловой системы.

**Блочное хранилище** делит данные на блоки одинакового размера, каждый из которых имеет свой собственный уникальный адрес, но без каких-либо метаданных для предоставления большего контекста о содержимом. Этот тип хранилища часто используется для приложений с высокой производительностью, таких как базы данных или системы ERP. Однако он может быть более дорогим и сложным в управлении, что делает его менее распространенным для общего использования озер данных.

Озера данных также могут интегрировать **гибридные решения** для хранения, которые объединяют элементы объектного, файлового и блочного хранения. Эти решения разработаны для обеспечения гибкости объектного хранилища с преимуществами производительности блочного хранилища. Например, распределенная файловая система Hadoop или HDFS или другие программно-определяемые решения для хранения могут эффективно управлять данными в различных типах хранилищ [2]. Каждый тип хранилища в озере данных имеет свой собственный набор преимуществ и идеальных вариантов использования.

**Хостинг хранилища.** Традиционные локальные решения для хранения данных включают физическую инфраструктуру, расположенную в центре обработки данных компании. Локальное хранилище обеспечивает полный контроль над средой хранения, что актуально для отраслей с высокими требованиями к безопасности и соблюдению требований. Преимуществами этого подхода являются повышенная безопасность и контроль над оборудованием, потенциально более высокая производительность для локализованной обработки данных. С точки зрения вариантов использования это чаще всего наблюдается в строго регулируемых отраслях, таких как финансы и здравоохранение, где безопасность являются приоритетом номер один. Облачное хранилище предоставляет масштабируемые и гибкие варианты хранения данных, размещенные на внешних платформах, предоставляемых такими поставщиками, как Amazon Web Services, Microsoft Azure или Google Cloud platform. Сильными сторонами этого подхода являются масштабируемость, экономичность и доступность из любой точки мира. Облачное хранилище также предлагает различные сервисы, которые можно интегрировать, такие как ИИ и аналитические инструменты. Кроме того, оно переводит затраты из капитальных в операционные, делая расходы более предсказуемыми и управляемыми. С точки зрения вариантов использования облачное хранилище используется в стартапах и предприятиях, которым нужны масштабируемые и гибкие решения без накладных расходов на поддержание физической инфраструктуры. Гибридные решения объединяют локальное и облачное хранилище, чтобы использовать преимущества обеих сред. Такой подход позволяет хранить конфиденциальные данные локально, используя облако для менее конфиденциальных данных. Сильными сторонами этого подхода являются гибкость, масштабируемость и возможность хранить критически важные данные внутри компании, при этом по-прежнему используя облачные технологии. С точки зрения вариантов использования гибридное хранилище отлично подходит для компаний, переходящих в облако; тех, у кого есть динамические потребности в хранении или которым требуется баланс между соответствием требованиям, безопасностью и масштабируемостью.

**Структуры папок.** Иерархии папок помогают организовать данные структурированным образом, упрощая навигацию, управление и извлечение. Хорошо продуманная структура очень важна в таких средах, как озера данных, где разнообразие и объем данных могут быстро привести к хаосу. Эффективные структуры папок сокращают время и вычислительные ресурсы, необходимые для поиска и доступа к данным. Хорошая структура также упрощает применение политик управления, таких как хранение данных, архивирование и управление жизненным циклом. Она также позволяет применять

политики безопасности на уровне папок, что невероятно важно для соответствия требованиям и защиты данных. Создание эффективной иерархии папок в вашем озере данных подразумевает понимание характера данных, потребностей пользователей и операционной динамики вашей компании. Основой хорошей иерархии папок является логическая группировка, которая классифицирует данные таким образом, чтобы они соответствовали тому, как пользователи взаимодействуют с ними. Также важно подумать о разделении данных на основе требований конфиденциальности и доступа, чтобы оптимизировать управление безопасностью. Еще одна концепция, которую следует учитывать при проектировании озера данных, - это архитектура медальона. Этот подход организует данные в разные слои, как правило, бронзовый, серебряный и золотой. Бронзовый слой — это место, где необработанные данные поступают из разных источников. Он хранится в своем исходном формате и часто используется для исторических записей. Серебряный слой содержит чистые и обогащенные данные. Здесь применяются преобразования, чтобы сделать данные более пригодными для анализа, а золотой слой — это последний тщательно отобранный слой данных. Он оптимизирован для бизнес-аналитики и аналитики, часто содержит агрегированные и специфичные для бизнеса данные. Еще один аспект, который следует учитывать, — это проектирование иерархий, которые будут достаточно гибкими для учета будущих изменений объема данных, новых типов данных и меняющихся бизнес-потребностей. Для данных, привязанных ко времени, таких как журналы или данные транзакций, организация папок по годам, кварталам, месяцам или дням может значительно ускорить запросы, привязанные к времени.

**Форматы файлов.** В озерах данных форматы файлов определяют, как данные хранятся и как осуществляется к ним доступ. Выбор формата файла может повлиять на все: от затрат до производительности запросов. Наиболее распространенные форматы:

**CSV** — один из самых простых и широко используемых форматов файлов. Он основан на тексте, прост для понимания и поддерживается почти всеми приложениями для обработки данных. Однако, поскольку в файлах CSV отсутствует какой-либо тип сжатия данных или принудительное применение схемы, они не оптимизированы для больших наборов данных или сложных запросов.

Файлы **JSON** очень гибкие и поддерживают иерархический, полуструктурированный формат данных, который идеально подходит для данных в различных схемах. JSON широко используется в веб-приложениях и для обмена данными между серверами и веб-клиентами. Хотя JSON легко читается и гибок, он может быть менее эффективным для хранения и запросов больших наборов данных.

**Apache Parquet** — это столбчатый формат файла хранения, оптимизированный для использования в озерах данных. Он предлагает отличные схемы сжатия и кодирования, которые сокращают накладные расходы на хранение и повышают производительность запросов, позволяя эффективно читать определенные столбцы. Parquet особенно хорошо подходит для сложных, вложенных структур данных и широко используется в фреймворках обработки больших данных, таких как Apache Hadoop, Spark и Presto.

**ORC** — это еще один столбчатый формат хранения, который обеспечивает высокоэффективный способ хранения и обработки данных. Как и Parquet, ORC поддерживает расширенное сжатие и кодирование, что делает его идеальным для высокоскоростных запросов. Файлы ORC предназначены для работы в средах, требующих тяжелых операций чтения и записи, что делает их популярными в системах, которые выполняют частое обновление данных.

**Apache Avro** предназначен для сериализации данных, предоставляя компактный и быстрый двоичный формат данных. Он также поддерживает эволюцию схемы, когда схемы хранятся внутри данных, что позволяет легко интегрировать сериализованные данные с приложениями. Avro — хороший выбор для конвейеров данных, которым требуются надежные, сложные структуры данных и быстрое получение данных.

Каждый формат файла предлагает разные преимущества и может быть лучшим выбором в зависимости от конкретных требований. CSV и JSON превосходны для простоты и гибкости. С другой стороны, Parquet, ORC и Avro обеспечивают лучшую производительность и эффективность для крупномасштабных операций с данными. Хотя это может быть оптимальным для хранения, возможно добавление практически любого типа файла в озеро данных.

**Преобразование данных.** Озера данных являются идеальной архитектурой для хранения необработанных данных, но необработанные данные все равно чаще всего необходимо подготовить для расширенной аналитики. Вот где вступает в дело преобразование данных. Преобразование данных в озерах данных включает изменение формата, структуры или значений данных. Основная цель — преобразовать необработанные данные в формат, который больше подходит для запросов и аналитики. Преобразование данных имеет несколько дополнительных преимуществ. Оно может улучшить качество данных за счет исправления ошибок, заполнения пропущенных значений и стандартизации форматов. Оно также может улучшить согласованность данных, упрощая интеграцию с другими источниками данных. Рассмотрим наиболее распространенные типы преобразования:

**Очистка** включает исправление или удаление неточных, неполных или противоречивых данных из набора данных. Примерами методов очистки являются стандартизация, которая требует корректировки данных для соответствия таким нормам, как форматирование дат, и дедупликация, которая включает в себя выявление и удаление дубликатов записей для предотвращения избыточности. Методы нормализации стандартизируют форматы данных, значения и диапазоны для добавления согласованности во всем наборе данных. Это особенно полезно в озерах данных из-за разнообразия источников и форматов данных. Примером нормализации будет преобразование всех валютных значений в единый валютный формат.

**Агрегация** включает в себя суммирование данных в более управляемую форму. Этот метод действительно помогает ускорить анализ и отчетность. Примером агрегации будет суммирование данных о продажах для предоставления ежемесячных итогов продаж.

**Объединение** означает объединение данных из разных источников или наборов данных для предоставления единого представления или единого производного набора данных. Этот метод также помогает ускорить и упростить аналитические запросы, поскольку объединения выполняются не при каждом анализе. Примером может служить объединение информации о клиентах из систем CRM с данными о транзакциях из баз данных продаж. Поскольку потребности в обработке данных становятся все более сложными, для обработки сложных сценариев часто используются передовые методы.

**Обогащение** данных дополняет исходный набор данных дополнительными источниками данных для предоставления более точной информации. Например, включение демографической информации в существующий набор данных для анализа моделей покупок в различных сегментах клиентов. Алгоритмы машинного обучения также могут применяться для преобразования или обогащения данных на основе изученных моделей и прогнозов. Примером будет использование методов кластеризации для сегментации клиентов на основе поведения. Создание чистых и хорошо отформатированных наборов данных, агрегирование метрик, а также слияние различных источников данных повышают элегантность и эффективность последующих аналитических запросов. Это также повышает точность моделей машинного обучения. По сути, преобразование данных действует как этап предварительной обработки, который позволяет в конечном итоге извлекать значимые идеи из необработанных данных.

**Обработка ошибок, ведение журнала и мониторинг.** Рано или поздно конвейеры данных сломаются. Обработка ошибок при приеме данных включает в себя прогнозирование, обнаружение и устранение ошибок, возникающих в процессе загрузки

данных. Это предотвращает попадание поврежденных данных в озеро данных и гарантирует, что процессы приема не будут нарушены. Некоторые примеры стратегий обработки ошибок включают реализацию логики повторных попыток для временных ошибок и настройку систем уведомлений для немедленных оповещений о сбоях. Одним из важнейших аспектов обработки ошибок является ведение журнала. Ведение журнала включает в себя запись событий, которые происходят во время процесса приема данных. Оно обеспечивает видимость производительности конвейеров данных и необходимо для диагностики проблем после их возникновения. Журналы должны включать такие сведения, как временная метка, источник данных, описание ошибки, как можно меньше и точнее, и часть процесса приема, где произошла ошибка. С точки зрения передовой практики, вы должны гарантировать, что все этапы процессов приема охватываются журналом, от точек ввода данных до окончательной загрузки. Вы также можете использовать инструменты управления журналами для агрегации, хранения и анализа данных журнала. Такие инструменты, как Elasticsearch, Logstash и Kibana, ELK Stack, могут предоставить мощные аналитические данные о журнале. Наконец, убедитесь, что журналы содержат достаточно информации, чтобы быть полезными, но не нарушают законы о конфиденциальности или правила защиты данных. Надежно храните журналы, чтобы предотвратить несанкционированный доступ. Непрерывный мониторинг процесса приема данных помогает обнаруживать и решать такие проблемы, как потеря данных, задержки или неточности, прежде чем они окажут большее влияние. Лучшей практикой здесь является внедрение инструментов мониторинга в реальном времени, которые могут предоставлять оповещения и панели мониторинга для визуализации работоспособности и производительности конвейеров приема данных. Этот проактивный подход помогает выявлять узкие места или сбои на ранних этапах процесса. Apache NiFi предлагает систему потока данных в реальном времени со встроенными панелями мониторинга для мониторинга пропускной способности данных, задержки и работоспособности системы, а Elastic Stack, включающий Elasticsearch, Logstash и Kibana, идеально подходит для регистрации процессов приема данных и визуализации потоков данных и показателей производительности. Prometheus и Grafana — это инструменты, которые обычно используются для мониторинга производительности систем приема данных путем извлечения метрик из конвейеров данных и отображения их на интуитивно понятных панелях мониторинга.

**Выводы.** Разница между классическим хранилищем данных и озером данных заключается в том, что хранилище данных предварительно классифицирует данные в точке входа, что приводит к потере времени и энергии, в то время как озеро данных принимает информацию с использованием нереляционной базы данных и не обрабатывает, и не классифицирует данные на входе. Это позволяет решить пробелу пропускной способности хранилища и проблему совместимости форматов входящих данных. В совокупности с возможностями проведения аналитики получаем удобный инструмент для предприятий с разнообразными данными и требованиями по их сбору, хранению и обработки.

Озеро данных может объединить данные о клиентах из платформы CRM с аналитикой социальных сетей, маркетинговой платформы, содержащей историю покупок, и инцидентов. Таким образом, компания сможет определить контингент клиентов, приносящих наибольшую прибыль, причины их оттока, а также подобрать акции или вознаграждения, которые повысят лояльность.

Озеро данных позволяет научно-исследовательским группам проверять гипотезы, уточнять предположения и оценивать результаты, например, выбирать правильные материалы при разработке продукта, что способствует повышению его производительности, проводить геномные исследования, в результате которых можно получить более эффективные лекарства, или понять готовность покупателей платить за различные атрибуты.

Интернет вещей (IoT) предоставляет больше возможностей для сбора данных о таких процессах, как производство, причем данные поступают в реальном времени от устройств, подключенных к Интернету. В озере данных удобно хранить и анализировать данные IoT, генерируемые машинами, и находить способы снижения операционных затрат и повышения качества работы.

### Список литературы

[1] Naresh Lokiny Comparative Study of Cloud Providers (AWS, Azure, Google Cloud) using Artificial Intelligence with DevOps. International Journal of Science and Research (IJSR) Volume 8 Issue 8, August 2019 p 2326 – 2329.DOI: g/10.21275/SR24724151213 2

[2] Raj, Anushree & D'Souza, Rio. (2019). A Review on Hadoop Eco System for Big Data. International Journal of Scientific Research in Computer Science, Engineering and Information Technology. 343-348. DOI:10.32628/CSEIT195172.

### Авторский вклад

**Алуев Евгений Александрович** – проведение исследования существующего рынка озер данных и их текущих реализаций.

## OVERCOMING BIG DATA CHALLENGES WITH DATA LAKE

***Eugene Alooeff***
*R&D Engineer ATEK,*
*Bachelor of Computer Science,*
*Bachelor.*

**Abstract.** The analysis of current problems of enterprises on storage and processing of unstructured data is made. The ways of solving these problems by means of using data lakes are analyzed, the features of their use are considered. Conclusions are made on the prospects of transition to using data lakes in some markets.

**Keywords**: Big Data, Data Lake, data processing, database.

УДК 004.06:004.08

# СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ БАЛАНСИРОВКИ ДАННЫХ ДЛЯ ЗАДАЧ МАШИННОГО ОБУЧЕНИЯ

***Е. Клицунова***
*Студентка кафедры информационных систем управления БГУ*
kateryna.klitsunova@gmail.com

***М.М. Лукашевич***
*Доцент кафедры информационных систем управления БГУ, ст. науч. сотр. лаборатории №222 ОИПИ, кандидат технических наук, доцент*
LukashevichMM@bsu.by

***Е. Клицунова***
*Является студенткой четвертого курса Белорусского государственного университета. Область научных интересов связана с решением практических задач в области машинного обучения.*

***М.М. Лукашевич***
*Является доцентом кафедры информационных систем управления БГУ. ст. науч. сотр. лаборатории №222 ОИПИ. Область научных интересов связана с разработкой методов и алгоритмов машинного обучения для решения задач анализа данных и компьютерного зрения.*

**Аннотация.** В работе представлены результаты сравнительного анализа методов и алгоритмов работы с несбалансированными данными при построении моделей машинного обучения. Представлены результаты экспериментальных исследований алгоритмов балансировки данных, основанных на увеличении меньшего класса и на уменьшении большего класса с использованием библиотеки imbalanced-learn. По результатам экспериментов оценено влияние изученных методов и алгоритмов на качество моделей, полученных в результате обучения с классическими классификаторами и ансамблевыми алгоритмами.

**Ключевые слова:** машинное обучение, несбалансированные данные, балансировка данных, oversampling, undersampling.

**Введение.** В актуальных задачах машинного обучения как во время исследований, так и на реальных практических примерах приходится сталкиваться с несбалансированным набором данных (Imbalanced Dataset). Набор данных называется несбалансированным, когда в нем наблюдается значительная диспропорция между количеством объектов различных классов. При этом соотношение классов может варьироваться от сравнительно небольшого смещения до значительных диспропорций, когда на один объект меньшего класса приходится сотня, тысяча или даже миллион объектов большего класса. Определение наличия дисбаланса классов в наборе данных можно осуществить визуально с помощью гистограммы, которая отражает количество объектов в каждом классе, или путем визуализации на графике при небольшой размерности пространства признаков [1]. Визуализация искусственно сгенерированного набора данных с соотношением объектов разных классов в пропорции 1:100 представлена на рисунке 1.

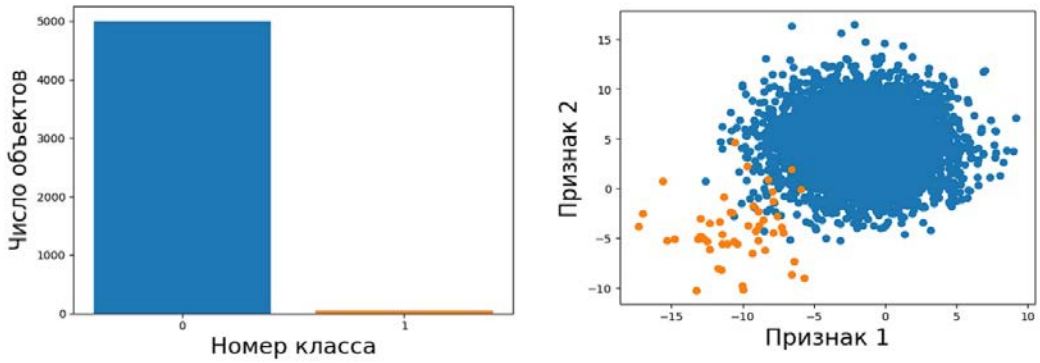



*Рисунок 1*. Визуализация набора данных с балансом классов 1:100 на гистограмме (слева) и графике (справа)

Наиболее характерна проблема дисбаланса классов для задач классификации. Она может привести к значительным трудностям при обучении моделей машинного обучения, поскольку классические алгоритмы классификации рассчитаны на равное соотношение между классами, и при работе с несбалансированными данными их эффективность значительно снижается [2]. Особенно важным это становится в задачах, где крайне важно получить несмещенные в сторону большего класса результаты — например, не пропустить пациента с серьезным заболеванием или мошеннические действия.

Для сравнительного анализа качества моделей в задачах классификации, как бинарной, так и мультиклассовой, используются стандартные метрики [3 - 5]. Однако некоторые из них применимы только для задачи со сбалансированным набором данных, но при использовании в задачах с несбалансированными данными могут привести к чрезмерно оптимистичным результатам и дать ложное представление об эффективности модели. К примеру, использование доли верно классифицированных объектов (Accuracy) для данных с выраженным дисбалансом может дать высокий уровень точности, приближающийся к 100%, даже если все объекты классифицируются в более крупный класс, а все объекты меньшего класса классифицированы неверно.

Согласно [6, 7] более подходящим является использование для задач классификации с несбалансированным набором данных сбалансированную точность (Balanced Accuracy), определяемую по формулам (1)-(3).

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specifity}}{2} \tag{1}$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Specifity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3}$$

где TP (True Positive) – количество корректных идентификаций положительного класса;

TN (True Negative) – количество корректных идентификаций отрицательного класса;

FN (False Negative) – количество случаев, когда объектам отрицательного класса были присвоены метки положительного класса;

FP (False Positive) – количество случаев, когда объектам отрицательного класса были присвоены метки положительного класса;

Целью исследования являлось экспериментальное сравнение качества некоторых моделей классификации до и после применения методов балансировки данных в зависимости от размера набора данных и величины дисбаланса классов. Необходимость исследования обусловлена тем, что каждый из методов имеет свои преимущества и недостатки, и их комбинации могут приводить к улучшению качества моделей.

**Методы балансировки данных.** Согласно [1, 8], методы и алгоритмы работы с несбалансированными данными можно разделить на следующие категории:

1 *Подходы на уровне алгоритма.* Данный тип методов заключается в модификации существующих классических алгоритмов таким образом, чтобы адаптировать их к несбалансированному набору данных и уменьшить смещение в сторону большего класса при обучении. Также сюда можно отнести обучение с учетом издержек классификации, что представляет собой модификацию алгоритмов путем добавления затрат на неправильную классификацию.

2 *Подходы на уровне данных.* Этот подход направлен на уменьшение дисбаланса непосредственно в наборе данных до применения к ним уже существующих алгоритмов классификации.

3 *Использование для обучения ансамблевых алгоритмов.* Такие алгоритмы лучше адаптируются к дисбалансу данных благодаря вычислению среднего результата нескольких классификаторов одного или разных типов.

4 *Комбинированные алгоритмы.* Приведенные выше методы можно комбинировать в разных соотношениях, чтобы получить модель с наилучшим качеством классификации. Также можно комбинировать различные инструменты из одной категории методов, к примеру, использовать разные варианты балансировки набора данных.

Рассмотрим основные методы подхода на уровне данных, а именно – балансировки данных.

*Увеличение меньшего класса.* Подход с увеличением количества объектов меньшего класса особенно актуален для работы с небольшими наборами данных, где удаление объектов большего класса может существенно сказаться на точности классификации. Выделяют следующие основные алгоритмы:

1 Алгоритм случайного увеличения (Random Oversampling) заключается в дублировании случайно выбранных объектов меньшего класса.

2 Алгоритм SMOTE создает синтетические примеры для меньшего класса, основываясь на существующими экземплярами, что увеличивает количество объектов редкого класса без создания копий существующего.

3 Алгоритм Borderline-SMOTE генерирует синтетические примеры только на границе классов, что улучшает способность модели различать классы, близкие к границе решения.

4 Алгоритм Borderline-SMOTE SVM использует векторные машины (SVM) для определения границ между классами и генерации синтетических данных на этих границах.

5 Алгоритм ADASYN создает синтетические примеры для меньшего класса с учетом их распределения и весов классов.

*Уменьшение большего класса.* Алгоритмы в рамках данного подхода можно разделить на три группы: ориентированные на выбор объектов, которые необходимо сохранить в выборке, ориентированные на выбор объектов для удаления и комбинированные методы.

1 Алгоритм случайного удаления (Random Undersampling) заключается в дублировании случайно выбранных объектов меньшего класса.

2 Алгоритм CNN направлен на поиск подмножества в выборке, который не приведет к потере качества модели.

3 Алгоритм Near Miss имеет три варианта (NearMiss-1, NearMiss-2 и NearMiss-3), в каждом из которых выбираются объекты большего класса, которые нужно сохранить в выборке, на основании метрики в пространстве признаков. В первом варианте остаются

объекты с минимальным средним расстоянием к трем ближайшим объектам меньшего класса, во втором варианте — к трем наиболее дальним объектам меньшего класса, в третьем варианте — к каждому объекту меньшего класса.

4 Алгоритм Tomek Links находит пары ближайших соседей из разных классов, после чего удаляется объект из большего класса.

5 Алгоритм ENN для каждого объекта вычисляет класс по трем ближайшим соседям, и если объект принадлежит большему классу и идентифицирован неправильно, то он удаляется.

6 Алгоритм OSS комбинирует алгоритмы Tomek Links для удаления шума и объектов большего класса на границе и CNN для удаления объектов, далеких от границы решения.

7 Алгоритм NCR комбинирует алгоритмы CNN для удаления избыточных объектов большего класса и ENN для удаления шума.

Методы увеличения меньшего класса следует применять с осторожностью во избежание переобучения модели. Также не обязательно по результатам увеличения меньшего класса получать именно столько объектов, сколько присутствует в преобладающем классе: для некоторых задач достаточно убрать дисбаланс до умеренного или небольшого. Уменьшение большего класса не во всех случаях приводит к значительному увеличению качества моделей (а иногда может привести и к ухудшению), однако его важно рассмотреть, поскольку алгоритмы из этой категории могут улучшить эффективность в случае комбинации их с методами увеличения меньшего класса.

**Постановка эксперимента.** Для проведения экспериментов были выбраны публичные наборы данных, представленные на платформе Kaggle. Подбор данных выполнялся по следующим критериям: различная величина наборов данных; различная величина дисбаланса; различная величина признакового пространства; различная тематика наборов данных. По данным критериям были отобраны пять различных наборов данных для следующих задач (таблица 1).

Таблица 1. Наборы данных для проведения экспериментов

| Название | Размер набора данных | Тип классификации | Дисбаланс |
|---|---|---|---|
| Классификация здоровья плода [9] | ~2 тыс. записей | мультиклассовая | умеренный |
| Прогнозирование церебрального инсульта [10] | ~43 тыс. записей | бинарная | чрезвычайный |
| Кредитный риск [11] | ~32 тыс. записей | бинарная | умеренный |
| Классификация кредитного рейтинга [12] | 100 тыс. записей | мультиклассовая | умеренный |
| Показатели здоровья при диабете [13] | ~254 000 записей | мультиклассовая | чрезвычайный |

Соотношение классов в выбранных наборах данных наглядно отображено на графиках, которые приводятся на рисунке 2.

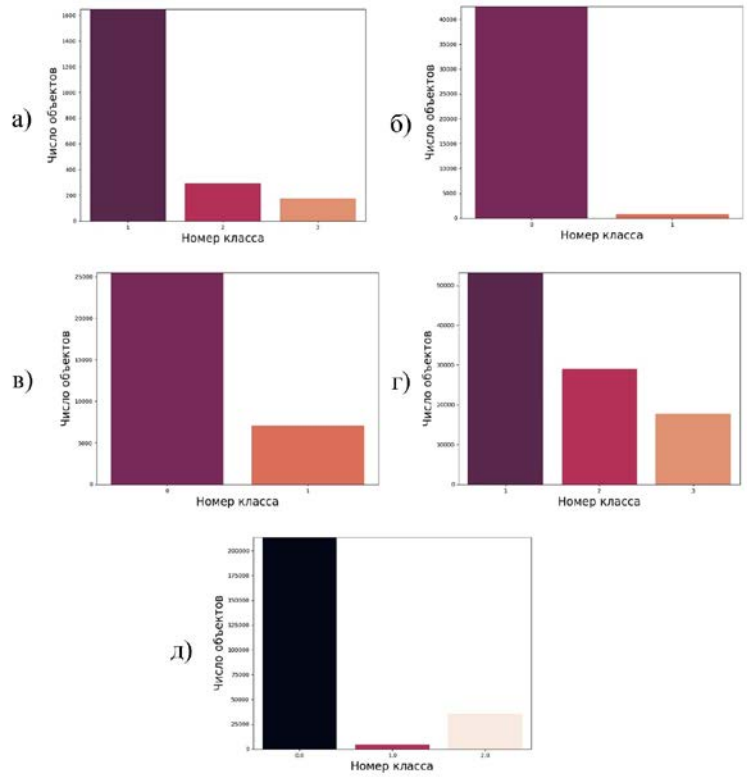



*Рисунок 2*. Соотношение классов в наборе данных: а) «Классификация здоровья плода» б) «Прогнозирование церебрального инсульта» в) «Кредитный риск» г) «Классификация кредитного рейтинга» д) «Показатели здоровья при диабете»

Для каждого из выбранных наборов данных выполнен разведочный анализ данных, замена пропусков средними значениями или их удаление, преобразование категориальных признаков в числовые и масштабирование данных. Выполнен отбор признаков с удалением на основе корреляционной матрицы и процента влияния на целевую переменную. Каждый набор был разделен на обучающую (70%) и тестовую (30%) выборки.

После выполнения предобработки на каждом выбранном наборе данных проводилось обучение модели с использованием следующих методов классификации:

- решающие деревья (Decision Trees, DT);
- метод k-ближайших соседей (k-Nearest Neighbors, kNN);
- Гауссовский наивный байесовский классификатор (Gaussian Naive Bayes, GaussianNB);
- многослойный перцептрон (Multilayer Perceptron, MLP).

Также было обучение проводилось на ансамблях с гиперпараметрами по умолчанию, а именно:

- случайный лес (Random Forest, RF);
- градиентный бустинг (Gradient Boosting, GB);
- AdaBoost;
- XGBoost;
- LightGBM.

В качестве языка программирования выбран Python благодаря наличию специализированной библиотеки imbalanced-learn [14], которая предоставляет алгоритмы обработки несбалансированных данных. Также использовалась библиотека алгоритмов машинного обучения scikit-learn, а также библиотеки XGBoost и LightGBM для реализации ансамблевых методов машинного обучения. Для всех методов классификации использовались значения параметров, установленные по умолчанию в указанных библиотеках. Для сравнения моделей в качестве метрики использовалась сбалансированная

точность (Balanced Accuracy). Каждая модель была обучена на обучающем наборе данных и выполнена оценка точности на тестовом наборе данных с применением указанной метрики (исходный набор). Далее для обучающего набора данных независимо друг от друга применялись указанные выше техники балансировки данных (увеличение меньшего класса и уменьшение большего класса). В рамках использования подхода по уменьшению большего класса эксперименты были проведены на значительном числе алгоритмов, но далее приведены результаты только некоторых более эффективных (для каждого набора данных перечень алгоритмов разный).

В таблицах 2-11 приводятся результаты экспериментов. «Полужирным» выделены максимальные значения точности при использовании исходного набора данных и различных техник балансировки данных. Голубой заливкой обозначены максимальные значения точности для набора в целом. Золотая заливка использовалась для акцента в тех случаях, когда наблюдалось снижение точности классификации по сравнению с точностью для исходного набора данных.

Таблица 2. Результаты экспериментов для набора данных «Классификация здоровья плода» (увеличение меньшего класса)

| **Модель** | **Исходный набор** | **Random Oversampling** | **SMOTE** | **B-SMOTE** | **B-SMOTE SVM** | **ADASYN** |
|---|---|---|---|---|---|---|
| kNN | 0,79 | 0,81 | 0,85 | 0,85 | 0,85 | 0,83 |
| DT | 0,87 | 0,88 | 0,86 | 0,86 | 0,86 | 0,87 |
| MLP | 0,67 | 0,69 | 0,85 | 0,80 | 0,85 | 0,82 |
| GaussianNB | 0,78 | 0,79 | 0,79 | 0,83 | 0,77 | 0,84 |
| RF | 0,89 | **0,93** | **0,93** | **0,93** | 0,91 | **0,92** |
| GB | 0,91 | 0,86 | 0,90 | 0,89 | 0,88 | 0,86 |
| AdaBoost | 0,82 | 0,91 | 0,92 | 0,90 | 0,91 | **0,92** |
| XGBoost | 0,89 | 0,92 | 0,92 | 0,92 | **0,92** | **0,92** |
| LightGBM | **0,92** | 0,92 | 0,92 | 0,92 | **0,92** | **0,92** |

Таблица 3. Результаты экспериментов для набора данных «Классификация здоровья плода» (уменьшение большего класса)

| **Модель** | **Исходный набор** | **Random Undersampling** | **NM-1** | **NM-3** | **CNN** | **Tomek Links** | **OSS** |
|---|---|---|---|---|---|---|---|
| kNN | 0,79 | 0,85 | 0,75 | 0,71 | 0,75 | 0,79 | 0,83 |
| DT | 0,87 | 0,87 | 0,85 | 0,83 | 0,80 | 0,84 | 0,87 |
| MLP | 0,67 | 0,82 | 0,77 | 0,76 | 0,77 | 0,67 | 0,84 |
| GaussianNB | 0,78 | 0,79 | 0,69 | 0,75 | 0,77 | 0,78 | 0,77 |
| RF | 0,89 | 0,91 | 0,90 | 0,88 | 0,87 | 0,88 | 0,91 |
| GB | 0,91 | **0,93** | **0,92** | **0,90** | **0,90** | 0,90 | **0,93** |
| AdaBoost | 0,82 | 0,78 | 0,70 | 0,84 | 0,79 | 0,81 | 0,88 |
| XGBoost | 0,89 | **0,93** | 0,91 | **0,90** | 0,87 | 0,89 | 0,94 |
| LightGBM | **0,92** | 0,92 | 0,88 | 0,87 | 0,86 | **0,93** | **0,93** |

Таблица 4. Результаты экспериментов для набора данных «Прогнозирование церебрального инсульта» (увеличение меньшего класса)

| Модель | Исходный набор | Random Oversampling | SMOTE | B-SMOTE | B-SMOTE SVM | ADASYN |
|---|---|---|---|---|---|---|
| kNN | 0,50 | 0,54 | 0,61 | 0,57 | 0,56 | 0,60 |
| DT | 0,51 | 0,52 | 0,55 | 0,54 | 0,54 | 0,54 |
| MLP | 0,50 | 0,73 | 0,65 | 0,66 | 0,56 | 0,64 |
| GaussianNB | **0,65** | 0,74 | **0,72** | **0,72** | **0,71** | **0,72** |
| RF | 0,50 | 0,51 | 0,53 | 0,52 | 0,53 | 0,53 |
| GB | 0,50 | 0,77 | 0,65 | 0,63 | 0,60 | 0,67 |
| AdaBoost | 0,50 | **0,78** | 0,67 | 0,67 | 0,62 | 0,67 |
| XGBoost | 0,50 | 0,60 | 0,56 | 0,54 | 0,53 | 0,57 |
| LightGBM | 0,50 | 0,66 | 0,57 | 0,55 | 0,53 | 0,56 |

Таблица 5. Результаты экспериментов для набора данных «Прогнозирование церебрального инсульта» (уменьшение большего класса)

| Модель | Исходный набор | Random Undersampling | NM-2 | NM-3 | Tomek Links | ENN | OSS | NCR |
|---|---|---|---|---|---|---|---|---|
| kNN | 0,50 | 0,74 | 0,51 | 0,65 | 0,50 | 0,50 | 0,50 | 0,50 |
| DT | 0,51 | 0,69 | 0,50 | 0,54 | 0,52 | 0,52 | 0,53 | 0,53 |
| MLP | 0,50 | 0,75 | **0,67** | 0,66 | 0,50 | 0,50 | 0,50 | 0,50 |
| GaussianNB | **0,65** | 0,74 | 0,52 | **0,71** | **0,65** | **0,65** | **0,59** | **0,66** |
| RF | 0,50 | 0,76 | 0,51 | 0,66 | 0,50 | 0,50 | 0,50 | 0,50 |
| GB | 0,50 | **0,77** | 0,50 | 0,70 | 0,50 | 0,50 | 0,50 | 0,50 |
| AdaBoost | 0,50 | **0,77** | 0,51 | 0,63 | 0,50 | 0,50 | 0,50 | 0,50 |
| XGBoost | 0,50 | 0,74 | 0,50 | 0,66 | 0,50 | 0,50 | 0,50 | 0,51 |
| LightGBM | 0,50 | 0,74 | 0,50 | 0,64 | 0,50 | 0,50 | 0,50 | 0,50 |

Таблица 6. Результаты экспериментов для набора данных «Кредитный риск» (увеличение меньшего класса)

| Модель | Исходный набор | Random Oversampling | SMOTE | B-SMOTE | B-SMOTE SVM | ADASYN |
|---|---|---|---|---|---|---|
| kNN | 0,72 | 0,74 | 0,73 | 0,74 | 0,74 | 0,73 |
| DT | 0,84 | 0,84 | 0,81 | 0,81 | 0,82 | 0,80 |
| MLP | 0,50 | 0,72 | 0,73 | 0,73 | 0,72 | 0,75 |
| GaussianNB | 0,67 | 0,72 | 0,69 | 0,68 | 0,71 | 0,68 |
| RF | 0,85 | 0,86 | 0,85 | 0,84 | 0,85 | 0,84 |
| GB | 0,84 | 0,84 | 0,83 | 0,84 | 0,84 | 0,84 |
| AdaBoost | 0,81 | 0,83 | 0,79 | 0,78 | 0,80 | 0,78 |
| XGBoost | **0,86** | **0,88** | **0,86** | **0,86** | **0,86** | **0,87** |
| LightGBM | **0,86** | 0,87 | **0,86** | **0,86** | **0,86** | 0,86 |

Таблица 7. Результаты экспериментов для набора данных «Кредитный риск» (уменьшение большего класса)

| **Модель** | **Исходный набор** | **Random Undersampling** | **NM-3** | **CNN** | **Tomek Links** | **ENN** | **OSS** | **NCR** |
|---|---|---|---|---|---|---|---|---|
| kNN | 0,72 | 0,74 | 0,70 | 0,73 | 0,73 | 0,75 | 0,72 | 0,74 |
| DT | 0,84 | 0,82 | 0,79 | 0,78 | 0,84 | 0,83 | 0,83 | 0,83 |
| MLP | 0,50 | 0,52 | 0,50 | 0,50 | 0,69 | 0,69 | 0,50 | 0,50 |
| GaussianNB | 0,67 | 0,71 | 0,71 | 0,71 | 0,68 | 0,70 | 0,69 | 0,70 |
| RF | 0,85 | 0,84 | 0,85 | 0,84 | 0,85 | 0,84 | 0,85 | 0,84 |
| GB | 0,84 | 0,84 | 0,84 | 0,84 | 0,84 | 0,84 | 0,84 | 0,84 |
| AdaBoost | 0,81 | 0,82 | 0,82 | 0,83 | 0,82 | 0,82 | 0,82 | 0,82 |
| XGBoost | **0,86** | **0,87** | **0,86** | **0,86** | **0,87** | **0,87** | **0,87** | **0,87** |
| LightGBM | **0,86** | 0,86 | **0,86** | **0,86** | 0,86 | 0,86 | 0,86 | 0,86 |

Таблица 8. Результаты экспериментов для набора данных «Классификация кредитного рейтинга» (увеличение меньшего класса)

| **Модель** | **Исходный набор** | **Random Oversampling** | **SMOTE** | **B-SMOTE** | **B-SMOTE SVM** | **ADASYN** |
|---|---|---|---|---|---|---|
| kNN | 0,71 | 0,75 | 0,76 | 0,76 | 0,75 | 0,77 |
| DT | 0,73 | 0,73 | 0,75 | 0,74 | 0,75 | 0,74 |
| MLP | 0,59 | 0,50 | 0,53 | 0,59 | 0,45 | 0,67 |
| GaussianNB | 0,64 | 0,63 | 0,63 | 0,63 | 0,63 | 0,63 |
| RF | **0,80** | **0,82** | **0,82** | **0,82** | **0,82** | **0,82** |
| GB | 0,71 | 0,72 | 0,73 | 0,73 | 0,73 | 0,73 |
| AdaBoost | 0,63 | 0,68 | 0,70 | 0,69 | 0,70 | 0,69 |
| XGBoost | 0,76 | 0,77 | 0,78 | 0,78 | 0,77 | 0,78 |
| LightGBM | 0,74 | 0,75 | 0,76 | 0,76 | 0,76 | 0,75 |

Таблица 9. Результаты экспериментов для набора данных «Классификация кредитного рейтинга» (уменьшение большего класса)

| **Модель** | **Исходный набор** | **Random Undersampling** | **NM-3** | **CNN** | **Tomek Links** | **ENN** | **OSS** | **NCR** |
|---|---|---|---|---|---|---|---|---|
| kNN | 0,71 | 0,68 | 0,58 | 0,56 | 0,71 | 0,67 | 0,64 | 0,71 |
| DT | 0,73 | 0,75 | 0,66 | 0,64 | 0,76 | 0,76 | 0,72 | 0,78 |
| MLP | 0,59 | 0,58 | 0,39 | 0,50 | 0,59 | 0,57 | 0,65 | 0,45 |
| GaussianNB | 0,64 | 0,63 | 0,67 | 0,65 | 0,64 | 0,64 | 0,64 | 0,64 |
| RF | **0,80** | **0,83** | **0,78** | **0,76** | **0,82** | **0,78** | **0,78** | **0,81** |
| GB | 0,71 | 0,73 | 0,71 | 0,70 | 0,72 | 0,72 | 0,71 | 0,73 |
| AdaBoost | 0,63 | 0,72 | 0,67 | 0,64 | 0,68 | 0,69 | 0,66 | 0,70 |
| XGBoost | 0,76 | 0,78 | 0,74 | 0,72 | 0,78 | 0,77 | 0,75 | 0,79 |
| LightGBM | 0,74 | 0,76 | 0,73 | 0,72 | 0,75 | 0,75 | 0,74 | 0,76 |

Таблица 10. Результаты экспериментов для набора данных «Показатели здоровья при диабете» (увеличение меньшего класса)

| **Модель** | **Исходный набор** | **Random Oversampling** | **SMOTE** | **B-SMOTE** | **B-SMOTE SVM** | **ADASYN** |
|---|---|---|---|---|---|---|
| kNN | 0,38 | 0,38 | 0,44 | 0,45 | 0,44 | 0,44 |
| DT | 0,40 | 0,39 | 0,40 | 0,40 | 0,39 | 0,40 |
| MLP | 0,39 | 0,43 | 0,40 | 0,44 | 0,47 | 0,42 |
| GaussianNB | **0,45** | **0,47** | **0,49** | **0,48** | **0,49** | **0,49** |
| RF | 0,39 | 0,38 | 0,39 | 0,39 | 0,39 | 0,39 |
| GB | 0,39 | 0,45 | 0,45 | 0,45 | 0,45 | 0,44 |
| AdaBoost | 0,39 | 0,45 | 0,46 | 0,47 | 0,46 | 0,46 |
| XGBoost | 0,39 | 0,44 | 0,40 | 0,40 | 0,40 | 0,40 |
| LightGBM | 0,38 | 0,45 | 0,41 | 0,40 | 0,41 | 0,40 |

Таблица 11. Результаты экспериментов для набора данных «Показатели здоровья при диабете» (уменьшение большего класса)

| **Модель** | **Исходный набор** | **Random Undersampling** | **CNN** | **Tomek Links** | **ENN** | **OSS** | **NCR** |
|---|---|---|---|---|---|---|---|
| kNN | 0,38 | 0,43 | 0,34 | 0,38 | 0,35 | 0,36 | 0,37 |
| DT | 0,40 | 0,40 | 0,36 | 0,40 | 0,39 | 0,39 | 0,39 |
| MLP | 0,39 | 0,43 | 0,39 | 0,38 | 0,37 | 0,38 | 0,37 |
| GaussianNB | **0,45** | **0,47** | **0,45** | **0,45** | **0,44** | **0,45** | **0,44** |
| RF | 0,39 | 0,41 | 0,38 | 0,38 | 0,36 | 0,38 | 0,37 |
| GB | 0,39 | 0,42 | 0,39 | 0,38 | 0,37 | 0,38 | 0,37 |
| AdaBoost | 0,39 | 0,41 | 0,39 | 0,38 | 0,38 | 0,38 | 0,37 |
| XGBoost | 0,39 | 0,42 | 0,38 | 0,38 | 0,38 | 0,38 | 0,37 |
| LightGBM | 0,38 | 0,42 | 0,37 | 0,38 | 0,38 | 0,38 | 0,37 |

**Обсуждение полученных результатов.** Сравнение полученных результатов позволяет сделать следующие выводы. В большинстве случаев классификаторы, обученные на сбалансированном наборе данных, показывали результаты не хуже, чем на исходном наборе. Во всех случаях удаётся достичь повышения точности классификации путём комбинации определенной модели классификатора и техники балансировки данных. Чем меньше степень дисбаланса, тем больше можно подобрать успешных комбинаций модели классификации и техники балансировки данных (таблица 12).

Таблица 12. Обобщение результатов экспериментальных исследований

| **Набор данных** | **Исходный набор** | | **Сбалансированный набор** | |
|---|---|---|---|---|
| | **Balanced Accuracy** | **Лучшая модель** | **Balanced Accuracy** | **Лучшая модель и техника балансировки** |
| «Классификация здоровья плода» | 0,92 | XGBoost<br>LightGBM | 0,93 | RF – Random Oversampling<br>RF – SMOTE<br>RF – B-SMOTE<br>GB – Random Undersampling |

| | | | | |
|---|---|---|---|---|
| | | | | GB – OSS<br>XGBoost – Random Oversampling<br>LightGBM – TomekLinks<br>LightGBM – OSS |
| «Прогнозирование церебрального инсульта» | 0,65 | GaussianNB | 0,78 | AdaBoost – Random Oversampling |
| «Кредитный риск» | 0,86 | XGBoost<br>LightGBM | 0,88 | XGBoost – Random Oversampling |
| «Классификация кредитного рейтинга» | 0,80 | RF | 0,83 | RF – Random Oversampling<br>RF – SMOTE<br>RF – B-SMOTE<br>RF – B-SMOTE SVM<br>RF – ADASYN<br>RF – Random Undersampling |
| «Показатели здоровья при диабете» | 0,45 | GaussianNB | 0,49 | GaussianNB – SMOTE<br>GaussianNB – B-SMOTE SVM<br>GaussianNB – ADASYN |

Наиболее противоречивым можно назвать алгоритм уменьшения большей выборки Near Miss. Наиболее чувствительным к нему оказался средний набор данных с крайне выраженным дисбалансом, причем для него вариант NearMiss-1 показал ухудшение сбалансированной точности вплоть до 0,24, а вот вариант NearMiss-3 улучшил сбалансированную точность в среднем на 0,13 — с 0,52 до 0,65. Для остальных же классификаторов и наборов данных применение этого метода дало результаты не лучше или даже существенно хуже, чем при обучении на исходном наборе.

Особенно хорошо показали себя методы балансировки данных на небольшом наборе данных и на наборе данных среднего размера, но с ярко выраженных дисбалансом классов. Для них удалось добиться значительного повышения качества моделей. А вот для больших наборов данных положительный эффект от балансирования класса оказался менее выраженным. При этом использование алгоритмов увеличения меньшего класса дало больший прирост сбалансированной точности, особенно для не-ансамблевых алгоритмов, чем использование алгоритмов уменьшения большего класса. Ансамблевые алгоритмы, как и ожидалось, оказались менее подверженными влиянию балансировки данных, однако и для них удалось добиться небольшого повышения точности. Тем не менее, для больших наборов данных со значительным дисбалансом результат классификации все равно является крайне низким — значение сбалансированной точности близко к 0,5, что не позволяет применять полученную модель в реальной практике.

**Заключение.** Сравнение результатов работы моделей на классических классификаторах показало улучшение качества моделей после балансировки данных. При этом методы увеличения меньшего класса дали более заметные результаты, а улучшение на наборах данных небольшого (до 10 тыс. записей) и среднего (до 50 тыс. записей) размера с большим дисбалансом в целом оказалось более выраженным. Также отмечено небольшое улучшение качества моделей при балансировке данных перед обучением с ансамблями такими как случайный лес, градиентный бустинг, XGBoost, AdaBoost и LightGBM. Наилучшие показатели сбалансированной точности были получены в результате комбинации балансирования данных и ансамблевых методов для всех наборов данных за исключением большого набора (250 тыс. записей) с чрезвычайным дисбалансом.

## Список литературы

[1] Brownlee J. Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning. Machine Learning Mastery, 2020. — 446 p.

[2] Kumar P., Bhatnagar R., Gaur K. Classification of imbalanced data: review of methods and applications. IOP conference series: materials science and engineering. 2021; 1099(1). DOI: 10.1088/1757-899X/1099/1/012077

[3] Михайличенко А.А. Аналитический обзор методов оценки качества алгоритмов классификации в задачах машинного обучения. Вестник Адыгейского государственного университета. 2022; 311(4). DOI: 10.53598/2410-3225-2022-4-311-52-59

[4] Ferri C., Hernendez-Orallo J., Modroiu R. An experimental comparison of performance measures for classification. Pattern Recognition Letters. 2009; 30(1). DOI: 10.1016/j.patrec.2008.08.010

[5] Sokolova M., Lapalme G. A systematic analysis of performance measures for classification tasks. Information processing & management. 2009; 45(4). DOI: 10.1016/j.ipm.2009.03.002

[6] Старовойтов В.В., Голуб Ю.И. Об оценке результатов классификации несбалансированных данных по матрице ошибок. Информатика. 2021; 18(1). DOI: 10.37661/1816-0301-2021-18-1-61-71

[7] Старовойтов В.В., Голуб Ю.И Как оценивать результаты классификации несбалансированных больших данных? BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня. Минск: Бестпринт. 2021 — с. 272–283.

[8] Krawczyk, B. Learning from imbalanced data: open challenges and future directions. Progress in artificial intelligence. 2016; 5(4). DOI: 10.1007/s13748-016-0094-0

[9] Ayres-de-Campos D., Bernardes, J., Garrido A., Marques-de-Sá J., Pereira-Leite L. A program for automated analysis of cardiotocograms. J. Matern. Fetal Med. 2000; 9. DOI: 10.1002/1520-6661(200009/10)9:5<311::AID-MFM12>3.0.CO;2-9

[10] Liu T., Fan W., Wu Ch. Data for A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical-datasets. Mendeley Data, 2019. 1. DOI: 10.17632/x8ygrw87jw.1

[11] Credit Risk Dataset [Электронный ресурс]. — Режим доступа: https://www.kaggle.com/datasets/laotse/credit-risk-dataset

[12] Dataset for Credit Score Classification [Электронный ресурс]. — Режим доступа: https://www.kaggle.com/datasets/ayushsharma0812/dataset-for-credit-score-classification

[13] Diabetes Health Indicators Dataset [Электронный ресурс]. — Режим доступа: https://www.kaggle.com/datasets/alexteboul/diabetes-health-indicators-dataset

[14] Lemaître G., Nogueira F., Aridas C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning; Journal of machine learning research. 2017; 18(17). DOI: 10.48550/arXiv.1609.06570

# COMPARATIVE ANALYSIS OF DATA BALANCING METHODS

***K. Klitsunova***
*Student of the Department of Information Management Systems of BSU*

***M.M. Lukashevich***
*Associate Professor of the Department of Information Management Systems of BSU, Senior Researcher of the Laboratory No. 222 of the UIIP PhD of Technical sciences, Associate Professor*

**Abstract.** The paper presents the results of a comparative analysis of methods and algorithms for working with imbalanced data in machine learning models. Experimental results of data balancing algorithms based on increasing a smaller class and decreasing a larger class using the imbalanced-learn library are presented. Based on the results of the experiments, the influence of the studied methods and algorithms on the quality of models obtained as a result of training with classical classifiers and ensemble algorithms was assessed.

**Keywords:** machine learning, imbalanced data, data balancing, oversampling, undersampling.

УДК 616.89-08:004.3:004.42:004.9

# ДЕКОМПОЗИЦИЯ ЦИФРОВЫХ СИГНАЛОВ ЭЛЕКТРОЭНЦЕФАЛОГРАММ ПРИ ИХ СРАВНИТЕЛЬНОМ АНАЛИЗЕ

***П.Ю. Бранцевич***
*Доцент кафедры программного обеспечения информационных технологий БГУИР, кандидат технических наук, доцент*
branc@bsuir.by

***П.Ю. Бранцевич***

*С 1985 года работал в области разработки архитектуры, методов, алгоритмов и программного обеспечения компьютерных систем и комплексов для определения метрологических характеристик виброизмерительных преобразователей и виброустановок, решения задач вибрационного контроля, мониторинга, диагностики и автоматики защиты сложных механизмов и агрегатов с вращательным движением. С 1995 по 2019 годы научный руководитель НИЛ «Систем вибродиагностики» БГУИР. Более сорока измерительно-вычислительных комплексов вибрационного контроля и мониторинга внедрены и введены в промышленную эксплуатацию на предприятиях энергетики Беларуси. Являлся научным руководителем и исполнителем четырех заданий Государственной научно-технической программы (ГНТП) «Энергетика», двух заданий ГНТП «Защита от чрезвычайных ситуаций», четырех заданий ГПНИ «Диагностика», более ста хозяйственных договоров. Автор трех монографий.*

**Аннотация.** Цифровая обработка электроэнцефалограмм (ЭЭГ), отражающих в определённом смысле состояние нейронной сети головного мозга человека, является современным методом определения их информативно-значимых параметров и характеристик. Рассмотрен способ представления ЭЭГ в виде суммы составляющих, выделенных в узких частотных полосах, которые по сути являются фрагментами модулированных гармоник. Приведены примеры сравнительного анализа ЭЭГ для оценки влияния электросудорожной терапии на состояние нейронной сети головного мозга человека.

**Ключевые слова:** сигнал, электроэнцефалограмма, параметр, характеристика, цифровая обработка.

**Введение.** Основным органом, управляющим поведением человека, является самообучающаяся нейронная сеть его головного мозга, в которой нейроны взаимодействуют посредством электрических токов и химических реакций. В таком аспекте человека можно представить, как органическую автоматическую систему принятия решений, в которую поступает информация о состоянии органов человека и об окружающей среде, а также имеются некоторые исполнительные органы, отражающие или исполняющие принимаемые решения.

Работа головного мозга вследствие протекающих в нем электрических токов идентифицируется изменением электромагнитного поля на поверхности головы, которое можно зафиксировать специальными первичными преобразователями и представить в виде

изменяющихся параметров тока или напряжения, что и происходит, когда снимают электроэнцефалограмму (ЭЭГ) [1−2]. ЭЭГ – это результат процедуры по определению параметров электрической активности головного мозга. ЭЭГ можно использовать для выявления симптомов нарушения работы головного мозга, оценки характера отклонений и степень их распространённости [3−4].

**Цифровая обработка сигналов электроэнцефалограмм.** Электроэнцефалограмма отражает колебания напряжения в результате ионного тока в нейронах головного мозга и является электрическим сигналом, как результатом спонтанной электрической активности мозга в течение определенного периода времени, полученным с нескольких электродов на поверхности головы или скальпа. Стандартной системой размещения электродов на поверхности головы, рекомендованной Международной федерацией электроэнцефалографии и клинической нейрофизиологии, является система «10−20%» [5].

Электроэнцефалограмма представляется в виде цифровых сигналы, которые можно обрабатывать и исследовать с помощью методов цифровой обработки сигналов (ЦОС) [6−8]. Целью исследования ЭЭГ является определение информативно значимых признаков, по изменению которых можно делать выводы о нарушении работы нейронной сети головного мозга человека, либо об эффективности применяемых методик лечения.

В настоящее время для оценки сигналов ЭЭГ анализируется мощность сигнала в частотных полосах, называемых: альфа (8-13 Гц); бета (13-35 Гц); гамма (35-70 Гц); дельта (0,3-4 Гц); тэта (4-8 Гц) [7].

Спектральный анализ показывает, что при изменении частотного разрешения изменяются параметры гармонических составляющих, представляющих сигнал ЭЭГ на разных временных интервалах, при этом заметно сохранение интенсивности сигнала в определённых частотных полосах [9−11].

Также представляют интерес результаты, получаемые при определении огибающих сигнала ЭЭГ для частотных полос и их амплитудные спектры [12].

Еще более наглядную информацию можно получить анализируя временные тренды среднеквадратического значения (СКЗ) сигнала ЭЭГ и СКЗ в частотных полосах [13].

Для более детального исследования ЭЭГ, выделения их информативно-значимых признаков, предлагается представление ЭЭГ в виде декомпозиции частотных составляющих с узкими частотными полосами. На рисунках 1-13 в качестве примера показаны исходная ЭЭГ и ее частотные оставляющие для отведения O1-AA частота дискретизации 250 Гц

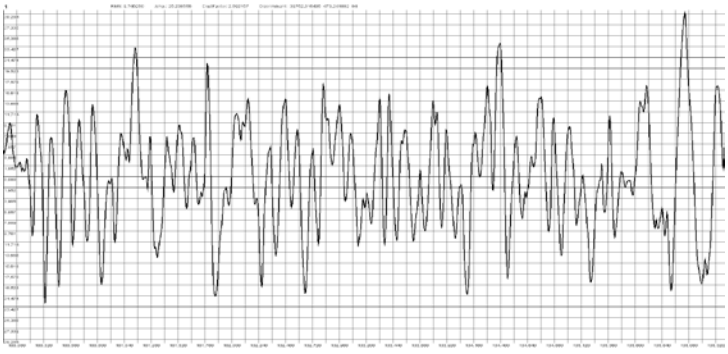


*Рисунок 1.* Сигнал ЭЭГ отведения O1-AA

*Рисунок 2.* Амплитудный спектр сигнал ЭЭГ отведения O1-AA, частотное разрешение 0,12207 *Гц*

*Рисунок 3.* Амплитудный спектр сигнал ЭЭГ отведения O1-AA, частотное разрешение 0, 061035 *Гц*

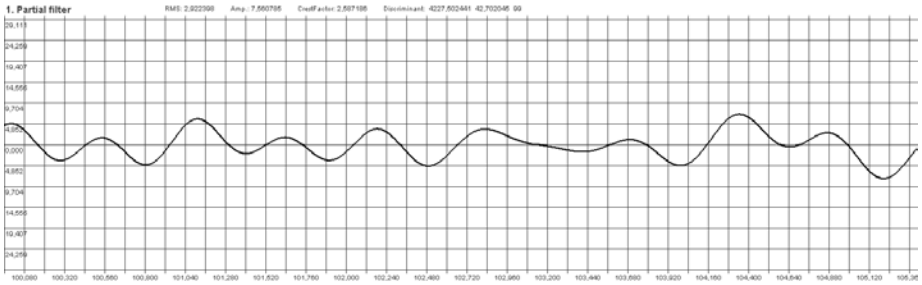



*Рисунок 4.* Составляющая сигнала ЭЭГ в частотной полосе 0,1-2 *Гц*

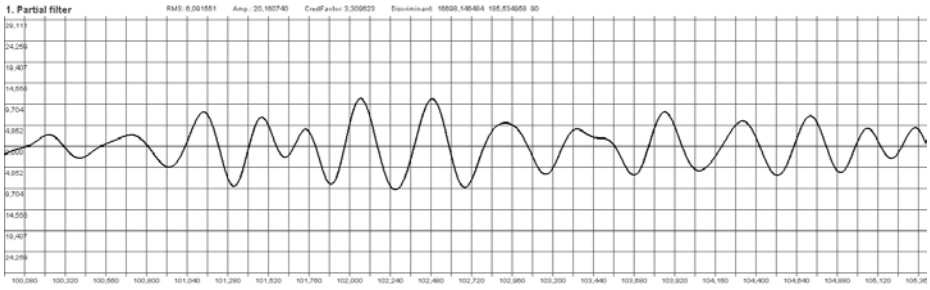



*Рисунок 5.* Составляющая сигнала ЭЭГ в частотной полосе 2-4 *Гц*

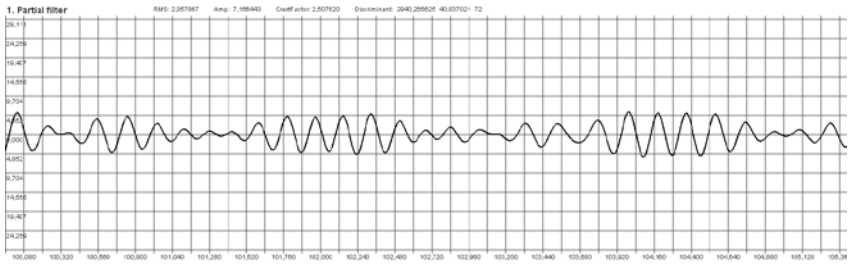



*Рисунок 6.* Составляющая сигнала ЭЭГ в частотной полосе 4-6 *Гц*

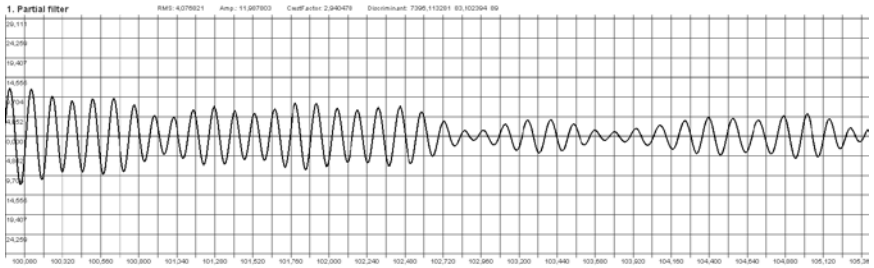



*Рисунок 7.* Составляющая сигнала ЭЭГ в частотной полосе 6-8 *Гц*

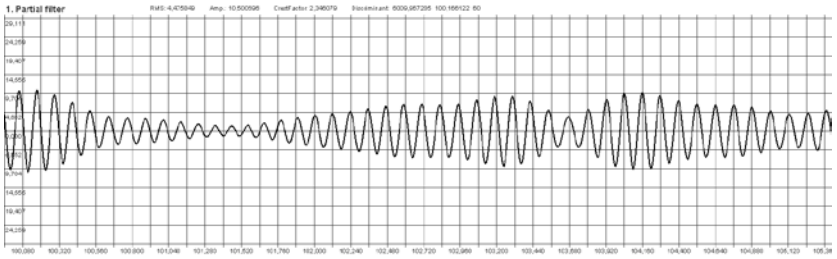



*Рисунок 8.* Составляющая сигнала ЭЭГ в частотной полосе 8-10 *Гц*

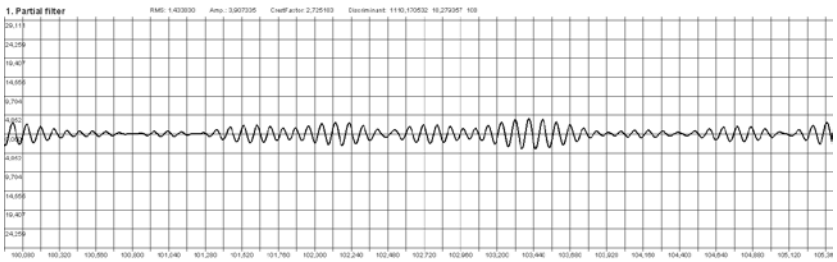



*Рисунок 9.* Составляющая сигнала ЭЭГ в частотной полосе 10-12 *Гц*

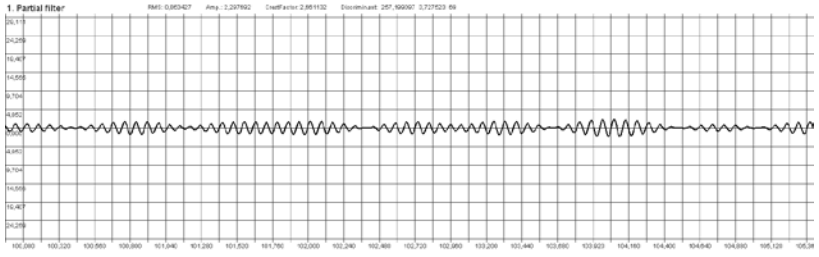



*Рисунок 10.* Составляющая сигнала ЭЭГ в частотной полосе 12-14 *Гц*

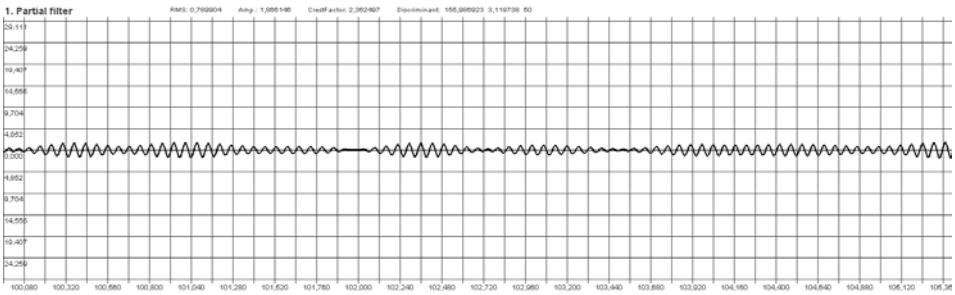



*Рисунок 11.* Составляющая сигнала ЭЭГ в частотной полосе 14-16 *Гц*

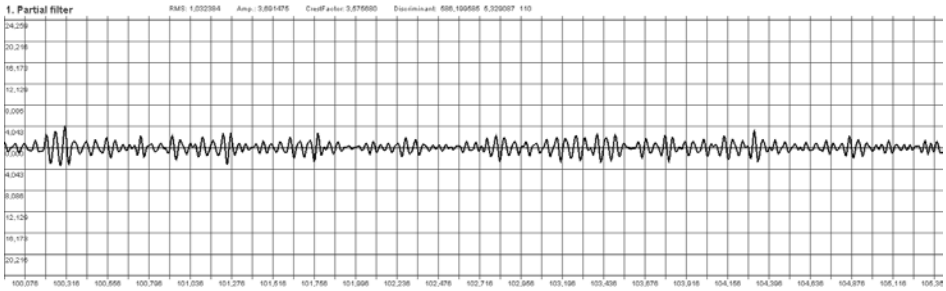



*Рисунок 12.* Составляющая сигнала ЭЭГ в частотной полосе 16-35 *Гц*

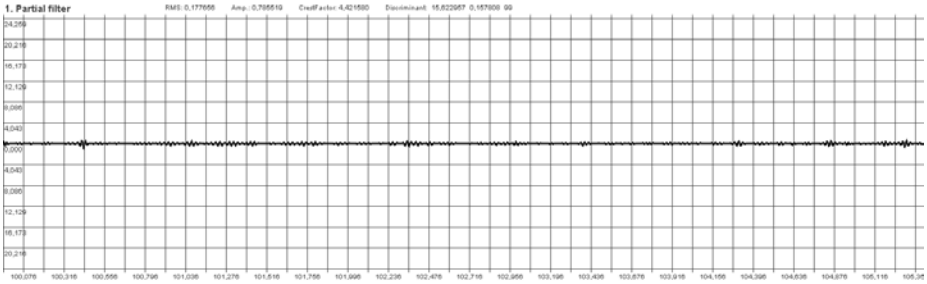



*Рисунок 13.* Составляющая сигнала ЭЭГ в частотной полосе 35-70 *Гц*

Можно заметить, что в узких частотных полосах составляющие ЭЭГ представляют собой фрагменты модулированных гармоник. Также можно предположить, что параметры этих составляющих ЭЭГ являются информативно-значимыми параметрами при сравнении ЭЭГ или оценке их изменений.

**Сравнительный анализ ЭЭГ.** Одним из применяемых методов лечения тяжелых депрессий и эпилепсий является электросудорожная терапия (ЭСТ), при которой осуществляется пропускание электрического тока через головной мозг пациента с целью достижения лечебного эффекта [14].

Нейронная сеть головного мозга человека подвергается сильному воздействию, однако пока не существует объективных информативно-значимых признаков на основе анализа ЭЭГ, которые показывали бы эффективность проведенных процедур ЭСТ.

Предложенный подход разложения на компоненты сигнала ЭЭГ использован при сравнительном анализе ЭЭГ, полученных до и после процедуры ЭСТ. В таблицах 1-3 представлены вычисленные параметры исследованных ЭЭГ.

Таблица 1. Параметры составляющих ЭЭГ до и после процедуры ЭСТ. Эксперимент 1

| Частотный диапазон, Гц | СКЗ, относительные единицы | | Пик фактор | |
|---|---|---|---|---|
| | До процедуры | После | До процедуры | После |
| 2−4 | 0,398 | 1,067 | 2,578 | 3,031 |
| 4−6 | 0,373 | 1,19 | 3.244 | 4,016 |

| | | | | |
|---|---|---|---|---|
| 6−8 | 0,404 | 1,07 | 3,047 | 3,329 |
| 8−10 | 1 | 1 | 3,765 | 2,582 |
| 10−12 | 1,218 | 0.763 | 2,087 | 2,48 |
| 12−14 | 0,383 | 0,315 | 2,614 | 2,497 |
| 14−16 | 0,244 | 0,246 | 2,588 | 2,68 |

Таблица 2. Параметры составляющих ЭЭГ до и после процедуры ЭСТ. Эксперимент 2

| Частотный диапазон, Гц | СКЗ, относительные единицы | | Пик фактор | |
|---|---|---|---|---|
| | До процедуры | После | До процедуры | После |
| 2−4 | 1,262 | 0,54 | 2,634 | 2,62 |
| 4−6 | 1,167 | 0,583 | 2,54 | 2,92 |
| 6−8 | 0,99 | 0,537 | 2,27 | 2,44 |
| 8−10 | 1 | 1 | 2,8 | 3,38 |
| 10−12 | 2,55 | 1,628 | 2,75 | 3,83 |
| 12−14 | 0,935 | 0,583 | 2,54 | 3,04 |
| 14−16 | 1,209 | 0,613 | 3,12 | 3,19 |

Таблица 3. Параметры составляющих ЭЭГ до и после процедуры ЭСТ. Эксперимент 3

| Частотный диапазон, Гц | СКЗ, относительные единицы | | Пик фактор | |
|---|---|---|---|---|
| | До процедуры | После | До процедуры | После |
| 2−4 | 0,331 | 0,453 | 2,432 | 2,625 |
| 4−6 | 0,303 | 0,443 | 2,918 | 2,863 |
| 6−8 | 0,323 | 0,366 | 3,245 | 2,553 |
| 8−10 | 1 | 1 | 2,84 | 2,553 |
| 10−12 | 0,363 | 0,42 | 2,69 | 2,897 |
| 12−14 | 0,297 | 0,234 | 3,401 | 3,806 |
| 14−16 | 0,176 | 0,202 | 2,545 | 3,075 |

Данные, представленные в таблицах 1 и 2 получены для случаев, когда процедуры ЭСТ дали положительный результат, а в таблице 3, когда положительный лечебный эффект не был достигнут.

Можно заметить, что при положительном результате от процедуры ЭСТ наблюдаются существенные изменения относительных величин СКЗ частотных составляющих, а при отрицательном результате эти изменения незначительны. Однако зафиксированные изменения относительных величин частотных составляющих ЭЭГ не имеют однозначного характера, поэтому требуется проведение большего числа экспериментов для получения обоснованных заключений.

**Заключение.** Применение электроэнцефалографии в медицинской практике позволяет получить различные новые данные о работе нейронной сети головного мозга человека, оценить эффективность применяемых методик лечения. Вычислительные мощности современных малогабаритных компьютеров, смартфонов, встроенных систем позволяют реализовать различные трудоёмкие алгоритмы цифровой обработки ЭЭГ для определения их информативно значимых параметров и характеристик. В данной работе представлены результаты исследований, проводимых с целью получения новых информативно-значимых параметров для сравнительного анализа ЭЭГ. Следует отметить, что полученные результаты требует подтверждения на большом объеме реальных данных.

## Список литературы

[1] Электроэнцефалография: руководство / М.В. Александров, Л.Б. Иванов, С.А. Лытаев [и др.]. Под ред. М.В. Александрова. 3-е изд., перераб. и доп. – Санкт-Петербург: СпецЛит, 2020. — 224 с.

[2] Зенков, Л.Р. Клиническая электроэнцефалография (с элементами эпилептологии). - 2-е изд., испр. и доп. - М.: МЕДпресс-информ, 2002. - 368 с.

[3] Докукина Т.В, Мисюк Н.Н. Визуальная и компьютерная ЭЭГ в клинической практике. - Минск: Кнігазбор, 2011. - 187 с.

[4] Павлова Л.П. Доминанты деятельного мозга человека. Системный психофизиологический подход к анализу ЭЭГ. СПб.: Информ-навигатор; 2017. - 430 с.

[5] Татум У.О., Хусейн А.М., Бенбадис С.Р., Каплан П.В. Клиническая интерпретация электроэнцефалографии. Пер. с англ. - М.: Издательский дом БИНОМ. 2020. - 264 с

[6] Лайонс Р. Цифровая обработка сигналов. – М.: ООО «Бином-Пресс», 2006. – 656 с.

[7] Кулаичев А.П. Компьютерная электрофизиология и функциональная диагностика. Москва: ИНФРА-М; 2018. – 469 с.

[8] Бранцевич П.Ю. Цифровая обработка вибрационных сигналов. – Минск: Бестпринт, 2022. – 297 с.

[9] Бранцевич П.Ю. Сравнительный анализ электроэнцефалограмм. BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics: сб. науч. ст. IX Междунар. науч.-практ. конф. В 2 ч. Ч. 1 (Республика Беларусь, Минск, 17–18 мая 2023 года). Редкол.: В. А. Богуш [и др.]. – Минск: БГУИР, 2023. – с. 132–144.

[10] Бранцевич П.Ю. Информативные параметры электроэнцефалограмм. Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния: материалы седьмой Междунар. науч.-практ. конф. 18-19 мая 2031 г., Минск, М-во образования Респ. Беларусь, НИУ «Ин-т приклад. физ. проблем им. А. Н. Севченко» Беларус. Гос. ун-та; редкол.: Ю.И. Дудчик (гл. ред.), И.М. Цикман, И.Н. Кольчевская. − Минск: ОДО «Рейплац», 2023. – с. 137–139.

[11] Бранцевич, П.Ю. Цифровая обработка электроэнцефалограмм / П.Ю. Бранцевич // Science and innovation. – Ташкент: LLC «Science and innovation», 2023. – Special Issue 3. – с. 930–934.

[12] Бранцевич, П.Ю. Цифровая обработка сигналов и сравнительный анализ электроэнцефалограмм / П.Ю. Бранцевич // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. X Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13 марта 2024 года ). В 2 ч. Ч. 1 / редкол. : В. А. Богуш [и др.]. – Минск: БГУИР, 2024. – С. 101–112.

[13] Бранцевич П.Ю. Докукина Т.В, Мисюк Н.Н., Королевич П.П. Информативные признаки и характеристики электроэнцефалограмм / Медэлектроника – 2024. Средства медицинской электроники и новые медицинские технологии : сб. науч. ст. XIV Междунар. науч.- техн. конф. (Республика Беларусь, Минск, 5-6 декабря 2024 года). – Минск : БГУИР, 2024. – С. 63-69.

[14] Метод лечения резистентных форм психических и поведенческих расстройств с использованием электросудорожной терапии. Инструкция по применению / Докукина Т.В. [и др.]. Утв. М-вом здравоохранения Респ. Беларусь 20.01.2015. – Минск: Министерство здравоохранения РБ, 2015. – 11 с.

## Авторский вклад

**Бранцевич Петр Юльянович** – цифровая обработка сигналов и сравнительный анализ электроэнцефалограмм, формирование статьи

# DECOMPOSITION OF DIGITAL SIGNALS OF ELECTROENCEPHALOGRAMS IN THEIR COMPARATIVE ANALYSIS

***P.J. Brancevich***
*Grand PhD courses of BSUIR,*
*PhD of Technical Sciences,*
*Associate Professor*

**Abstract.** Digital processing of electroencephalograms (EEG), reflecting in a certain sense the state of the neural network of the human brain, is a modern method for determining their informative and significant parameters and characteristics. The article considers a method for representing EEG as a sum of components isolated in narrow frequency bands, which are essentially fragments of modulated harmonics. Examples of comparative EEG analysis are given to assess the effect of electroconvulsive therapy on the state of the neural network of the human brain.

**Keywords:** signal, electroencephalogram, parameter, characteristic, digital processing

УДК 004.75

# STATISTICAL METHOD AND GRAPH THEORY-BASED ALGORITHM FOR DETECTING NETWORK TRAFFIC

***M.B. Bekiyeva***
*Senior teacher of Applied Mathematics and Informatics Department, Oguz han Engineering and Technology University of Turkmenistan, Candidate of Physical and Mathematical Sciences*
successbmb@gmail.com

***G.O. Orazdurdyyeva***
*Instructor of Computer Sciences and Information Technologies Department, Oguz han Engineering and Technology University of Turkmenistan*
gulshatorazdurdyyewa3@gmail.com

***M.B. Bekiyeva***
*I graduated from the International Turkmen-Turkish University. My research interests include the development of mathematical modeling, the design of algorithms for solving problems related to the finite element method, as well as the organization of educational and research processes in a technical university.*

***G.O. Orazdurdyyeva***
*I graduated from the Oguz han Engineering and Technology University of Turkmenistan. My research interests are related to cybersecurity, mathematical modeling, and network traffic analysis.*

**Abstract.** With the continuous evolution of cyber threats, robust mechanisms for network traffic analysis and anomaly detection are crucial for maintaining cybersecurity. This paper introduces a novel approach that fuses statistical modeling with graph theory to uncover irregular network behaviors. By conceptualizing network interactions as graph structures and applying advanced statistical analysis, the algorithm detects deviations from typical traffic patterns, significantly boosting the detection of both conventional and novel threats.

**Keywords:** Network anomaly detection, Graph theory, Statistical modeling, Network security, Traffic analysis, Cybersecurity, Kernel density estimation, Z-score analysis.

**Introduction.** The ever-growing complexity of modern networks has led to a surge in sophisticated cyber-attacks, underscoring the necessity for innovative detection techniques. Traditional monitoring tools often struggle to adapt to evolving threat landscapes, leaving critical vulnerabilities exposed. This paper proposes a hybrid approach that leverages statistical analysis and graph theory to identify anomalies within network traffic. By modeling data flow as graph structures and applying statistical methods, the algorithm enhances the identification of both subtle and large-scale deviations, offering a proactive defense mechanism against cyber threats.

**Methodology.**

**1 Traffic Data Acquisition and Preprocessing**: Traffic data acquisition and preprocessing are critical steps in building an effective anomaly detection framework. Network traffic data is collected from multiple sources, such as packet capture tools (e.g., Wireshark, tcpdump), flow-based systems (e.g., NetFlow, sFlow), and network intrusion detection systems (e.g., Snort, Suricata). The raw traffic data typically includes vast amounts of redundant, irrelevant, or noisy information, which can hinder analysis. To address this, preprocessing involves several stages: packet deduplication to remove repetitive data, protocol parsing to extract relevant fields, and filtering to exclude irrelevant traffic (e.g., broadcast and multicast packets). Noise reduction

techniques, such as applying statistical filters or traffic normalization, ensure the dataset accurately reflects normal and anomalous network behaviors. Additionally, session reassembly and feature extraction are performed to convert raw traffic into structured data suitable for analysis, including metrics like packet size, inter-arrival time, and byte counts. These preprocessing steps ensure the data is clean, consistent, and ready for use in advanced anomaly detection algorithms.

**2 Statistical Modeling**: Advanced statistical methods, including Kernel Density Estimation (KDE) and Z-score analysis, are leveraged to scrutinize traffic patterns. KDE provides a smooth estimate of data distribution, while Z-score analysis pinpoints outliers that diverge significantly from standard traffic flows.

You are monitoring network traffic to detect anomalies in packet inter-arrival times (the time between successive packets reaching a server). A typical pattern shows that inter-arrival times follow a normal distribution with a mean (μ – mu) of 2 milliseconds (ms) and a standard deviation (σ – sigma) of 0.5 ms.

Step-by-step Z-score analysis:

**2.1 Data collection**: Monitor network traffic over a period and record the inter-arrival times of packets. Assume the observed data includes these values (in milliseconds): 2.1, 1.9, 2.3, 1.8, 2.0, 5.0, 1.7, 2.2, 2.0, 0.5, 2.1, 1.9, 2.3, 1.8, 2.0, 5.0, 1.7, 2.2, 2.0, 0.5.

**2.2 Calculate the Z-score for each data point**: The Z-score is calculated using the following formula:

$$Z\frac{(x-\mu)}{\sigma},$$

where:

$x$ – observed inter-arrival time;

$\mu = 2.0$ – mean inter-arrival time;

$\sigma = 0.5$ – standard deviation of inter-arrival times.

Calculate Z-scores for the dataset:

For $x = 2.1$:

$$Z\frac{(2.1-2.0)}{0.5} = 0.2;$$

For $x = 5.0$:

$$Z\frac{(5.0-2.0)}{0.5} = 6.0;$$

This calculation is applied to all data points in the dataset.

**Z-scores**: 0.2, − 0.2, 0.6, − 0.4, 0.0, 6.0, − 0.6, 0.4, 0.0, − 3.0, 0.2, – 0.2, 0.6, –0.4, 0.0, 6.0, – 0.6, 0.4, 0.0, – 3.0.

**2.3 Identify anomalies**: Set a threshold for anomalies. Commonly, data points with $|Z| > 3|Z| > 3$ are considered outliers.

In this practice, $Z = 6.0$ (corresponding to $x = 5.0$) and $Z = -3.0$ (corresponding to $x = 0.5$) are anomalies.

**2.4 Interpretation**:

$x = 5.0$: An unusually long inter-arrival time could indicate a delay caused by network congestion, server overload, or a Denial of Service (DoS) attack.

$x = 0.5$: A very short inter-arrival time might signify bursty traffic or a scanning attack.

**2.5 Action**:

Investigate the causes of these anomalies.

Mitigate issues by implementing rate-limiting, blocking malicious IPs, or optimizing server resources, Fig. 1.

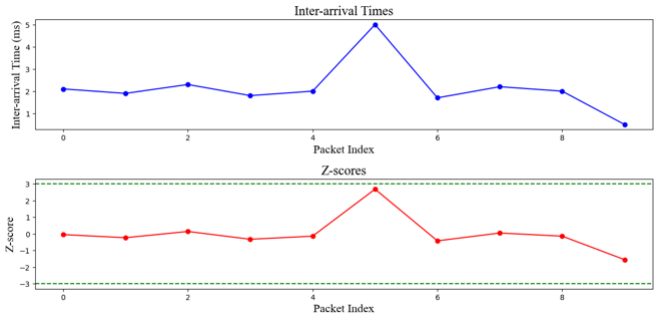



*Figure 1*. Monitoring network traffic to detect anomalies in packet inter-arrival times

**3 Graph construction and metrics**: The network is translated into a graph framework, where nodes symbolize endpoints and edges depict communication links. Metrics such as edge betweenness, eigenvector centrality, and clustering coefficients are employed to detect anomalies in network topology. In a typical network, nodes represent endpoints such as computers, servers, or routers, and edges represent the communication links between them. By analyzing the network as a graph, we can detect anomalous behavior that might indicate security issues like malicious activities, unauthorized data access, or network misconfigurations.

**Network graph construction:** Consider a simple network with 5 nodes and 6 communication links:

– nodes: A, B, C, D, E;
– edges: (A-B), (A-C), (B-C), (B-D), (C-E), (D-E).

This network can be visualized as a graph where nodes are the devices, and edges are the communication links.

**Graph Metrics:**

*1 Edge betweenness:* This metric measures the number of shortest paths that pass-through a given edge. Anomalous edges might appear in a graph if they have unusually high betweenness, suggesting critical communication paths that could be targeted by an attacker.

*2 Eigenvector centrality:* This metric identifies the most influential nodes in the network. Nodes with high eigenvector centrality are often crucial for communication across the network. A node with unexpectedly high centrality might be a point of interest for detecting unauthorized access.

*3 Clustering coefficients:* This measures the tendency of nodes to cluster together. Anomalies in the clustering coefficient could indicate abnormal groupings of network entities, such as an unexpected spike in communication between previously isolated nodes, which could signify a security breach or misconfiguration.

Practice:

Assume the following hypothetical scenario:

– node A is a server that should only communicate with nodes B and C;
– node B is a typical endpoint that communicates with A, C, and D;
– node C is a regular endpoint, connected to A, B, and E;
– node D is an endpoint that should only communicate with B and E;
– node E is an endpoint that communicates only with C and D.

Now imagine a scenario where an attack is happening. Node B starts communicating with E, which is unusual because B and E should not normally communicate.

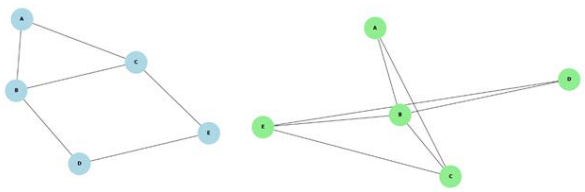


*Figure 2*. Network graph construction

***4 Anomaly detection framework**:* By amalgamating statistical insights with graph metrics, the algorithm delivers a comprehensive assessment of network anomalies. Machine learning classifiers further enhance detection precision by refining the model through iterative learning.

***5 Experimental setup and results*** to validate the proposed algorithm, extensive experiments were conducted using real-world network traffic datasets. The testing environment consisted of a controlled network laboratory equipped with simulation tools designed to mimic diverse attack scenarios, including Distributed Denial of Service (DDoS), port scanning, and data exfiltration attempts. The dataset was partitioned into training and testing subsets, ensuring an unbiased evaluation of the algorithm's performance.

**Results.**

1 The algorithm demonstrated a 92% accuracy in detecting network anomalies, surpassing traditional rule-based detection systems by 15%.

2 The false positive rate was reduced to 4.5%, indicating enhanced precision in distinguishing between normal and malicious traffic.

3 Graph-based metrics, particularly eigenvector centrality, played a pivotal role in identifying previously undetected anomalies linked to lateral movement within the network.

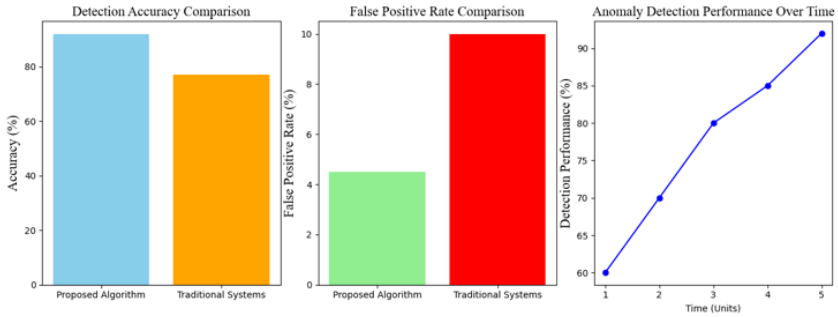



*Figure 3.* Anomaly detection performance over time

Despite its promising performance, the algorithm's computational overhead during large-scale deployments remains a challenge. Future work will focus on optimizing graph construction processes and exploring distributed computing techniques to enhance scalability. Additionally, further research will be conducted to refine machine learning classifiers, ensuring improved detection accuracy across diverse network environments.

**Conclusions.** As cyber threats continue to evolve in sophistication and scale, traditional methods of network traffic analysis often fall short in detecting novel and complex attack patterns. This paper introduces a novel approach to network anomaly detection that integrates statistical methodologies with graph theory, providing a robust solution to this challenge. By combining advanced statistical techniques such as Kernel Density Estimation (KDE) and Z-score analysis with graph-based metrics like eigenvector centrality and edge betweenness, the proposed algorithm is able to identify subtle deviations from typical traffic behavior. These deviations, which often signal cyber threats, can be detected at an early stage, allowing for faster and more accurate responses. The algorithm constructs a comprehensive view of network traffic by treating network interactions as graph structures, where nodes represent network entities and edges represent communication links. Statistical insights are then applied to measure the normality of traffic patterns, while graph metrics are employed to analyze the topology of the network and identify anomalous behaviors, such as lateral movements or unexpected connections. Experimental results, using real-world network traffic datasets and simulated attack scenarios, validate the efficacy of the proposed method. The algorithm demonstrates superior performance in detecting various types of cyber threats, including Distributed Denial of Service (DDoS) attacks, data exfiltration attempts, and internal reconnaissance activities. Performance metrics such as detection accuracy, precision, recall, and Z-score reveal that this approach significantly outperforms traditional anomaly detection systems. Furthermore, it reduces false positives and improves detection rates for both known and unknown attacks.

The promising results of this research suggest that the integration of statistical analysis and graph theory offers a highly effective framework for enhancing network security. This hybrid approach can serve as a valuable addition to existing security tools, providing network administrators with a powerful mechanism for identifying and mitigating complex cyber threats in real-time.

## References

[1] Smith, J.: Advanced Network Security Techniques.Cybersecurity Journal, 45(3), 112-125, 2023.

[2] Lee, K., & Park, H.: Graph Theory Applications in Network Intrusion Detection. IEEE Transactions on Network Security, 59(6), 847-860, 2022.

[3] Johnson, M.: Statistical Methods for Anomaly Detection in Network Traffic. Journal of Information Security, 38(2), 203-218, 2021.

[4] Brown, T.: Machine Learning Approaches for Cybersecurity. ACM Computing Surveys, 55(4), 1-36, 2023.

[5] Wang, Y.: Graph-Based Anomaly Detection in Large Networks. Journal of Network and Systems Management, 30(2), 329-350, 2022.

[6] Nguyen, D.: Cyber Threat Intelligence and Data Analysis. Journal of Cyber Research, 29(5), 455-470, 2021.

[7] Patel, R.: Network Forensics: Techniques and Applications. IEEE Security & Privacy, 21(1), 58-67, 2023.

[8] Kim, H.; Real-Time Detection of Network Intrusions using Graph Metrics. Computer Communications, 190, 101-113, 2022.

[9] Zhou, L.: Evaluating Anomaly Detection in IoT Networks. Journal of Internet Technology, 24(3), 389-401, 2023.

[10] Clark, E.: Advanced Persistent Threat Detection Using Graph Analysis. IEEE Transactions on Information Forensics, 18(7), 1347-1360, 2022.

[11] Bekiyeva, M.B.: Modeling network traffic dynamics under DDoS attacks using differential equations, https://edaconf.bsuir.by/m/12_136156_1_190888.pdf, 2024.

[12] Orazdurdyyeva, G.O.: Simulating network conditions and DDoS attack scenarios using NS-3 technology, https://edaconf.bsuir.by/m/12_136156_1_190888.pdf, 2024.

## Author contribution

**Bekiyeva Maral Batyrowna** – Development of a statistical method for detecting network traffic, formulation of conclusions based on statistical methods, preparation for publication and formatting the work.

**Orazdurdyyeva Gulshat Orazmuhammedowna** – Development of an algorithm based on graph theory, data analysis and conducting experiments, writing and editing the paper.

# СТАТИСТИЧЕСКИЙ МЕТОД И АЛГОРИТМ НА ОСНОВЕ ТЕОРИИ ГРАФОВ ДЛЯ ОБНАРУЖЕНИЯ СЕТЕВОГО ТРАФИКА

***М.Б. Бекиева***
*Старший преподаватель кафедры Прикладной математики и информатики Инженерно-технологического университета Туркменистана имени Огуз хана, кандидат физико-математических наук*

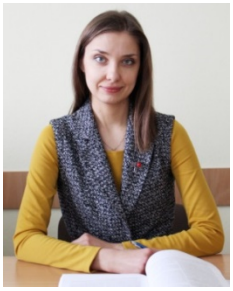
***Г.О. Ораздурдыева***
*Преподаватель кафедры Компьютерных наук и информационных технологий Инженерно-технологического университета Туркменистана имени Огуз хана*

**Аннотация.** С постоянной эволюцией киберугроз надежные механизмы анализа сетевого трафика и обнаружения аномалий играют ключевую роль в обеспечении кибербезопасности. В данной статье представлен новый подход, объединяющий статистическое моделирование и теорию графов для выявления нестандартного поведения в сети. Представляя сетевые взаимодействия в виде графовых структур и применяя продвинутый статистический анализ, разработанный алгоритм выявляет отклонения от типичных шаблонов трафика, значительно повышая эффективность обнаружения как известных, так и новых угроз.

**Ключевые слова:** обнаружение аномалий в сети, теория графов, статистическое моделирование, сетевая безопасность, анализ трафика, кибербезопасность, оценка ядерной плотности, Z-оценка.

УДК 796.012:612.76]:514.742.2

# АЛГОРИТМ КОРРЕКЦИИ ТРАЕКТОРИИ ДВИЖЕНИЯ СПОРТСМЕНОВ НА БАЗЕ МАТРИЦ ПОВОРОТА



***Г.Г. Турантаева***
*Аспирант кафедры «Спортивная инженерия БНТУ»*
ga_et@inbox.ru

***Н.С. Давыдова***
*Доцент кафедры инфокоммуникационных технологий БГУИР, кандидат технических наук, доцент* davydova-ns@bsuir.by

***В.Е. Васюк***
*Ведущий специалист центра координации научно-методической и инновационной деятельности БГУФК, кандидат педагогических наук, доцент*

***К.И. Татарко***
*Доцент кафедры социальной и организационной психологии БГУ, кандидат психологических наук, доцент*

**Аннотация.** В статье рассматривается процесс коррекции траектории движения спортсменов на основе координат, полученных с помощью безмаркерных систем захвата движения, в соответствие к глобальной системе координат. Описан алгоритм коррекции траектории движения, основанный на векторной алгебре, в частности на базе матриц поворота. Предложенный подход обеспечивает получение точных количественных параметров для последующего биомеханического анализа спортивных движений.

**Ключевые слова:** система захвата движения, матрица поворота, коррекция траектории движения.

**Введение.** Современные технологии анализа движений спортсменов, в частности безмаркерные системы захвата движения, открывают новые горизонты для исследования и оптимизации спортивных показателей. Эти системы предоставляют возможность отслеживать движения спортсменов без необходимости использования специализированных датчиков или маркеров, что функционально упрощает процесс и делает его более естественным [1]. Помимо удобства, такие технологии обладают значительными преимуществами, включая высокую точность регистрации движений и возможность получения данных в реальном времени. Тем не менее, для достижения достоверных результатов необходимо не только правильно регистрировать данные, но и корректно обрабатывать их. Одним из ключевых этапов данного процесса является коррекция координат траектории движения спортсмена в соответствие с глобальной системой координат, т.к. в процессе регистрации движения происходит отклонение траектории спортсмена от заданной. При этом, глобальная система координат задается во время калибровки видеокамер безмаркерных систем захвата движения. В работе представлен алгоритм коррекции траектории движения спортсменов, основанный на методах векторной алгебры [2].

Необходимо откорректировать полученные координаты так, чтобы они соответствовали глобальной системе координат. Такое преобразование является критически важным для обеспечения точности количественных данных, используемых в дальнейших расчетах параметров движения, особенно в сагиттальной плоскости, где сосредоточены большинство аналитических показателей. Современные технологии

анализа движений в спорте требуют высокой точности при обработке данных [3, 4]. Проблема преобразования координат заключается в необходимости адаптировать локальные данные о движении (например, координаты суставов) к заранее установленной глобальной системе координат, что критически важно для обеспечения достоверности последующих расчетов параметров движения.

**Основная часть.** В рамках кинематического анализа движений спортсменов большинство показателей сосредоточено в сагиттальной плоскости [5]. Однако, для корректного преобразования координат по сагиттальной оси необходимо также учитывать углы наклона в горизонтальной плоскости. Это связано с тем, что во время регистрации движений спортсмены могут двигаться не по прямой линии, а с небольшим отклонением в сторону. Поэтому для корректного преобразования координат сагиттальной оси (Y) необходимо учитывать и рассчитывать углы наклона в горизонтальной плоскости (XY).

Для коррекции координат траектории движения спортсмена необходимо выполнить несколько последовательных шагов, используя математические методы, такие как векторная алгебра, матричные преобразования и тригонометрические функции. Прежде всего, необходимо определить общий центр тяжести (ОЦТ) спортсмена. Нахождение общего центра тяжести является ключевым аспектом для кинематической оценки движений спортсменов. Системы захвата движения предоставляют данные о координатах суставов, что позволяет применять их для вычисления ОЦТ.

Алгоритм расчета ОЦТ следующий:

1 *Определение масс звеньев тела.*

Как правило, массы звеньев тела спортсмена (рисунок 1) рассчитываются на основании академических данных [6] относительной массы и расположения центров масс звеньев тела (таблица 1):

Таблица 1. Относительные массы и расположения центров масс звеньев тела

| Звено | Процент от массы тела | Расположение центра масс (от 1 целой части проксимального сустава) |
|---|---|---|
| Туловище | 43% | 0,44 |
| Плечо | 3% | 0,47 |
| Предплечье | 2% | 0,42 |
| Бедро | 12% | 0,44 |
| Голень | 5% | 0,42 |
| Стопа | 2% | 0,44 |

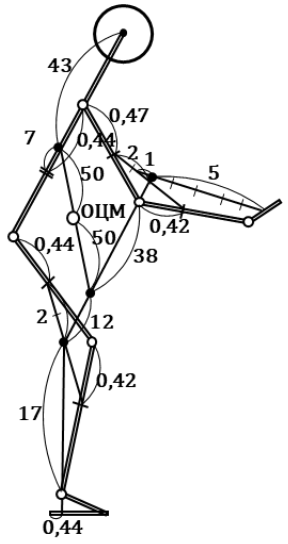



*Рисунок 1.* Схема расчета общего центра тяжести спортсмена

2 *Расчет координат центров тяжести звеньев тела.*

Для каждого звена тела рассчитывается центр тяжести в горизонтальной плоскости (YX) в соответствии с формулами [7]:

$$Y_{\text{цт}n} = \frac{y_{i1} + \lambda * y_{i2}}{1 + \lambda} \quad (1)$$

$$X_{\text{цт}n} = \frac{x_{i1} + \lambda * x_{i2}}{1 + \lambda} \quad (2)$$

$$\lambda = \frac{m_1}{m_2} \quad (3)$$

где n – обозначения звеньев тела,

i – обозначения суставов,

$y_{i1}$ – координаты проксимального сустава по горизонтальной оси,

$y_{i2}$ – координаты дистального сустава по горизонтальной оси,

$x_{i1}$ – координаты проксимального сустава по поперечной оси,

$x_{i2}$ – координаты дистального сустава по поперечной оси,

$m_1$ и $m_2$ – массы проксимального и дистального звеньев тела, их заменяем относительным расположением центров звена.

3 *Расчет координат общего центра тяжести.*

Общий центр тяжести определяется на основании ЦТ всех звеньев тела спортсмена в соответствии с формулами [6]:

$$Y_{\text{ОЦТ}} = \frac{\sum_{n=1}^{m} y_n * P_n}{\sum_{n=1}^{m} P_n} \quad (4)$$

$$X_{\text{ОЦТ}} = \frac{\sum_{n=1}^{m} x_n * P_n}{\sum_{n=1}^{m} P_n} \quad (5)$$

где $y_n$ – координата центр тяжести n-го звена тела по горизонтальной оси,

$x_n$ – координата центр тяжести n-го звена тела по поперечной оси,

m – общее количество звеньев тела,

$P_n$ – это вес n-го звена тела (берется относительное соотношение масс звеньев, например для голени 0,05P).

В данном исследовании применяется метод матрицы поворота, также известный как матрица направляющих косинусов [2]. Этот математический инструмент служит для выполнения ортогонального преобразования в евклидовых пространствах, что позволяет осуществлять вращение векторов без изменения их длины.

Схема поиска угла поворота и коррекции локальных координат траектории движения спортсмена в соответствии с глобальными координатами представлена на рисунке 2, алгоритм которой состоит из следующих этапов:

1 Построение вектора траектории движения спортсмена, используя начальные и конечные координаты ОЦТ;

2 Расчет угла поворота ( $\alpha$ ) вектора траектории движения ОЦТ относительно глобальной системы координат:

$$\alpha = \gamma - \beta \quad (6)$$

$$\gamma = \cos^{-1}\left(\frac{x_{\text{ОЦТн}}}{L}\right) \quad (7)$$

$$\beta = \cos^{-1}\left(\frac{x_{\text{ОЦТк}}}{L}\right) \quad (8)$$

где $\gamma$ – угол между вектором L’ и горизонтальной осью,

$\beta$ - угол между вектором $L_{ОЦТк}$ и горизонтальной осью,

$ОЦТ_н\,(x_{ОЦТн}y_{ОЦТн}), ОЦТ_к(x_{ОЦТк}y_{ОЦТк})$ - начальная и конечная координаты перемещения ОЦТ.

При этом, $L = L_{ОЦТк} = L’$ – длина векторов $ОЦТ_к$ и ОЦТ’, определяются по формуле вычисления евклидова расстояния между двумя точками:

$$L = \sqrt{(x_{ОЦТк}{}^2 + y_{ОЦТк}{}^2)} \tag{9}$$

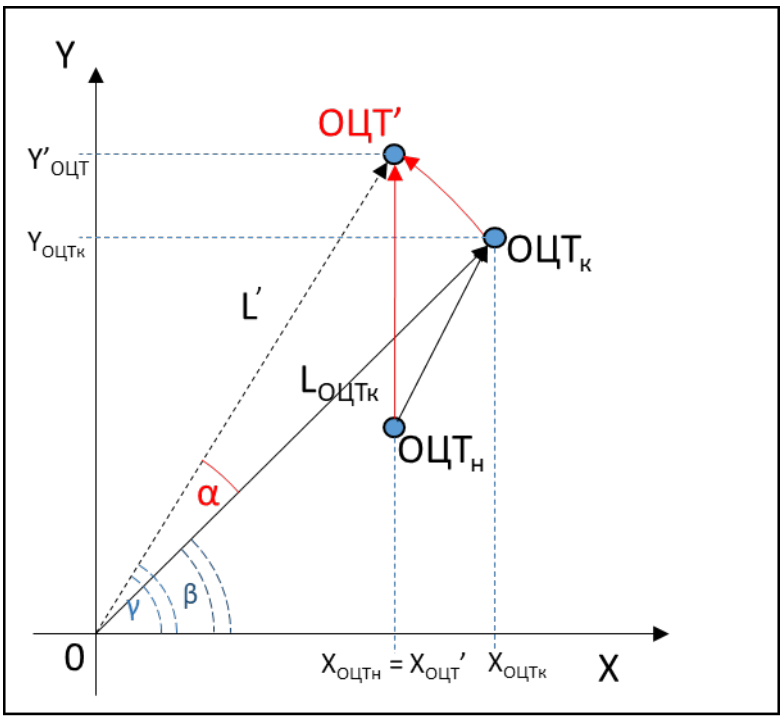



*Рисунок 2.* Схема поворота вектора траектории движения ОЦТ (где $\alpha$ – угол поворота вектора, $ОЦТ_Н$ $ОЦТ_К$ – исходная траектория движения ОЦТ, $ОЦТ_НОЦТ’$ – откорректированная траектория движения ОЦТ)

3 Коррекция координат к глобальной системе осуществляется с использованием матрицы поворота:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha & \mp\sin\alpha \\ \pm sin\,\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{12}$$

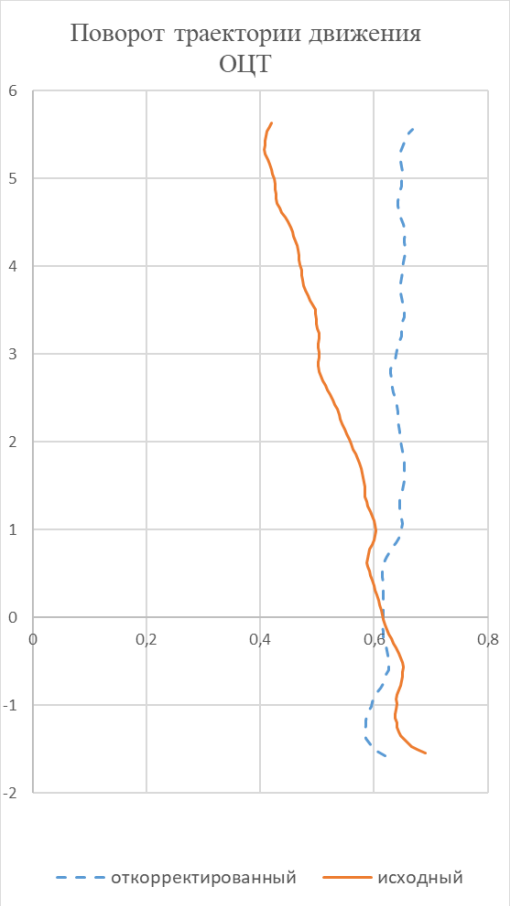



*Рисунок 3.* Пример откорректированной траектории ОЦТ спортсмена

Новые координаты ( $x'y'$) в результате поворота точки ($x_{\text{ОЦТк}}y_{\text{ОЦТк}}$) будут иметь следующий вид:

$$\begin{aligned} x' &= x_{\text{ОЦТк}} \cos\alpha \mp y_{\text{ОЦТк}} \sin\alpha \\ y' &= y_{\text{ОЦТк}} \cos\alpha \pm x_{\text{ОЦТк}} \sin\alpha \end{aligned} \qquad (13)$$

Знак синуса в формулах 12-13 зависит от направления вектора поворота: при повороте по часовой стрелке - используются верхние знаки, против часовой – нижние знаки.

После нахождения угла поворота и применения соответствующих уравнений к сагиттальным и поперечным координатам для всех временных моментов, мы можем быть уверены в том, что полученные данные обладают необходимой точностью и надежностью. Пример откорректированной траектории ОЦТ спортсмена представлен на рисунке 3.

**Заключение.** Коррекция координат в анализе движения спортсменов является критически важным этапом, обеспечивающим получение точных и достоверных результатов. Представленный алгоритм коррекции траектории движения спортсменов может быть использован в рамках кинематического анализа движений и для разработки точных моделей движения. Это в свою очередь, может существенно повысить качество и эффективность оценки и анализа двигательных действий. Потенциальное приложение данного подхода включает интеграцию в системы мониторинга и анализа движений в реальном времени, что позволит тренерам и спортсменам оперативно адаптироваться и корректировать тренировочные режимы.

Такие решения могут значительно улучшить процесс подготовки спортсменов и повысить спортивные достижения.

## Список литературы

[1] Cronin NJ, Feasibility of OpenPose markerless motion analysis in a real athletics competition / NJ Cronin, J Walker, CB Tucker [et al.] // Front. Sports Act. Living. – 2024. – Vol. 5:1298003.

[2] Rotation-matrix [Electronic resource]. – Mode of access: https://www.geeksforgeeks.org/rotation-matrix. – Date of access: 16.01.2025.

[3] Preatoni, E. Motion analysis in sports monitoring techniques: Assessment protocols and application to racewalking / E. Preatoni, La Torre [et al.] // Medicina Dello Sport. – 2010. – Vol. 63. – P. 327-342.

[4] Johnson, J.D. Validation of machine vision and action sport cameras for 3D motion analysis model reconstruction / J.D. Johnson, M. Hales, R. Emert // Sci Rep. – 2023. – Vol. 13. – Mode of access: https://doi.org/10.1038/s41598-023-46937-9. Date of access: 05.02.2025.

[5] Dorschky E. CNN-Based Estimation of Sagittal Plane Walking and Running Biomechanics From Measured and Simulated Inertial Sensor Data / E. Dorschky, M. Nitschke, F. Martindale Christine [et al.] // Bioeng. Biotechnol. – 2020. – Vol. 8.

[6] Бондаренко, К. К. Биомеханика: практическое пособие / К. К. Бондаренко, А. Е. Бондаренко // Гомельский гос. ун-т им. Ф. Скорины. – 2019. – С. 45.

[7] Центр масс [Electronic resource]. – Mode of access: https://edu-potential.ru/images/catalog/physics/Massa_center.pdf. Date of access: 05.02.2025.

**Авторский вклад**

**Турантаева Гарсия Гаврильевна** – разработка алгоритма анализа данных, сбор и систематизация фактических данных, написание статьи.

**Давыдова Надежда Сергеевна** – постановка задач исследования и формулировка гипотез, разработка структуры статьи**.**

**Васюк Валерий Евстафьевич** – анализ и интерпретация результатов исследования, рецензирование статьи**.**

**Татарко Кристина Игоревна -** обзор литературы и аргументация выбора методов исследования, форматирование статьи.

# ALGORITHM OF ATHLETES' TRAJECTORY CORRECTION BASED ON ROTATION MATRICES,

***G.G. Turantaeva***
*PhD student of sports engineering department of BNTU*

***N.S. Davydova***
*Associate Professor of the Academic department of infocommunication technologies, PhD of Technical Sciences, Associate Professor*

***V.E. Vassiouk***
*Leading specialist of the center for coordination of scientific, methodological and innovative activities of BSUPC, Candidate of Pedagogical Sciences, Associate Professor*

***K.I. Tatarko***
*Candidate of Psychological Sciences, Associate Professor of the Department of Social and Organizational Psychology BSU*

**Abstract.** The article deals with the process of converting athletes' motion trajectory coordinates obtained by markerless motion capture systems into correspondence to the global coordinate system. The algorithm of athletes' motion trajectory transformation based on vector algebra, in particular on the basis of rotation matrices, is described. The proposed approach provides accurate quantitative parameters for subsequent biomechanical analysis of sports movements.

**Keywords:** motion capture system, rotation matrix, transformation.

УДК [304.2]

# BIG DATA В HR-АНАЛИТИКЕ

***В.В. Шаталова***
*директор БГУИР филиал МРК, канд. техн.наук, доцент*
shatalova@bsuir.by

***Т.В. Казак***
*заведующий кафедрой инженерной психологии и эргономики БГУИР, член-корреспондент Международной академии психологических наук, доктор психологических наук, профессор*
kazak@bsuir.by

***В.В. Шаталова***

*Окончила Белорусский государственный университет информатики и радиоэлектроники, факультет компьютерного проектирования. Работает в должности директора БГУИР филиал «Минский радиотехнический колледж».*

***Т.В. Казак***

*Заведующий кафедрой инженерной психологии и эргономики Белорусского государственного университета информатики и радиоэлектроники, доктор психологических наук, профессор, член-корреспондент Международной академии психологических наук*

**Аннотация.** В статье рассмотрены возможности бизнес-аналитики (Business Intelligence, BI) и анализа больших объемов информации (BigData) для улучшения работы с персоналом, повышения эффективности HR-процессов, производительности предприятия и улучшения бизнес-результатов.

**Ключевые слова:** BigData, бизнес-аналитика, HR-аналитика, бизнес решения, человеческие ресурсы.

Сегодня HR-данные становятся одним из наиболее ценными активами любого предприятия, поэтому использование бизнес-аналитики (Business Intelligence, BI) для анализа больших объемов информации (BigData) позволяет HR-специалистам перейти от интуитивных решений к обоснованному подходу в управлении человеческими ресурсами. Это не только повышает эффективность HR-процессов, но и помогает улучшить бизнес-результаты. Технологии BigData позволяют анализировать ценные сведения, выявлять скрытые закономерности и тренды, прогнозировать будущее события и принимать более обоснованные управленческие решения на основе фактов и данных.

BigData и BI тесно связаны между собой. Большие данные это огромные объемы информации, которые могут быть использованы для анализа и принятия решений. BI-аналитика предлагает инструменты и методы для исследования этой информации с целью получения данных, обеспечения их целостности и сохранности, предоставления информации для принятия обоснованных решений для улучшения работы с персоналом и повышения производительности предприятия.

BigData в контексте HR-аналитики - это использование огромных объемов структурированных и неструктурированных данных, собранных из различных источников, для более глубокого анализа и прогнозирования процессов управления персоналом. Эти

данные могут включать информацию о производительности сотрудников, их вовлеченности, текучести кадров, обучении, взаимодействиях в команде, личных предпочтениях и многое другое.

HR-аналитика - это процесс сбора, обработки, анализа и интерпретации данных о персонале для принятия обоснованных управленческих решений. HR-аналитика позволяет решать задачи во всех направлениях управления персоналом [1]:

- организационно-штатном сопровождении;
- результативность и эффективность сотрудников;
- подборе и адаптации сотрудников;
- компенсациях и льготах;
- обучении и развитии;
- управлении эффективностью.

Виды HR-аналитики:

1 Дескриптивная (описательная) - сбор и обобщение первичных данных, которые помогают понять, что происходило в прошлом и имеет место в настоящее время (что происходит здесь и сейчас?)

2 Стратегическая - на основе имеющихся показателей составляется стратегический план развития (что нужно делать?)

3 Предиктивная (предсказательная) - строятся модели и сценарии развития событий в будущем, используя методы описательной статистики или нейросети (что могло бы произойти?)

4 Прогнозная - позволяет более точно прогнозировать развитие ситуации, основываясь на показателях компании и компаний-конкурентах (что можно сделать, чтобы это случилось?)

Стандартные задачи, которые позволяет решать HR-аналитика: оценка текучести кадров, прогнозирование фонда оплаты труда, расчет показателей качества подбора персонала, структурирование персонала по уровню классификации, отслеживание времени присутствия сотрудников на рабочем месте, исследование заработных плат по интересующим позициям на рынке труда и другое. Другие, менее стереотипные, но более интересные задачи, решение которых позволяет отладке HR-процессов: анализ вовлеченности сотрудников, выявление наиболее перспективных, с высоким потенциалом, сотрудников, оценка вероятности увольнения кадров, определение факторов текучести персонала, оценка вероятности увольнения сотрудников в период испытательного срока, обозначение опережающих и запаздывающих показателей и другое.

HR-аналитика позитивно влияет на управленческие решения за счет скорости (автоматизированный сбор и анализ данных), качества (сокращение доли ошибок, связанных с субъективными оценками), обоснованности (выявление множества скрытых взаимосвязей в управлении персоналом). Основное отличие HR-аналитики от простейшего статистического анализа заключается в следующем:

1 Объект аналитики - не только корпоративные базы данных, но и различные внешние источники.

2 Применение передовых методов и инструментов, включая математическую лингвистику, искусственный интеллект, программы статистического анализа и другое.

3 Использование разнообразных инструментов визуализации бизнес-данных.

4 Создание прогностических моделей, определяющих степень влияния решений в области управления персоналом на бизнес-процессы.

Данные, которые поступают из различных источников имеют разные форматы:

– структурированные данные - данные, которые могут быть организованы в виде таблицы и базы данных, записи в системах управления персоналом, результаты опросов, данные о зарплатах и бонусах, информация о производительности сотрудников;

– неструктурированные данные - данные, которые не поддаются традиционному формализованному анализу: текстовые данные из мессенджеров, электронные письма, отзывы сотрудников, записи из социальных сетей и другое.

Для анализа BigData в HR используются современные инструменты и технологии: машинное обучение, искусственный интеллект, алгоритмы предсказания, глубокое обучение, системы обработки естественного языка, которые позволяют собирать, очищать, обрабатывать и анализировать данные, превращая их в нужную информацию.

Эти технологии позволят осуществить комплексный сбор и всесторонний анализ человеческого капитала с учетом общей эффективности бизнеса и отдельных направлений деятельности предприятия. В частности, покрыть каждую из 9 областей HR-аналитики [2]:

1 Прогнозирование текучести кадров - наиболее критичная проблема для предприятий, особенно работающих в конкурентных областях. Биг Дата позволяет проанализировать, какие сотрудники с наибольшей вероятностью могут покинуть предприятие на основе таких факторов как:

– производительность (ее снижение, может быть индикатором, что сотрудник не удовлетворен своей работой);

– история увольнений (анализ данных о сотрудниках, которые ранее уволились из компании, позволяет выявить повторяющиеся паттерны, такие как: возраст, стаж работы, уровень заработной платы, стажировки, обучение и другое);

– уровень вовлеченности (анализ данных вовлеченности, опросы, мониторинг рабочих процессов для выявления признаков демотивации).

2 Оптимизация подбора и найма - помогает сделать подбор кандидатов более эффективным, ускоряя поиск и улучшая точность подбора, а также минимизировать человеческий фактор, который часто встречается при принятии решений о кандидатах:

– анализ резюме (алгоритмы машинного обучения могут автоматически анализировать резюме, соискателей, сопоставляя их с требованиями вакансии и предсказывая вероятность успешной работы кандидата на той или иной должности);

– анализ социальных сетей и профессиональных платформ (помогаю выявить дополнительные качества кандидатов);

– прогнозирование успеха кандидатов (с помощью алгоритмов, которые анализируют прошлые успешные и неудачные наймы, можно создать профиль идеального кандидата для конкретной позиции и предсказать вероятность успешной адаптации нового сотрудника).

3 Управление производительностью сотрудников - помогает точно оценивать производительность сотрудников и корректировать стратегии управления:

– отслеживание ключевых показателей (собираются данные о результатах работы сотрудников, их активности, временных затратах на выполнение задач, взаимодействие с коллегами и клиентами);

– анализ трендов (с помощью аналитики можно увидеть, какие сотрудник или группы сотрудников работают наиболее эффективно, а какие нуждаются в дополнительной мотивации и обучении);

– использование поведенческих данных (мониторинг данных о взаимодействии сотрудников, вовлеченности в рабочие процессы и времени, которые проводят на различных платформах, внутренних форумах, мессенджерах, помогает предсказать возможные проблемы с производительностью и предложить пути решения).

4 Обучение сотрудников - помогает выявить пробелы в знаниях и навыках сотрудников, а также оптимизировать образовательные программы:

– персонализированное обучение (анализирует данные о прошлых обучающих сессиях и успехах сотрудников, возможны персонализированные программы обучения, которые будут наиболее эффективные для каждого сотрудника);

– анализ эффективности обучения (с помощью больших данных можно отслеживать как обучение влияет на производительность сотрудников, корректировать образовательные стратегии в реальном времени);

– оценка потребностей в обучении (данные о текущих проблемах и пробелах в навыках могут быть использованы для предсказания того, какие навыки будут востребованы в будущем, и на основе данных создавать новые программы для подготовки сотрудников).

5 Анализ вовлеченности и удовлетворенности сотрудников - мониторинг вовлеченности сотрудников важная область применения BigData в HR, которые позволяет собирать и анализировать данные о том, как сотрудники взаимодействуют с предприятием, их уровень удовлетворенности, а также факторы, которые влияют на их вовлеченность:

– социальные сети и коммуникации (мониторинг внутренних коммуникаций позволяет выявить, насколько удовлетворены сотрудники своей работой, какие проблемы их волнуют, какими темами интересуются);

– опросы и обратная связь (собирают и анализируют данные с помощью регулярных опросов, можно получить объективную картину того, насколько сотрудники удовлетворены своей работы, условиями труда и корпоративной культурой).

6 Инклюзивность и разнообразие - помогает в анализе данных, связанных с разнообразием и инклюзивностью, а также в создании более справедливой рабочей среды, для формирования эффективных команд с учетом зависимости общей результативности группы от входящих в ее состав индивидов:

– анализ данных о составе команды (использование данных о поле, возрасте, этнической принадлежности, образовании и других демографических характеристиках позволяет отслеживать разнообразие в компании и выявить области для улучшения);

– обнаружение предвзятости (алгоритмы могут выявить скрытые предвзятости в процессах набора, оценки и продвижения сотрудников, что помогает HR-отделы корректировать их стратегии для повышения справедливости и инклюзивности).

7 Аналитика лидерства (руководства) - помогает оценить вклад руководителей с помощью анализа результативности всей организации и обратной связи от сотрудников, полученной методом анонимных опросов, фокус-групп и личных интервью.

8 Аналитика корпоративной культуры – изучение неформальных (и часто негласных) правил поведения и взаимодействия в компании, чтобы сформулировать целевую культуру, влиять на нее, прослеживая и отмечая тенденции, а также своевременно выявлять нежелательные изменения.

9 Аналитика индивидуальной результативности сотрудников – для поощрения каждого работника в зависимости от его вклада, а также определения зависимости между показателями деятельности специалиста и результатами оценки его навыков и компетенций.

Преимуществами BigData в HR-аналитике являются:

1 Улучшение принятия решений - с помощью данных и анализа HR-менеджеры могут принимать более обоснованные решения, основанные на фактах, а не на интуиции или субъективных оценках.

2 Повышение эффективности - автоматизации и оптимизация HR-процессов (найм, обучение, удержание сотрудников, позволяют сэкономить время и ресурсы).

3 Прогнозирование и предотвращение проблем - использование больших данных позволяет предсказывать текучесть кадров, снижение производительности труда.

4 Индивидуальный подход - с помощью BigData можно создавать персонализированные программы для сотрудников, что повышает их вовлеченность и мотивацию.

BigData в HR-аналитике является мощным инструментом для повышения эффективности работы с персоналом и оптимизации бизнес-процессов, что позволяет

предприятиям принимать более обоснованные решения, прогнозировать проблемы и улучшать внутренние процессы.

## Список литературы

[1] Применение Big Data в HR-менеджменте. [Электронный ресурс]. – Режим доступа: https://trendfactor.ru/primenenie-big-data-v-hr-analitike

[2] Big Data в кадровом менеджменте. [Электронный ресурс]. – Режим доступа: https://www.hr-director.ru/article/67163-big-data-v-menedjmente-18-m6

[3] Н. Кхан, М. Дэйв HR-аналитика: Практическое руководство по работе с персоналом на основе больших данных. М.: Альпина Паблишер, 2022. – 384 с.

# BIG DATA IN HR - ANALYTICS

***V.V. Shatalova***
*Deputy Dean of the Faculty of Computer-Aided Design, PhD, Associate Professor*

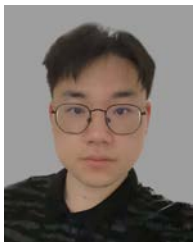
***T.V. Kazak***
*Head of the Department of Engineering Psychology and Ergonomics, Corresponding Member of the International Academy of Psychological Sciences, Doctor of Psychological Sciences, Professor*

**Abstract.** The article discusses the possibilities of business intelligence (BI) and analysis of large amounts of information (BigData) to improve work with staff, increase the efficiency of HR processes, enterprise productivity and improve business results.

**Key words:** Big Data, business analytics, HR analytics, business solutions, human resources.

UDC 004.896:004.942

# ANALYSIS AND REVIEW OF MULTIMODAL TRAJECTORY PREDICTION METHODS IN COMPLEX DYNAMIC SCENES: EVOLUTION FROM CLASSICAL MODELS TO DEEP LEARNING



***Yi Tang***
*Ph.D. student from the*
*Belarusian State University*
*of Informatics and*
*Radioelectronics*
tangyijcb@163.com

***Yi Tang***
*A Ph.D. student from the Belarusian State University of Informatics and Radioelectronics. The research area is about autonomous vehicle under dynamic and complex environments.*

**Abstract.** This article provides a systematic review of the mainstream algorithms and methods in the field of trajectory prediction for autonomous vehicles, categorizing them into four major approaches: traditional statistical methods, machine learning-based methods, deep learning-based methods, and hybrid models. Through a comprehensive analysis of the principles, strengths, weaknesses, and relevant literature of each approach, this study offers a detailed comparison of their performance characteristics and delves into their respective application scenarios. Furthermore, based on the current state of research, the article explores future directions for trajectory prediction technologies and proposes corresponding research recommendations.

**Keywords**: Trajectory Prediction; Autonomous driving; Machine learning; Deep learning.

**Introduction.** With the continuous growth in global vehicle ownership, issues such as traffic congestion, environmental pollution, and road safety have become increasingly severe. The emergence of autonomous driving technology offers new solutions to these challenges. By integrating advanced sensors, artificial intelligence algorithms, and communication technologies, autonomous vehicles can achieve self-navigation, path planning, and environmental perception, thereby improving traffic efficiency, reducing human driving errors, and lowering energy consumption [1]. Among these technologies, trajectory prediction is a critical component of autonomous driving systems. It enables the vehicle to anticipate the motion trends of surrounding traffic participants, providing the decision-making system with sufficient time to plan safe and efficient driving paths. This capability effectively mitigates potential collision risks and enhances the overall performance and reliability of autonomous vehicles.

In recent years, significant progress has been made in the field of vehicle trajectory prediction, with researchers proposing a variety of prediction methods that can be broadly categorized into four main types as illustrated in Figure 1: traditional statistical methods, machine learning-based methods, deep learning-based methods, and hybrid models. Traditional statistical methods, such as Kalman filtering and Monte Carlo methods, have been widely used in early trajectory prediction research due to their computational efficiency and simplicity. However, these methods exhibit notable limitations when dealing with complex traffic scenarios and vehicle interactions. Machine learning methods, such as Support Vector Machines (SVM), Gaussian

Processes (GP), and Hidden Markov Models (HMM), learn patterns and rules from historical data, making them more effective in handling complex traffic scenarios. Nonetheless, they require high-quality labeled data, extensive feature engineering, and are sensitive to data quality. Deep learning methods, including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Graph Neural Networks (GNNs), excel in feature extraction and can handle large-scale datasets [2]. They demonstrate outstanding prediction accuracy, particularly in complex traffic scenarios. However, these methods come with high model complexity, long training times, and significant computational resource demands. Hybrid models combine the strengths of traditional statistical methods, machine learning, and deep learning, further improving prediction accuracy and robustness. Nevertheless, their model construction and optimization processes are complex, with substantial computational overhead.

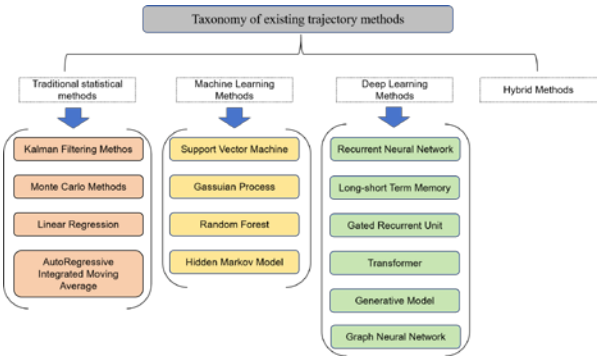



*Figure 1.* Classification of existing trajectory prediction methods

This study aims to comprehensively analyze and compare different trajectory prediction methods, exploring their underlying principles, strengths, weaknesses, and applicable scenarios. Through a systematic review of the existing literature, this paper summarizes the latest advancements and research outcomes in various methods, providing a comprehensive reference for researchers and technology developers. Finally, the paper discusses future directions for trajectory prediction technologies and proposes corresponding research recommendations, offering theoretical guidance and practical insights for subsequent research.

**Task Definition.** The task of trajectory prediction focuses on estimating possible future trajectories for target agents based on their observed motion history and surrounding environmental context. In the context of autonomous driving scenarios involving multiple dynamic agents (including autonomous vehicles), the surrounding map information is denoted as $\mathbf{M}$, while the observed trajectories of all agents are collectively represented as $\mathbf{X} = \left\{\mathbf{x}_0, ..., \mathbf{x}_{N_a}\right\}$. Specifically, $\mathbf{x}_i = \left\{\mathbf{x}_i^{-H+1}, ..., \mathbf{x}_i^0\right\}$ denotes the historical trajectory of the agent over the preceding time steps. To address the inherent multimodal nature of future motion, a multi-agent motion predictor is implemented to generate plausible future trajectories for all agents in the scene, represented as $\mathbf{Y} = \left\{\mathbf{y}_0, ..., \mathbf{y}_{N_a}\right\}$. For each individual agent, possible future trajectories and their associated probability scores are predicted, enabling the modeling of diverse potential outcomes. These multimodal trajectories are expressed as $\mathbf{y}_i = \left\{\mathbf{y}_i^1, ..., \mathbf{y}_i^k\right\}$, where $\mathbf{y}_i^k = \left\{\mathbf{y}_{i,1}^k, ..., \mathbf{y}_{i,T}^k\right\}$ represents the predicted trajectory for the agent over a prediction horizon. The probability scores corresponding to these predicted trajectories are captured as $\mathbf{s}_i = \left\{s_i^1, ..., s_i^k\right\}$, providing a quantitative measure of their likelihood.

Table 1. Comparison of commonly used trajectory prediction datasets. Here ‘size’ denotes the number of trajectories in the dataset; ‘History’ and ‘Prediction’ denote the observed and future trajectories for each agent; ‘HD Map’ stands for High-definition maps that include lane markings, traffic rules, and other semantic information.

| Dataset | Type | Size | History | Prediction | Participant | Environment |
|---|---|---|---|---|---|---|
| NGSIM [3] | Highway | 6,000+ | 3s | 5s | Vehicles | - |
| HighD [4] | Highway | 110,000 | 3s | 5s | Vehicles | - |
| Argoverse [5] | Urban | 324,557 | 2s | 3s | Vehicles | HD Map |
| INTERACTION [6] | Urban | 11,000+ | 2s | 3s | Vehicles, Pedestrians | HD Map |
| ETH/UCY [7] | Pedestrian | 1,536 | 3.2s | 4.8s | Pedestrians | - |
| Lyft Level 5 [8] | Urban | 1,000+ | 1s | 3s | Vehicles | HD Map |
| ApolloScape [9] | Urban | 100,000+ | 2s | 3s | Vehicles, Pedestrians | HD Map |
| NuScenes [10] | Urban | 40,000 | 2s | 4s | Vehicles, Pedestrians, Cyclists | HD Map |

**Dataset description and comparison.** The choice of datasets in trajectory prediction research is critical for evaluating algorithm performance and training models effectively. Trajectory prediction datasets can be broadly classified into three categories based on their focus and characteristics: Highway Datasets, Urban Datasets, and Pedestrian Datasets. Highway Datasets (e.g., NGSIM, HighD) focus on vehicle behavior in highway scenarios, featuring simple traffic interactions, longer historical and prediction times (e.g., 3s history, 5s prediction), and are ideal for studying long-term vehicle maneuvers and lane changes. Urban Datasets (e.g., Argoverse, INTERACTION, Lyft Level 5, ApolloScape, NuScenes) emphasize complex traffic interactions in urban environments, characterized by shorter historical and prediction times (e.g., 1s–2s history, 3s prediction), inclusion of multiple traffic participants (vehicles, pedestrians, cyclists), and often provide HD maps for contextual information, making them suitable for studying multi-agent interactions and collision avoidance. Pedestrian Datasets (e.g., ETH/UCY) are dedicated to pedestrian movements in open spaces, featuring larger time steps (e.g., 0.4s) due to slower pedestrian speeds, a focus on social interactions and group behavior, and are ideal for studying pedestrian trajectory prediction and crowd dynamics. Table 1 presents a comparative analysis of commonly used trajectory prediction datasets, highlighting their key characteristics and differences.

**Evaluation metrics.** In trajectory prediction tasks, evaluation metrics are used to quantify the differences between the predicted results and the ground truth trajectories. Trajectory prediction performance metrics are essential for evaluating the accuracy and reliability of predictive models. Commonly used metrics include the Final Displacement Error (FDE), which measures the Euclidean distance between the predicted and true trajectory endpoints, and the Average Displacement Error (ADE), defined as the mean Euclidean distance between the predicted and actual trajectories over all time steps. Additionally, the Root Mean Square Error (RMSE) quantifies the square root of the average squared differences between predicted and true trajectory points, whereas the Mean Absolute Error (MAE) calculates the mean of absolute

differences, providing a more intuitive measure of prediction deviation. Furthermore, the Miss Rate (MR) evaluates the proportion of predicted trajectory endpoints that deviate from the true endpoints by more than a predefined threshold, serving as a critical indicator of model reliability in safety-critical applications such as autonomous driving [11]. These metrics collectively offer a comprehensive framework for assessing both endpoint prediction accuracy and overall trajectory consistency. Below are commonly used evaluation metrics and their corresponding calculation formulas are shown below in formula (1):

$$\begin{cases} \text{FDE} = \sqrt{(x_T - \hat{x}_T)^2 + (y_T - \hat{y}_T)^2} \\ \text{ADE} = \dfrac{1}{T}\sum_{t=1}^{T}\sqrt{(x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2} \\ \text{RMSE} = \sqrt{\dfrac{1}{T}\sum_{t=1}^{T}\left[(x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2\right]} \\ \text{MAE} = \dfrac{1}{T}\sum_{t=1}^{T}\left[|\, x_t - \hat{x}_t \,| + |\, y_t - \hat{y}_t \,|\right] \\ \text{MR} = \dfrac{1}{N}\sum_{i=1}^{N}\mathbb{I}(\text{FDE}_i > \text{thres}\mathit{hold}) \end{cases} \tag{1}$$

Where $(x_t, y_t)$ represents the position of the true trajectory at time step $t$, and $(\hat{x}_t, \hat{y}_t)$ represents the position of the predicted trajectory at time step $t$, with $t \in (0, T]$, where $T$ is the total number of time steps. Additionally, $N$ denotes the total number of samples, and $\mathbb{I}(\cdot)$ is an indicator function. The value is 1 if the condition is true, else 0.

**Traditional statistical methods.** Kalman filtering, linear regression, time series models (e.g., ARIMA), and Monte Carlo simulation are widely used methods in early trajectory prediction research. These approaches predict future trends based on historical trajectory data and are favored for their computational efficiency, simplicity, and ease of implementation [12]. They are particularly suitable for handling low-dimensional and structured trajectory data scenarios.

For example, the Kalman filter effectively addresses the uncertainty in vehicle states by recursively updating the state estimates. It models the uncertainty of the current vehicle state and its physical dynamics using Gaussian distributions, producing the mean and covariance matrix of the vehicle state at each future time step. This allows the calculation of an average trajectory with associated uncertainties. While the Kalman filter accounts for the uncertainty of predicted trajectories, its unimodal Gaussian distribution is insufficient to represent diverse operational modes.

The Monte Carlo method, on the other hand, generates potential future trajectories by randomly sampling input variables and applying physical models to approximate the state distribution. When generating trajectory samples, constraints such as limiting lateral acceleration below feasible thresholds or considering the physical limitations of the vehicle are applied to ensure the feasibility of operations. This method can predict the trajectories of traffic participants from either fully known states or uncertain states estimated through filtering algorithms.

Linear regression is one of the simplest and most intuitive methods among traditional statistical approaches. It constructs a predictive model by assuming a linear relationship between the dependent variable (target variable) and the independent variables (feature variables). For trajectory prediction, linear regression can fit the motion trend of a vehicle based on historical

trajectory data (e.g., position, velocity, time) and predict its future positions. The mathematical expression is below in (2):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon \quad (2)$$

Where $y$ represents the target variable (e.g., future position), $x_1, x_2, ..., x_n$ represent the feature variable (e.g., time or velocity), $\beta_1, \beta_2, ..., \beta_n$ denote the regression coefficients, and $\varepsilon$ is the error term. The advantage of linear regression lies in its simplicity and interpretability, making it suitable for structured trajectory data. However, as it assumes a linear relationship, it struggles to capture nonlinear motion patterns in complex traffic scenarios and thus performs limitedly in dynamic traffic environments.

**Machine learning based methods.** Unlike traditional statistical methods, machine learning-based approaches rely on data-driven models for trajectory prediction. Classic machine learning methods include GP, SVM, Random Forests, and HMM.

Gaussian Process is an effective tool commonly used in prototype trajectory approaches. It treats trajectories as samples from a Gaussian Process and predicts possible trajectories by measuring the similarity between historical trajectories and prototype sets. When applying Gaussian Processes, parameters are determined based on the samples, and different methods vary in how prototype trajectories are constructed. However, models trained with specific trajectory samples often struggle to generalize to other scenarios.

Support Vector Machines can learn and recognize driver behaviors in complex environments. The key lies in finding support vectors satisfying classification requirements and identifying the optimal hyperplane to maximize the margin between classified data. In trajectory prediction, driving behaviors such as left turn, right turn, or going straight are usually categorized, and a kernel function is used to map the input data to a high-dimensional space for linear classification to predict trajectories. However, this method requires predefined driving behaviors, and such definitions can influence the final prediction outcomes.

Random Forests, as a nonlinear model based on an ensemble of decision trees, learn the complex features in trajectory data by constructing multiple decision trees. The core idea is to split input trajectory features (e.g., position, velocity, acceleration) into subspaces, where each decision tree independently learns these subspaces and outputs trajectory predictions through an ensemble voting mechanism. In trajectory prediction, Random Forests possess nonlinear modeling capability and exhibit high robustness to noise and outliers. Additionally, they are fast to train and can manage large-scale trajectory data effectively. However, the results are based on a collection of trees, which weakens interpretability. Furthermore, Random Forests heavily depend on the selection of input features, and they may perform less effectively in complex multi-modal interaction scenarios [13].

The HMM is a statistical model used to represent systems with hidden states that can be inferred from observable data. In the context of trajectory prediction, HMM models the behavior of traffic participants as a sequence of hidden states and observations. The model assumes that the system is a Markov process, meaning the future state depends only on the current state and not on the sequence of events that preceded it. The process is shown below in equation (3):

$$p(\mathbf{s}, \mathbf{o}) = p(s_1) \prod_{t=1}^{T-1} p(s_{t+1} \mid s_t) \prod_{t=1}^{T} p(o_t \mid s_t) \quad (3)$$

Where $p(s_1)$ represents the probability distribution of the initial state, $p(s_{t+1} \mid s_t)$ denotes the state transition probabilities that describe how states evolve over time, $p(o_t \mid s_t)$ represents

the observation probabilities that describe how states generate observation data, $\mathbf{s} = (s_1, s_2, \ldots, s_T)$ refers to the hidden state sequence, and $\mathbf{o} = (o_1, o_2, \ldots, o_T)$ denotes the observation sequence.

**Deep learning based methods.** In recent years, the application of deep learning techniques in the field of trajectory prediction has been extensively studied. Due to their powerful feature learning capabilities, deep learning methods such as RNN, LSTM, Transformers, Gated Recurrent Units (GRUs), and GNN have demonstrated exceptional performance in capturing complex trajectory patterns. Figure 2 illustrates the general workflow of a sequential network. Historical trajectories are fed into the temporal model, which processes the data over multiple time steps. Each sequential unit automatically extracts features from the input data, and the final output generates the predicted future trajectory. These models are particularly effective in handling large-scale, complex, and highly diverse trajectory datasets, making them a robust choice for trajectory prediction tasks [14].

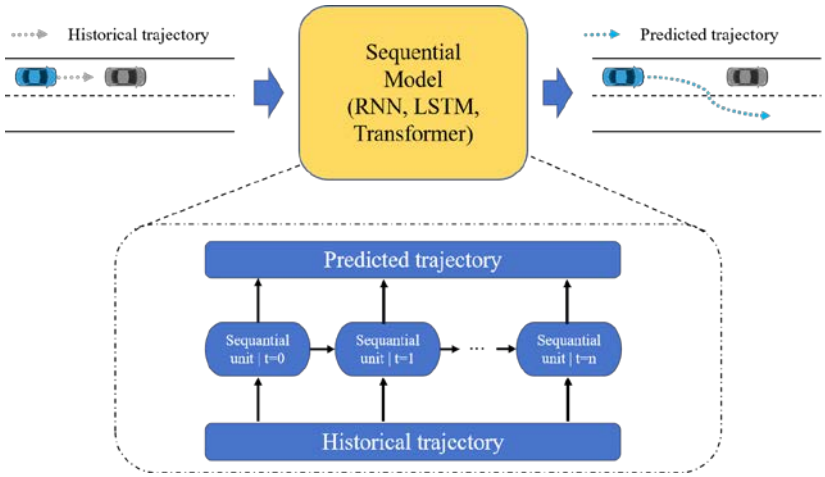



*Figure 2.* Process of sequential model

RNNs are specifically designed to handle sequential data by storing information from the previous time step and combining it with the current input to determine the output. However, when dealing with long sequences, RNNs are prone to issues such as vanishing or exploding gradients. Gated Recurrent Neural Networks, such as LSTM and GRU, effectively address these problems. Trajectory prediction models using RNNs can be categorized into single RNN models and multiple RNN models. Single RNNs are used for basic operations, unimodal trajectory prediction, or auxiliary models, while multiple RNNs are employed for more complex prediction tasks, such as predicting trajectories that consider the interaction with surrounding vehicles.

The attention mechanism, inspired by human cognitive processes, allows for the rapid selection of high-value information from a large dataset and has been widely applied in deep learning tasks. In trajectory prediction tasks, attention mechanisms are used to extract lane and vehicle attention to output the distribution of future trajectories or to model interactions among traffic participants. The Transformer model, which heavily relies on attention mechanisms, performs sequence-based tasks (e.g., machine translation) without using RNNs. Its advantages in handling time-series data have also been applied to trajectory prediction.

GNNs are suitable for vehicle trajectory prediction problems that rely on interaction-related factors, as shown in Figure 3; they treat each object in the environment as a node to form a graph, thereby capturing the dependencies among objects. Graph Convolutional Networks (GCNs) are the most popular GNN approach, extending convolutional operations from traditional image data to graph data. GCNs learn mapping functions to extract interaction-aware features from the

characteristics of nodes and their neighbors. For example, the GRIP model treats each vehicle at each sampling time as a node, considers temporal and spatial relationships as well as interaction states, and uses a GCN model combined with an LSTM encoder-decoder to predict the trajectories of surrounding vehicles. Additionally, vector map-based GNN methods treat vehicles and vector maps as nodes, leveraging GNNs to capture interaction features between vehicles and between vehicles and maps, thereby improving trajectory prediction accuracy.

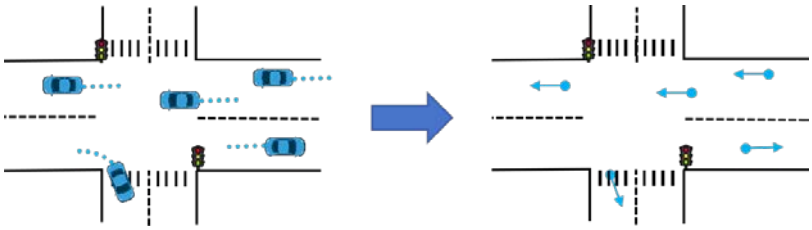


*Figure 3*. Illustration of graph-based representation. In this context, each agent is represented as a node in a graph, and the arrows of nodes indicate motion directions, velocities, and other relevant dynamic properties.

**Hybrid methods.** Hybrid methods aim to combine the strengths of various approaches to achieve higher prediction accuracy and robustness. For example, a Kalman filter can be used for initial trajectory prediction, followed by deep learning techniques to refine the results. Alternatively, combining Graph Neural Networks (GNNs) with Bayesian models can simultaneously capture complex trajectory relationships and uncertainty information. Such methods excel at integrating multimodal data and modeling complex relationships, but their algorithmic complexity and optimization challenges, coupled with high computational costs, limit their application in scenarios with strict real-time requirements.

Hybrid approaches integrate multiple types of methods to fully leverage their respective advantages, thereby improving the accuracy and robustness of trajectory prediction. For instance, physics-based models can be combined with learning-based methods, utilizing the prior knowledge of physics models and the data-driven capabilities of learning approaches to handle complex traffic scenarios more effectively.

**Analysis and discussions.** Trajectory prediction methods vary significantly in their assumptions, computational requirements, and the complexity of scenarios they can handle. To understand their relative strengths and limitations, it is essential to compare these approaches systematically. As illustrated in Table 2, We provide a concise analysis of the main trajectory prediction methods, focusing on their advantages, disadvantages, and potential application scenarios in dynamic environments.

**Challenges and future trends.** Trajectory prediction faces several key challenges that continue to hinder its performance in real-world scenarios. One significant issue is the long-tail distribution problem, where certain rare trajectory patterns appear infrequently in training data, making them difficult to predict accurately. Additionally, the real-time requirements of trajectory prediction impose strict constraints on computational efficiency, necessitating models that can deliver accurate predictions within milliseconds. Another challenge is data sparsity, where missing or incomplete data at certain time steps can compromise the model's ability to generalize effectively. Furthermore, the integration of multimodal data sources, such as visual inputs, radar signals, and high-definition maps, remains a critical challenge, as it requires designing mechanisms to effectively fuse heterogeneous data while preserving their complementary information. Lastly, adaptive dynamic modeling poses a significant challenge, as models must be

capable of adjusting their parameters in real time to account for variations in traffic scenarios, environmental conditions, and the behavior of surrounding agents.

Table 2. Comparative analysis of trajectory prediction approaches.

| Method | Description | Advantages | Disadvantages | Application Scenarios |
|---|---|---|---|---|
| Traditional Methods | Rely on physical models and kinematic assumptions (e.g., Kalman filters, ARIMA, Monte Carlo). | Computation efficient; suitable for simple traffic scenarios. | Struggle with nonlinear dynamics, complex interactions, and long-term dependencies. | Simple and static traffic scenarios; real-time prediction for low-complexity systems. |
| Machine Learning | Leverages labeled data for trajectory prediction (e.g., SVM, GP, Random Forest, HMM). | High accuracy in moderately complex scenarios; handles noise and stochasticity. | Sensitive to data quality; requires feature engineering; struggles with long-term dependencies. | Moderate-complexity traffic with sufficient labeled data; offline applications. |
| Deep Learning | Automatically extracts features and models complex interactions (e.g., RNN, CNN, Transformer, GNN). | Superior performance in high-dimensional, complex scenarios. | High computational cost; sensitive to hyperparameters; challenging deployment feasibility. | Complex and dynamic traffic scenarios; requires high computing resources. |
| Hybrid Methods | Combines traditional, ML, and DL methods to enhance prediction accuracy and robustness. | Leverages complementary strengths; robust in complex scenarios. | Intricate model design; increased computational demands; challenges in real-time deployment. | Highly dynamic and interactive scenarios; trade-offs between accuracy and efficiency. |

Looking toward the future, several promising trends and research directions are emerging to address these challenges. Multimodal data fusion is expected to play a pivotal role, as combining diverse data sources can significantly enhance prediction accuracy and robustness in complex environments. Adaptive dynamic modeling is another critical avenue, enabling models to dynamically recalibrate parameters based on real-time data, thereby improving their adaptability to diverse traffic conditions. Reinforcement learning, with its capacity to optimize sequential decision-making, offers potential for improving the model's ability to anticipate future trajectories by learning from feedback in dynamic environments. Federated learning is gaining traction as a privacy-preserving approach to model training, allowing collaborative learning across distributed datasets while safeguarding sensitive user data. Finally, enhancing the interpretability of trajectory prediction models is becoming increasingly important, as interpretable models not only foster trust among users but also provide insights into the underlying decision-making processes, which is crucial for safety-critical applications such as autonomous driving. Together, these advancements are expected to drive the field toward more robust, efficient, and trustworthy trajectory prediction systems.

**Conclusion.** In this work, we provide a comprehensive analysis of existing algorithms and methodologies in trajectory prediction, systematically exploring their application scenarios, strengths, and limitations. In the future, we will focus on developing efficient, multimodal trajectory prediction frameworks that address challenges such as long-tail distributions, real-time performance, and data sparsity in complex and dynamic environments.

## Reference list

[1] Parekh D, Poddar N, Rajpurkar A, et al. A review on autonomous vehicles: Progress, methods and challenges[J]. Electronics, 2022, 11(14): 2162.

[2] Rudenko A, Palmieri L, Herman M, et al. Human motion trajectory prediction: A survey[J]. The International Journal of Robotics Research, 2020, 39(8): 895-935.

[3] Coifman B, Li L. A critical evaluation of the Next Generation Simulation (NGSIM) vehicle trajectory dataset[J]. Transportation Research Part B: Methodological, 2017, 105: 362-377.

[4] Krajewski R, Bock J, Kloeker L, et al. The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems[C]//2018 21st international conference on intelligent transportation systems (ITSC). IEEE, 2018: 2118-2125.

[5] Chang M F, Lambert J, Sangkloy P, et al. Argoverse: 3d tracking and forecasting with rich maps[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 8748-8757.

[6] Zhan W, Sun L, Wang D, et al. Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps[J]. arXiv preprint arXiv:1910.03088, 2019.

[7] Pellegrini S, Ess A, Schindler K, et al. You'll never walk alone: Modeling social behavior for multi-target tracking[C]//2009 IEEE 12th international conference on computer vision. IEEE, 2009: 261-268.

[8] Li G, Jiao Y, Knoop V L, et al. Large car-following data based on lyft level-5 open dataset: Following autonomous vehicles vs. human-driven vehicles[C]//2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2023: 5818-5823.

[9] Huang X, Cheng X, Geng Q, et al. The apolloscape dataset for autonomous driving[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 954-960.

[10] Caesar H, Bankiti V, Lang A H, et al. nuscenes: A multimodal dataset for autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11621-11631.

[11] Shi S, Jiang L, Dai D, et al. Mtr++: Multi-agent motion prediction with symmetric scene modeling and guided intention querying[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.

[12] Huang Y, Du J, Yang Z, et al. A survey on trajectory-prediction methods for autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(3): 652-674.

[13] Bharilya V, Kumar N. Machine learning for autonomous vehicle's trajectory prediction: A comprehensive survey, challenges, and future research directions[J]. Vehicular Communications, 2024: 100733.

[14] Pavel M I, Tan S Y, Abdullah A. Vision-based autonomous vehicle systems based on deep learning: A systematic literature review[J]. Applied Sciences, 2022, 12(14): 6831.

## Author’s contribution

**Yi Tang -** conducted a comprehensive analysis of trajectory prediction methods, categorized existing approaches, and evaluated their applicability, advantages, and limitations across various scenarios.

УДК 004.021:004.75

# ИСПОЛЬЗОВАНИЕ МЕТОДА РАНЖИРОВАНИЯ ПРИ ОПРЕДЕЛЕНИИ СПИСКА ПОТЕНЦИАЛЬНЫХ ЗНАКОМЫХ ПОЛЬЗОВАТЕЛЯ

***Е.А. Бугаев***
*Инженер-программист*
*ООО «Whitesnake»*
buhaev.yauheni@gmail.com

***Е.А.Бугаев***

*Окончил Белорусский государственный технологический университет. Область научных интересов связана с разработкой эффективных и быстрых алгоритмов для обработки больших объемов данных. Интересы включают в себя теоретические и практические аспекты оптимизации, машинного обучения, параллельных и распределенных вычислений, анализа данных и искусственного интеллекта.*

**Аннотация.** Выполнен анализ подхода к планированию системы ранжирования при определении списка потенциальных знакомых. Описан основной алгоритм решения данной проблемы.

Разработана принципиальная схема работы сервиса ранжирования, показаны возможные варианты улучшения системы. Проведена оптимизация входных и выходных данных системы. Проведена оценка и выполнена оптимизация.

**Ключевые слова:** ранжирование, модель машинного обучения, граф, GNN, ROC-AUC.

**Введение**. В эпоху цифровых технологий, когда социальные связи играют решающую роль в личной и профессиональной жизни, поиск новых контактов и друзей становится актуальной задачей. Интернет-платформы и социальные сети предлагают множество возможностей для знакомства, однако определить наиболее подходящих людей может быть непросто.

Методы ранжирования социальных контактов предоставляют возможность упорядочивать и оценивать потенциальных знакомых с учетом различных критериев. Эти методы используют сложные алгоритмы для создания списка наилучших кандидатов, что значительно упрощает поиск и выбор новых социальных связей.

**Методы машинного обучения.** Поточечное ранжирование — один из методов машинного обучения, который используется для оптимизации алгоритмов рекомендаций. Этот метод основывается на бинарной классификации. Он анализирует каждый отдельный объект и определяет, относится ли он к той или иной категории на основе заданных критериев. Этот процесс помогает системе рекомендаций выделить наиболее подходящих потенциальных знакомых пользователю, увеличивая вероятность успешного совпадения интересов и предпочтений. Принципиальная схема работы классификатора представлена на рисунке 1.

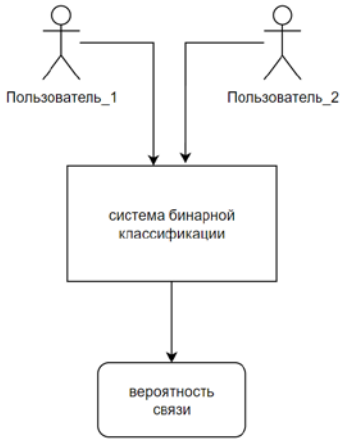



*Рисунок 1.* Бинарная классификация двух пользователей

**Проблемы интеграции социального контекста в рамках бинарной классификации.**

В связи с тем, что при определении потенциальных контактов, социальный контекст играет важную роль, его использование необходимо, что заставляет нас рассматривать при классификации не только непосредственно пользователей, но и их окружение. На рисунке 2 представлен простой пример социального контекста двух людей в двух коленах.

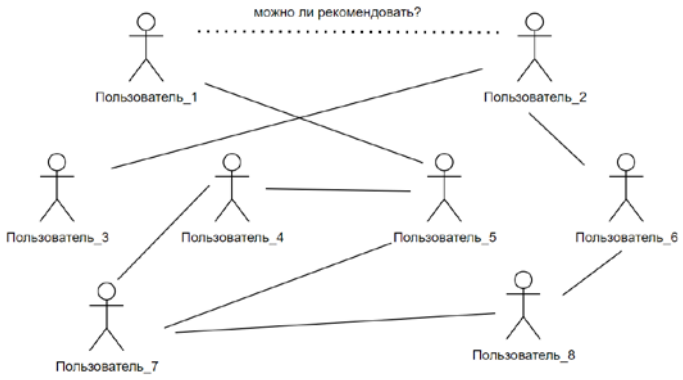



*Рисунок 2.* Пример социального контекста

Интеграция социального контекста в процессы бинарной классификации сталкивается с рядом проблем, среди которых можно выделить следующие:

1 Сложность измерения и оценки: социальный контекст включает в себя множество разнообразных и часто субъективных факторов, таких как интересы, предпочтения, стиль общения и отношения;

2 Многомерность данных: социальный контекст является многомерным и многообразным, что затрудняет его обработку и интеграцию в алгоритмы бинарной классификации;

3 Изменчивость контекста: социальный контекст не является статичным и может меняться со временем. Интересы, предпочтения и социальные связи пользователей могут изменяться, что требует постоянной адаптации и обновления алгоритмов классификации.

4 Конфиденциальность данных: использование социального контекста требует обработки большого объема личных данных пользователей. Это вызывает вопросы конфиденциальности и защиты данных, что требует внедрения строгих мер безопасности и соблюдения нормативных требований.

5 Интерпретируемость моделей: сложные модели машинного обучения, используемые для анализа социального контекста, могут быть трудными для интерпретации и объяснения. **Предсказание связей.** Для решения проблем, описанных выше, модель получает на вход граф, включающий в себя весь социальный контекст, что дает возможность использовать дополнительные данные. На рисунке 3 представлена схема передачи контекста, где пользователи – узлы, а их социальные контакты – ребра.

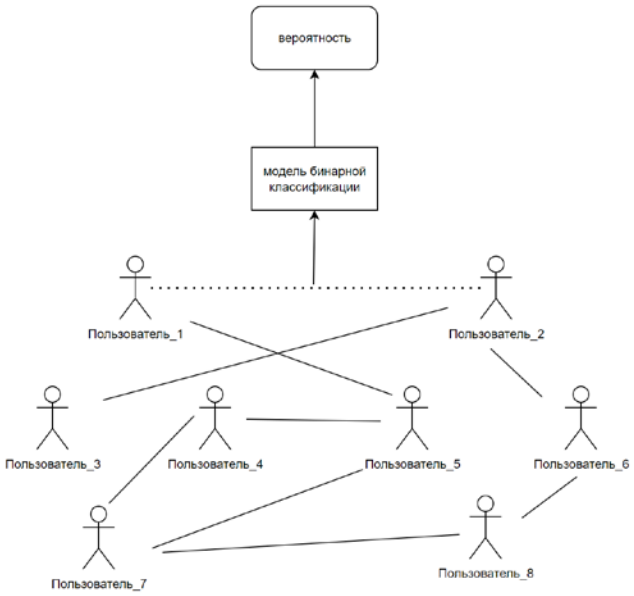



*Рисунок 3.* Передача социального вектора модели

Любая система машинного обучения состоит из подсистем, выполняющих конкретные функции. Можно выделить такие система, как:

- система данных;
- модель (система обучения);
- система оценки работы модели.

**Система данных и конструирование признаков.** Данные играют основную роль при разработке и обучению моделей машинного обучения. При работе с системой ранжирования мы работаем, в базовом виде, с тремя абстрактными типами данных, а именно:

- данные о пользователе;
- связи пользователей с другими пользователями;
- взаимодействия пользователей друг с другом.

| | affiliate_channel | affiliate_provider | age | country_destination | date_account_created | date_first_booking | first_affiliate_tracked | first_brow |
|---|---|---|---|---|---|---|---|---|
| 0 | direct | direct | NaN | NDF | 2010-06-28 | NaN | untracked | Chrome |
| 1 | seo | google | 38 | NDF | 2011-05-25 | NaN | untracked | Chrome |
| 2 | direct | direct | 56 | US | 2010-09-28 | 2010-08-02 | untracked | IE |
| 3 | direct | direct | 42 | other | 2011-12-05 | 2012-09-08 | untracked | Firefox |
| 4 | direct | direct | 41 | US | 2010-09-14 | 2010-02-18 | untracked | Chrome |

*Рисунок 4.* Пример данных о пользователе

На рисунке 4 представлен пример данных о пользователе. В большинстве своем они хранят данные, относящиеся непосредственно к конкретному пользователю.

Данные о связях позволяют понять отношения одного пользователя к другому. Взаимодействия, в свою очередь описывают непосредственно действия между, как самим пользователями, так и их активностью. Прекрасными примерами взаимодействия в разрезе социальных сетей могут быть:

- реакция на пост;
- заявка в друзья;
- просмотр профиля;
- комментарий.

На основании данных необходимо определить признаки, на базе которых будет производиться обучение модели.

Признаки можно разделить на сотни, если не тысячи различных типов, но в общем стоит рассматривать:

1 индивидуальные признаки пользователя: пол, рост, возраст, страна;

2 общие признаки пользователей: места, хобби, учебные заведения;

3 социальные признаки пользователей: посещаемость страницы профиля, общие друзья.

Корректность выделенных признаков индивидуальна для каждой, отдельно взятой, модели.

**Разработка модели машинного обучения.** Графовые нейронные сети (GNN) [1] — это тип нейронных сетей, предназначенных для обработки данных, представленных в виде графов. Графы состоят из узлов (вершин) и рёбер (связей), что позволяет GNN эффективно моделировать и анализировать структуру данных, содержащих сложные взаимосвязи между объектами.

**Ключевые особенности GNN.**

1 Адаптивность к структурированным данным: GNN способны обрабатывать данные, представленные в виде графов, такие как социальные сети, молекулярные структуры, транспортные сети и т. д.

2 Передача и агрегация информации: узлы в графе могут обмениваться информацией с соседними узлами, что позволяет сети "обучаться" и выявлять скрытые зависимости.

3 Обобщающая способность: GNN могут обобщать знания, извлечённые из обучающих данных, на новые, невиданные структуры, что делает их мощным инструментом для множества задач машинного обучения.

**Принцип работы GNN.** Основой работы GNN являются две основные операции: передача сообщений и агрегация. На каждом этапе (итерации) узлы графа передают и принимают сообщения от своих соседей, обновляя свои представления (векторные состояния) на основе полученной информации. Этот процесс повторяется до достижения сетью стабильного состояния.

**Этапы обучения GNN**:

1 Подготовка данных: сбор и предварительная обработка данных, которые будут представлены в виде графа. Узлы и рёбра графа должны быть определены и снабжены соответствующими признаками.

2 Инициализация параметров: установка начальных значений параметров модели, которые будут оптимизироваться в процессе обучения.

3 Форвардный проход: на каждом этапе обучения узлы графа передают и получают сообщения от соседей, обновляя свои векторные представления на основе полученной информации.

4 Вычисление функции потерь: оценка текущей производительности модели с использованием метрики, подходящей для решаемой задачи (например, кросс-энтропия для классификации).

5 Обратное распространение ошибки: вычисление градиентов функции потерь по параметрам модели для их последующего обновления.

6 Обновление параметров: корректировка параметров модели с использованием алгоритма оптимизации (например, стохастического градиентного спуска).

7 Повторение процесса: повторение этапов форвардного прохода, вычисления функции потерь, обратного распространения ошибки и обновления параметров до тех пор, пока модель не достигнет требуемого уровня производительности или не исчерпает заданное количество эпох.

Этот процесс позволяет графовым нейронным сетям адаптироваться к структуре и признакам графа, обучаясь на основе доступных данных и улучшая свою способность обобщать знания для новых, невиданных структур.

**Этапы работы GNN.**

1 Инициализация: каждый узел графа получает начальное представление, основанное на его начальных признаках.

2 Передача сообщений: узлы обмениваются информацией с соседями через рёбра графа.

3 Агрегация: каждый узел обновляет своё представление, агрегируя полученные от соседей сообщения.

4 Финальное представление: процесс передачи и агрегации повторяется несколько раз, после чего сеть выдаёт конечные представления узлов для окончательной оценки.

Поскольку модель GNN предсказывает наличие рёбер, её можно рассматривать как бинарный классификатор. Для оценки её качества будем использовать метрику ROC-AUC.

**Применение метрики ROC-AUC для GNN.** ROC-AUC (Receiver Operating Characteristic - Area Under Curve) [1] — это метрика, широко используемая для оценки качества бинарных классификаторов. Она основывается на построении кривой зависимости истинно положительных срабатываний (True Positive Rate) от ложно положительных (False Positive Rate) при различных порогах классификации. Пример кривой представлен на рисунке 5.

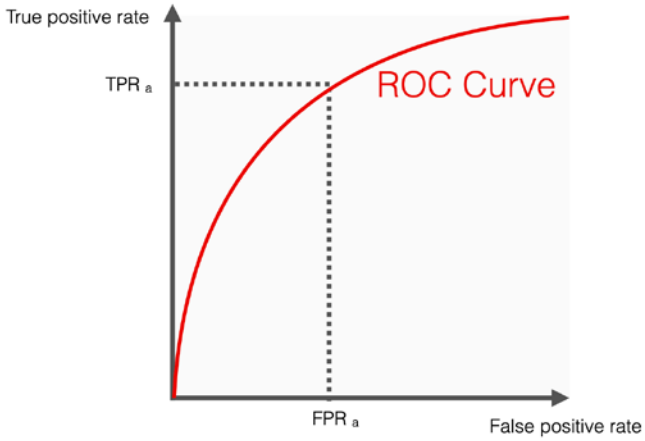



*Рисунок 5.* Пример ROC-AUC кривой

Применение ROC-AUC при работе с GNN:

1 Классификация узлов: в задачах, где GNN используется для классификации узлов графа (например, для определения категории узла в социальной сети или типологии молекулы), ROC-AUC позволяет оценить, насколько хорошо модель различает положительные и отрицательные классы узлов при различных порогах.

2 Классификация связей: в задачах, связанных с прогнозированием наличия или отсутствия связи между двумя узлами (например, для рекомендаций друзей или предсказания химических взаимодействий), ROC-AUC помогает понять, насколько

эффективно модель идентифицирует истинные положительные связи при минимальном количестве ложных срабатываний.

Использование метрики ROC-AUC для оценки GNN позволяет получить обобщённое представление о качестве модели вне зависимости от выбора конкретного порога классификации, что делает её особенно полезной при работе с несбалансированными данными, где классы могут быть представлены в разных пропорциях.

**Схема системы и возможные улучшения.** На рисунке 6 представлена базовая версия системы ранжирования.

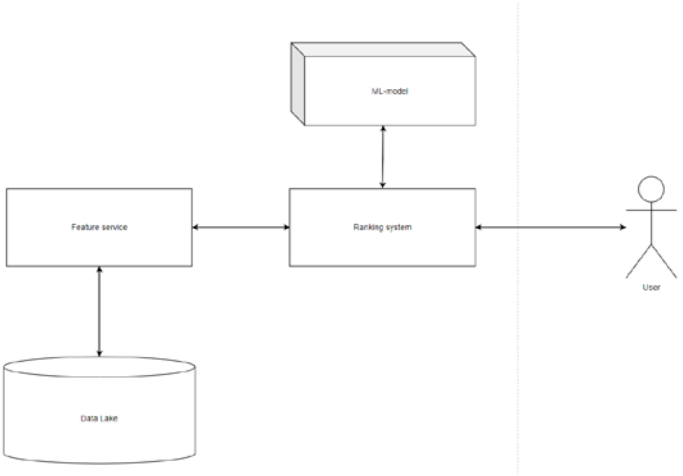



*Рисунок 6.* Принципиальная схема системы ранжирования

Принцип ее действия прост. Сервис ранжирования получает запрос, и собирает социальный вектор. Вектор отправляется на вход модели машинного обучения, которая, в свою очередь, отдает свои предсказания, которые передаются пользователю в качестве рекомендаций.

Однако подобная система не сможет справиться с большим количеством запросов, а количество узлов социального графа может насчитывать миллионы. На основании данных исследования [2], новые дружеские связи с вероятностью в 92% формируются через «друзей друзей». На основании данного факта можно заключить, что анализ следует проводить только в отношении «друзей друзей», что позволит значительно сократить количество узлов.

На рисунке 7 представлена базовая схема с добавлением сервиса определения «друзей друзей».

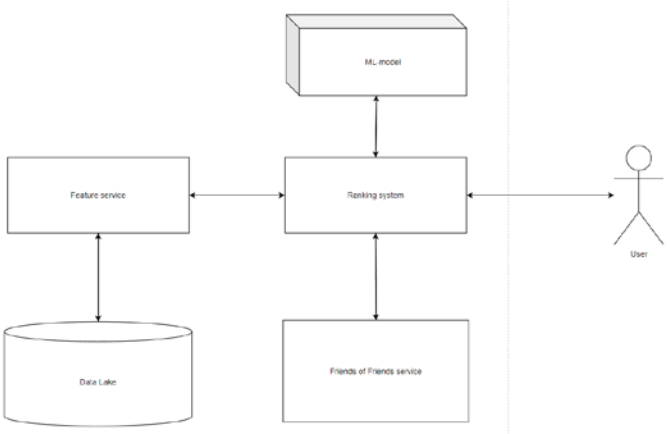



*Рисунок 7.* Принципиальная схема системы ранжирование с добавочным сервисом определения потенциальных друзей.

Дальнейшая оптимизация может быть направлена на отказ от работы системы по запросу. Для это достаточно разработать сервис предварительных расчетов, который будет обновлять предсказания в зависимости от приходящих изменений и выдавать пользователям заблаговременно подготовленные ответы. Принципиальная схема отображена на рисунке 8.

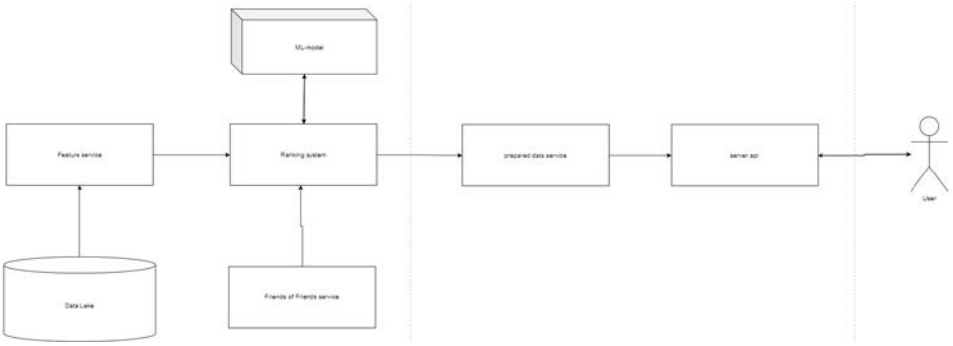


*Рисунок 8.* Принципиальная схема системы ранжирование с добавочным сервисом определения потенциальных друзей и предварительной генерацией рекомендация.

**Заключение**. Грамотно реализованная система, корректно подобранные признаки и рационально смоделированная модель машинного обучения позволяет добиться высоких результатов при выполнении предсказаний и предоставлении пользовательских рекомендаций. Уменьшение размерности графов способствует увеличению производительности системы. Использование системы ранжирования позволяет улучшить пользовательский опыт пользователя приложения социальной направленности, что, в свою очередь, благоприятно влияет на популярность системы.

## Список литературы

[1] Рашка С., Лю Ю., Мирджалили В. Машинное обучение с PyTorch и Scikit-Learn: учебное пособие. Москва: Издательство, 2023. – 600 с.

[2] интернет ресурс https://youtu.be/Xpx5RYNTQvg?t=1823

## Авторский вклад

**Бугаев Евгений Анатольевич** – анализ и описание системы ранжирования. Разработка схемы системы, практическое описание.

# USING THE RANKING METHOD TO DETERMINE A LIST OF POTENTIAL USER ACQUAINTANCES

***E. A. Bugaev***
*Software Engineer*
*Whitesnake*

**Annotation.** An analysis of the approach to planning the ranking system when determining the list of potential acquaintances is carried out. The main algorithm for solving this problem is described. A basic scheme of the ranking service has been developed, possible options for improving the system are shown. Optimization of the input and output data of the system was carried out. Assessed and optimized**.**

**Keywords:** ranking, machine learning model, graph, GNN, ROC-AUC**.**

УДК 37.013.75

# ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В ОБРАЗОВАНИИ

***Н.Ч.Реджепмурадова***
*Преподаватель кафедры информационных систем*
*Института телекоммуникаций и информатики Туркменистана*
nurjahanrejepmuradowa92@gmail.com

***Н.Ч.Реджепмурадова***

*Окончила Института телекоммуникаций и информатики Туркменистана. Область научных интересов сосредоточены в исследование и применение технологий искусственного интеллекта для улучшения образовательного процесса, персонализации обучения и автоматизации преподавания.*

**Аннотация.** Основной целью статьи является определить роль преподавателя в адаптивном обучении и возможность его замещения искусственным интеллектом. Теоретическая основа исследования базируется на комплексе научных положений ученых в области «Online" образования и искусственного интеллекта. Методы теоретического уровня использовались для анализа и обобщения существующих интеллектуальных систем адаптирования процесса обучения, а также методологии алгоритмов обработки данных, применяемых для этого.

**Ключевые слова:** базы данных, образовательный сервис, мультиагентный подход, геймификация обучения.

**Введение:** Личностные качества каждого индивидуальны, а соответственно, и степень усвоения материала разная. Создание и применение интеллектуальной адаптивной платформы позволит решить проблемы, связанные с индивидуализацией процесса образования. Интеллектуальная составляющая данных систем основывается на технологиях: Big Data, Data Mining, Learning Analytics. Их применение в образовании создает свои положительные аспекты:

1 **Для обучающихся** (использовать виртуального преподавателя для подбора материала, определять темп обучения (нагрузку), выбирать индивидуальную траекторию обучения, виртуально моделировать ситуацию);

2 **Для преподавателя** (аргументировать те или иные нововведения (выводы), выстраивать индивидуальную последовательность учебного плана и связанные с ним учебные навыки для каждого учащегося, искать скрытые закономерности, эффективнее использовать элементы групповой работы, грамотно выстраивать занятия, осуществлять прогноз успешности обучения и своевременное внесение изменения).

**Методология.** Когнитивные методы обучения (методы учебного понимания) – это приемы обучения, основанные на таких видах деятельности, которые позволяют личности познавать себя и окружающий мир. Суть когнитивной технологии в обучении состоит в обеспечивающей понимание студентами учебной информации. Основная цель исследования когнитивного обучения, заключается в развитии всех умственных способностей и стратегий, делающих возможным процесс обучения и адаптации к новым ситуациям. Акцент обучения должен сместиться в сторону ускорения обработки информации из-за его обилия, то есть, для перехода к обществу знаний, мы должны создать

такую систему образования, где в массовом порядке будут преподаваться когнитивные навыки, повышающие индивидуальные способности обработки информации.

Когнитивные технологии создают условия для понимания каждым студентом входящей информации. К примеру, формировать информационную компетентность учащихся, под которой понимается совокупность умений использовать информацию, поступающую из различных источников:

– **умение конспектировать**, т. е. излагать собственные мысли в соответствии с нормами языка и правилами логики;

– **умение аннотировать**, т. е. осуществлять краткую структурированную запись содержания книг, статей, выступлений, теле- видео- и других материалов;

– **умение осуществлять сбор информации по заданной проблеме**;

– **умение осуществлять сопоставление информации**, полученной из различных источников, по заданным критериям; – умение использовать технические средства получения информации;

– **умение использовать программные средства получения информации**.

В новом образовании должна быть создана такая методология обучения, которая позволит массово готовить специалистов, знающих себя и свои способности, что позволит ему учиться на протяжении всей своей жизни и тем самым непрерывно повышать свою квалификацию.Онлайн-обучение нисколько не улучшает понимание учениками информации, а лишь делает процесс обучения проще и доступней. Чтобы избежать проблем в понимании, необходимо подходить к каждому студенту индивидуально или же разделить их на группы на основе их социотипов.

Психика людей может быть представлена в виде 16 возможных вариантов восприятия и обработки информации, согласно соционической концепции. Поэтому необходимо понимать, что в голове у вашего собеседника может быть установлена совершенно другая система обработки информационных данных, отличающиеся от вашей, и соответственно в любое сказанное вами слово может быть вложен абсолютно противоположный смысл для ученика. Таким образом, основой электронного обучения в первую очередь должно быть не педагогика, а соционика.

Способности у всех к изучению чего-либо разный. Разный жизненный опыт, разные мотивации и уровень знаний. Поэтому кому-то из обучающихся программа всегда кажется слишком легкой, кому-то невыносимо трудной. Кому-то удобно воспринимать материал в одной форме, кому-то в другой. В результате, преподаватель всё время оказывается перед нерешаемой задачей: как обеспечить приемлемую успеваемость «слабых» учеников и не отбить желание учиться легким материалом у сильных?. Поэтому при подготовке к заданиям необходимо учитывать разный уровень когнитивного развития студентов и создавать задания соответствующего уровня сложности. В этом также заключается и принцип итогового контроля, в котором оценка знаний соответствует определенному уровню знаний [5, с. 49].

Новым инструментом учебного процесса уже в ближайшем будущем могут стать видеоигры. Геймификация и образовательные игры мотивируют на обучение, привносят дух соревновательности и дают освоить конкретные навыки. С помощью игр ученики с большим интересом осваивают новые области знания и получают возможность моделировать и испытывать на себе разные ситуации. Это помогает определиться с профессиональными склонностями и интересами. Игры, позволяющие создавать сложные структуры, помогают развитию креативности, а кооперативные, социальные и бизнес-игры способствуют коллективному решению задач и развитию эмоционального интеллекта.

На сегодняшний день применение искусственного интеллекта входит и в сферу образования. Создаются первые системы на основе данных технологий в области прокторинга, когда машины используя алгоритмы распознавания лиц, отслеживают сдачу студентами контрольных тестов и заданий при сдаче их в электронной среде. Одним из направлений в области образования является использование технологий искусственного интеллекта в адаптивном обучении, использование которого позволяет определить возможности формирования для каждого отдельно взятого студента индивидуальной образовательной траектории.

С точки зрения информационного подхода, искусственный интеллект – это область информационных технологий, основная задача которой разработка интеллектуальных компьютерных систем, способных решать задачи, которые традиционно решались с помощью человеческого разума, например, понимание языка, рассуждения, решение проблем. На сегодняшний день различают несколько направлений в развитии технологий искусственного интеллекта:

– нейронные сети – это математическая модель, а также ее аппаратная и программная реализация, построенная по образу и подобию биологических нейронных сетей;

– экспертные системы – это компьютерные системы, способные на основе технологий баз знаний и соответствующих алгоритмов заменить человека эксперта в различных областях;

– естественно-языковые системы – это автоматические информационные системы, в которых для общения с пользователем используется естественный язык;

– нечеткие множества – это базовое понятие нечеткой логики, в которой вместо двух значений ноль и единица рассматривается весь диапазон ;

– эволюционные методы и генетические алгоритмы –это эвристические алгоритмы, использующиеся для задач оптимизации, в основе которых лежат механизмы аналогичные естественному отбору в природе;

– системы извлечения знаний – это новые инновационные технологии извлечения и обработки больших объемов информации, полученные из глобальных компьютерных сетей.

Сфера образования является одним из перспективных направлений использования систем искусственного интеллекта. В рейтинге перспективных направлений, в которых за счет внедрения данных технологий можно получить наилучшие показатели увеличения факторной производительности, сфера образования входит в пятерку лидеров.

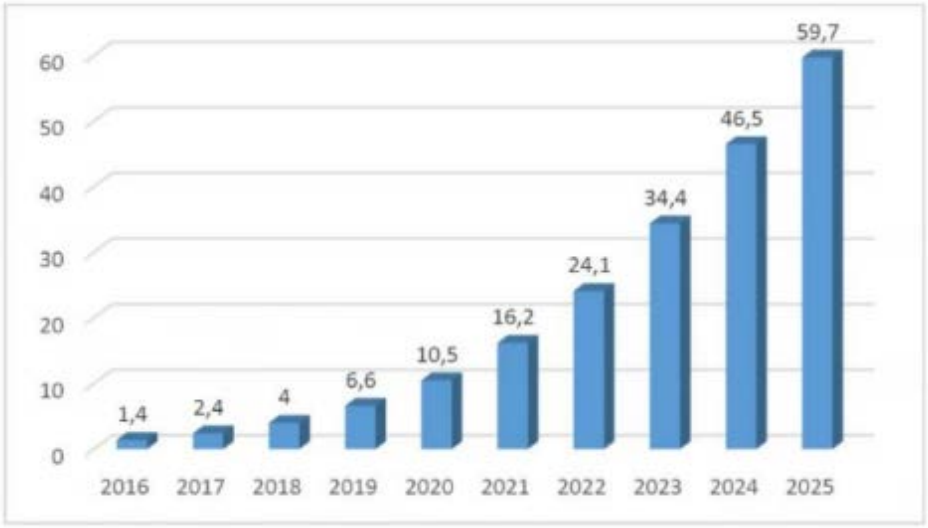



*Рисунок 1.* Динамика мирового рынка систем искусственного интеллекта в образовании.

Основным направлением применения искусственного интеллекта в сфере образования является прокторинг – это системы удаленного наблюдения за поведением испытуемого во время проведения проверочных испытаний .

**Обсуждение.** В основе искусственного интеллекта в образовании лежит основная цель сделать вычислительно точными и явными формы образовательных, психологических и социальных знаний, которые часто остаются неявными, то есть представить эти знания в формализованном виде, чтобы с помощью компьютерных программ проанализировать полученные результаты и на основе проведенного исследования получить соответствующую модель обучения. Одним из направлений искусственного интеллекта является возможность формирования адаптивных сред обучения, позволяющих на основе анализа определить индивидуальную траекторию обучаемого.

Адаптивные среды обучения – это цифровая среда обучения, адаптирующая подходы и материалы к обучению и изучению возможностей и потребностей отдельных учащихся.

Использование искусственного интеллекта в образовании должно базироваться на трех базовых подходах: описание процесса обучения (педагогическая модель); описание дисциплины (модель предметной области); описание студента (модель обучаемого)

Представим модель адаптивного обучения на основе технологий искусственного интеллекта, структура которой будет иметь следующий вид . В процессе обучения система обрабатывает процесс получения знаний в виде анализа достижений, ошибок, эмоционального состояния обучаемого. На основе проведенного сбора информации идет процесс анализа всех параметров процесса обучения, на основе которого происходит самообучение и усовершенствование самой системы.

На основе анализа большого объема информации система адаптивного обучения на основе технологий искусственного интеллекта способна обеспечить индивидуальную траекторию образования отдельного взятого студента по каждому предмету с учетом его возможностей и способностей.

Одним из основных направлений использования искусственного интеллекта в сфере образования являются возможности адаптивного обучения, позволяющего на основе трех базовых моделей: педагогической модели, модели обучаемого и модели предметной области, с использованием алгоритмов обработки знаний, предоставить адаптированный

контент для каждого отдельно взятого студента, за счет обучения системы искусственного интеллекта на основе анализа и обработки больших объемов информации, что позволит выстроить индивидуализированную траекторию обучаемого на основе его способностей и возможностей как для изучения отдельно взятого предмета, так и для процесса формирования специалиста в той или иной области, тем самым отойти от усредненного подхода в обучении, который существует на сегодняшний день и не позволяет раскрыть все возможности обучаемого.

К важным понятиям, связанным с адаптивной технологией обучения, относят:

– адаптирующее образование – программы и курсы, ориентированные на формирование комплекса способностей, необходимых для того, чтобы индивид мог комфортно жить и эффективно действовать в новых для него социальных условиях;

– адаптивная образовательная среда- система, способная каждому обучающемуся помочь достичь оптимального уровня интеллектуального развития в соответствии с его природными задатками и способностями;

– адаптивная образовательная модель – модернизированная массовая общеобразовательная школа для учащихся со смешанными способностями.

В ряде публикаций под «электронным обучением» понимают адаптивное (индивидуализированное) обучение, реализованное на базе всестороннего применения мульти- и гипермедиа, удалённого доступа к распределённым образовательным ресурсам на основе веб-технологий, с автоматизированным контролем и анализом результатов обучения и широким использованием разнообразных сетевых средств взаимодействия обучающихся между собой и с преподавателем.

Под адаптивностью в обучении в современной научной трактовке понимают персонификацию процесса обучения на основе создания электронных курсов, учитывающих индивидуальные особенности обучаемых, в том числе психологические особенности, восприятие, уровень начальных знаний, а также индивидуальные цели и задачи обучения. Адаптация рассматривается также как адаптивное планирование (статическая адаптация) и адаптивное взаимодействие (динамическая адаптация). Адаптивное планирование в вузе на этапе подготовки индивидуализированного учебнометодического материала позволяет реализовать адаптацию, как в группе студентов, так и к отдельному студенту, обучающемуся по индивидуальному плану. В процессе информационного взаимодействия при динамической адаптации осуществляется как изменение содержание и способов представления учебно-методических материалов, так и всесторонняя настройка системы для конкретного студента.

Адаптивное обучение в вузе обладает рядом преимуществ по сравнению с классическими методиками обучения:

– снижение нагрузки на студента и уменьшение общего времени обучения за счёт отображения только действительно необходимой студенту информации;

– возможность реализации более удобных средств контроля над действиями студента.

Адаптация достигается на основе использования модели обучаемого, составляющими которой являются цели обучения, начальные знания в области обучения, индивидуальные особенности обучаемого. Реализация моделей адаптивного обучения базируется на современных подходах программированного обучения на основе пакета прикладных программ, метод экспертных оценок, мультиагентный подход, создание комплексных систем обучения. Программированное обучение в электронных учебно-методических материалах на основе использования пакета прикладных программ осуществляется по принципу разделения библиотеки стандартных программ и программ, управляющих ресурсами машины и собственно библиотекой. Для воздействия и

пользователя с системой используется диалоговый компонент со специальным входным языком, позволяющим давать чёткие команды вызова обучающей системе.

Метод экспертных систем позволяет реализовывать возможность не закладывать априори последовательность шагов обучения, поскольку она строится самой системой в процессе её функционирования, что позволяет формировать для каждого обучаемого индивидуальный план обучения. Данные обучающие системы способны выполнять параметрическую и структурную адаптации.

Мультиагентный подход позволяет реализовать возможность адаптации всех уровней (параметрическая, структурная, адаптация объекта управления, адаптация целей), что позволяет обеспечить управление объектом- обучаемым на всех этапах процесса обучения. Основа этого подхода- построение системы как совокупности агентов. Каждый из агентов имеет семантическое описание своего поля деятельности (свою структуру, свои знания), и соответствует экспертной системе с традиционной структурой. Агент обладает всеми свойствами экспертных систем, а также памятью своей деятельности, имеет собственные ресурсы для достижения собственных целей, взаимодействия с другими агентами и разрешения конфликтов с целями других агентов для достижения общих целей. Это позволяет свободно выбирать те цели, которые преследуются на данный момент объектом управления, и соответственно целям выбирать тот эталон (представленный соответствующим агентом), соответствие которому достигается моделью обучаемого на данный момент. Таким образом, для каждой конкретной задачи обучения составляется определённый коллектив агентов, что говорит о смене структуры и целей решающей системы в зависимости от поставленной задачи.

Важную роль в развитии современных высокоинтеллектуальных образовательных технологий играют также системы адаптивной гипермедиа, которые применяют различные виды моделей пользователя для приспосабливания под его индивидуальные характеристики содержимого и ссылок страниц гипермедиа.

Выделяются две технологии в адаптивной гипермедиа:

– адаптивное представление (поддержка посредством создания (не)видимых ссылок ориентации и навигации обучающегося в гиперпространстве),

– адаптивная поддержка в навигации (приспособление содержимого страницы гипермедиа к целям, знаниям пользователя, а также к другой информации, хранящейся в модели пользователя).

В настоящее время проводятся исследования, направленные на эволюцию базовых моделей, отражавших традиционный квалификационный подход к организации процесса, в сторону компетентностного подхода, учитывающего способности обучаемых применять знания, умения и личностные качества для успешной профессиональной деятельности. В связи с этим особую актуальность приобретают вопросы, связанные с построением «психологического портрета личности», а также с эффективным учётом личностных характеристик студента при построении моделей обучаемого и модели обучения, поскольку это напрямую связано с реализацией компетенций, определяющих, по сути, набор видов деятельности, которые должен осуществлять обучаемый в будущем.

При построении базовых эвристических моделей, в частности модели обучаемого, предусмотрен учёт влияния личностных характеристик на адаптивную модель обучения. Суть этого процесса заключается в динамической модификации стратегии обучения в соответствии с текущей

моделью обучаемого, то есть осуществляется подбор и упорядочивание наиболее эффективных на данном этапе обучающих воздействий с учётом выявленных личностных характеристик.

**Заключение.** Таким образом, интеллектуально ориентированные образовательные технологии раскрывают новые пути повышения качества образовательных услуг, оказываемых вузом в условиях информатизации современного общества за счёт обеспечения индивидуального подхода к студенту, поддержки в решении задачи интеллектуального анализа решений с интерактивной обратной связью. Использование указанных технологий значительно экономит и рационализирует время преподавателя, усиливает управленческие и коммуникативные аспекты учебного процесса за счёт технологии оптимального подбора моделей обучаемых.

## Список литературы

[1] Овездурдыева И, Мырадов М, Технологии работы с большими данными “big data”: сбор, хранение и обработка больших данных Научное издательство «Наука и мировоззрение» 2024 70-74 с

[2] Мурадов М Назарова С, Современные технологии распознавания речи Научное издательство «Наука и мировоззрение» 2024 301-305 с

[3] Мурадов М., Распознавание речи Сборник статей Международной научно-практической конференции Научный потенциал - 2023 Петрозаводск 2023 405-409 с

## Авторский вклад

**Реджепмурадова Нурджахан Чарымырадовна -** В данной работе проведено исследование современных подходов к использованию искусственного интеллекта (ИИ) в образовании, а также проанализированы его преимущества и вызовы. Автором были рассмотрены ключевые технологии, такие как адаптивное обучение, интеллектуальные системы автоматизированного оценивания знаний.

# ARTIFICIAL INTELLIGENCE IN EDUCATION

***N.Ch.Rejepmuradova***
*Lecturer of the Department of Information Systems of the Institute of Telecommunications and Informatics of Turkmenistan*

**Abstract:** The main purpose of the article is to determine the role of the teacher in adaptive learning and the possibility of replacing it with artificial intelligence. The theoretical basis of the research is based on a complex of scientific positions of scientists in the field of "Online" education and artificial intelligence. Theoretical methods were used to analyze and generalize existing intelligent systems for adapting the learning process, as well as the methodology of data processing algorithms used for this.

**Keywords:** databases, educational service, multi-agent approach, learning gamification.

УДК 612.845.552: 004.932

# ТЕХНОЛОГИИ БОЛЬШИХ ДАННЫХ В ОБРАБОТКЕ ИЗОБРАЖЕНИЙ ДЛЯ ЛЮДЕЙ С АХРОМАТОПСИЕЙ

***В.В. Синицына***
*Аспирант кафедры инженерной психологии и эргономики БГУИР, магистр*
vladasinitsina1@gmail.com

***В.В. Синицына***
*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой алгоритмов реколоризации изображений для людей с аномалиями цветового зрения.*

**Аннотация.** На сегодняшний день существует довольно небольшое количество методов, преобразующих цветовую информацию изображений в соответствии с корректным ее восприятием людьми с ахроматопсией. Однако проблема ахроматопического зрения заключается не только в невозможности отличать цвета, которые нормальный трихромат воспринимает различно, но и в преобразовании в данных методах всей цветовой составляющей изображения, в то время как часто возникает необходимость преобразовывать лишь некоторые области изображения, которые являются наиболее важными для ахроматопов с точки зрения ценности визуальной информации. Именно для решения данных задач предложен метод препроцессинга изображений с целью их последующей реколоризации для людей с ахроматопсией, в котором важным этапом является обработка больших данных.

**Ключевые слова:** ахроматопсия, реколоризация, алгоритмы, препроцессинг, большие данные.

**Введение.** Как известно, аномалии цветового зрения являются достаточно распространенными на земном шаре, так как наблюдаются у 8% мужского и 0,5% женского населения. Выделяют несколько видов данных аномалий, среди которых наиболее распространенным считается аномальная трихромазия, встречающаяся примерно у 6% населения Земли, однако менее распространенным и в то же время наиболее тяжелым видом аномалий является ахроматопсия.

Ахроматопсия встречается только у 1 из 30 000 человек на земном шаре, а также бывает полной, при которой наблюдается абсолютное отсутствие восприятия цвета и такие сопутствующие заболевания, как нистагм и светобоязнь, и неполной, которая характеризуется небольшим цветоразличением и либо отсутствием нистагма и светобоязни, либо их присутствием в более легкой форме [1]. Наиболее распространенной является полная ахроматопсия.

Кроме того, ахроматопсия встречается как врожденная, так и приобретенная. Приобретенная ахроматопсия именуется церебральной (или центральной), связана она с повреждением коры головного мозга [2]. Стоит также отметить, что на сегодняшний день ахроматопсия не поддается лечению, в связи с чем актуальными являются исследования и проектирование различных методов и алгоритмов для помощи данным цветоаномалам с восприятием цветовой информации объектов.

Ахроматопсия представляет собой такую аномалию цветового зрения, при которой наблюдается нарушение работы колбочковой системы, но корректное функционирование палочек – клеток-рецепторов сетчатки, чувствительных к силе света и действующих преимущественно при сумеречном или ночном зрении – в глазах человека. В связи с этим ахроматопы воспринимают мир

исключительно в оттенках серого. Важно отметить, что именно благодаря палочкам люди с ахроматопсией хорошо ориентируются в пространстве в темное время суток.

Основное ограничение ахроматопического зрения заключается в невозможности отличать объекты, оригинальные цвета которых различны, а преобразованные в оттенки серого – неразличимы. Вследствие наличия у ахроматопов таких сопутствующих заболеваний, как светобоязнь, нистагм, а также зачастую гиперметропия или миопия, люди с ахроматопсией обычно используют солнцезащитные очки, а также очки для коррекции гиперметропии или же миопии.

Таким образом, цель работы – определить, каким образом следует осуществить предварительную обработку изображений для людей с ахроматопсией с целью последующей реколоризации данных изображений.

В качестве задач следует выделить следующие:

– анализ существующих на сегодняшний день методов и средств реколоризации изображений для людей с ахроматопсией;

– формулировка и подробное описание препроцессинга изображений до применения к ним алгоритмов или методов реколоризации;

– анализ полученных в предыдущих пунктах результатов и подведение итогов.

**Анализ существующих методов и средств реколоризации изображений для людей с ахроматопсией.** В литературе упоминаются два основных метода, которые предназначены для помощи людям с ахроматопсией в корректном восприятии визуальной информации.

В методе Rasche et al.’s [3] предполагается, что воспринимаемая разница между любыми двумя цветами пропорциональна их соответствующей воспринимаемой разнице в оттенках серого цвета. Так, происходит минимизация функции ошибок как гарантия того, что неразличимые цвета будут иметь различные значения яркости. Метод Rasche et al.’s [4] дополняет предыдущий, используя целевую константу пропорциональности *K*. В данном методе цель заключается в сохранении постоянства яркости и контраста, где контраст не меняет своего значения благодаря сохранению относительного расстояния между каждой парой цветов. Кроме того, в Zhu et al.’s исследовании [5] указано, что для решения проблемы восприятия визуальной информации монохроматами также предлагается использование методов преобразования изображений в оттенки серого для черно-белой печати.

В качестве примера существующих средств реколоризации изображений для ахроматопов также стоит отметить такую имеющуюся во многих мобильных устройствах функцию цветовой коррекции, как «Оттенки серого», которая, как указано в настройках мобильных устройств, помогает сделать отчетливее цвета, которые пользователю сложно различать. Однако данная функция изменяет все цвета на экранах мобильных устройств, не позволяя никоим образом пользователю выбрать некоторую конкретную область для реколоризации.

Так, на основании анализа существующих на сегодняшний день методов и средств реколоризации изображений для людей с ахроматопсией важно отметить, что в данных методах и средствах не предусмотрена возможность предварительной обработки изображений таким образом, чтобы реколоризация происходила только для наиболее значимых для ахроматопов областей изображений.

**Предварительная обработка изображений с целью их последующей реколоризации.** Стоит отметить, что реколоризация лишь выбранных областей изображения очень важна для людей с любым видом аномалий, так как не всегда реколоризация всего экрана бывает удачной, преобразуя довольно часто некорректно цвета, которые пользователь до реколоризации воспринимал верно, а также в связи с особым вниманием цветоаномала к определенным объектам на изображении. Кроме того, необходимо упомянуть о том, что благодаря препроцессингу и последующей реколоризации изображений многие люди с ахроматопсией смогут корректировать цвета в рамках своих бытовых потребностей.

Так, из вышеупомянутых утверждений следует, что в связи с необходимостью различения объектов в той или иной области изображения следует выделять на экране наиболее важные объекты для реколоризации, а только затем осуществлять преобразование цветов определенным образом согласно некоторому выбранному алгоритму или методу.

Предварительная обработка изображений для людей с ахроматопсией включает следующие шаги:

1 Выбор ахроматопом некоторых областей на изображении, в пределах которых расположена наиболее актуальная для человека с данным видом цветоаномалии визуальная информация. Так, выбор областей может быть осуществлен на экранах устройств как выделением курсором компьютерной мыши, так и при помощи пальцев посредством заключения необходимых объектов в нарисованные замкнутые области.

2 Далее, координаты каждой точки некоторых выбранных пользователем замкнутых областей направляются в алгоритм или же метод реколоризации, так как последующие преобразования следует осуществлять лишь в пределах данных областей.

3 При помощи некоторого алгоритма или же метода реколоризации происходит преобразование цветов лишь в выбранных областях изображения с наиболее важными для пользователя объектами.

Так, в предлагаемом методе предварительной обработки изображений выделяются некоторые области, в пределах которых расположены объекты, представляющие для человека с ахроматопсией наибольший интерес с точки зрения имеющейся в них визуальной информации. Количество областей и цветов в рамках этих областей зачастую будет велико, что потребует последовательной обработки каждой выделенной области, а также присутствующих в каждой такой области цветов. Кроме того, данную обработку изображений допустимо применять на большом количестве изображений с целью последующей реколоризации множества областей и цветов в рамках данных областей.

**Заключение.** Таким образом, для возможности реколоризации только наиболее актуальных с точки зрения ценности визуальной информации областей на изображениях для людей с ахроматопическим зрением был предложен метод предварительной обработки изображений, за которым последует дальнейшая обработка цветовой информации таких областей некоторым выбранным методом или же алгоритмом преобразования цветовой составляющей. Так, данный препроцессинг не только облегчит реколоризацию, так как выбором областей и расположенных в рамках данных областей цветов ограничит количество нуждающихся в реколоризации имеющихся у выбранных объектов цветов, но также позволит людям с ахроматопсией лучше справляться со многими повседневными задачами, в которых зачастую концентрация внимания происходит лишь на некоторых наиболее значимых для человека с ахроматопическим зрением объектах.

## Список литературы

[1] Hirji N., Aboshiha J., Georgiou M., Bainbridge J., Michaelides M. Achromatopsia: clinical features, molecular genetics, animal models and therapeutic options. Ophthalmic Genetics. 2018;39(2):149-157. DOI: https://doi.org/10.1080/13816810.2017.1418389.

[2] Bartolomeo P., Bachoud-Levi A.-C., Schotten M. T. The anatomy of cerebral achromatopsia: a reappraisal and comparison of two case reports. Cortex. 2014;56:138-144. DOI: https://doi.org/10.1016/j.cortex.2013.01.013.

[3] Rasche K., Geist R., Westall J. Detail preserving reproduction of color images for monochromats and dichromats. IEEE Computer Graphics and Applications. 2005;25(3):22-30. DOI: https://doi.org/10.1109/MCG.2005.54.

[4] Rasche K., Geist R., Westall J. Re-coloring images for gamuts of lower dimension. Computer Graphics Forum. 2005;24(3):423-432. DOI: https://doi.org/10.1111/j.1467-8659.2005.00867.x.

[5] Zhu Z., Mao X. Image Recoloring for Color Vision Deficiency Compensation: a Survey. The Visual Computer. 2021;37:2999-3018. DOI: https://doi.org/10.1007/s00371-021-02240-0.

## Авторский вклад

**Синицына Влада Владиславовна –** постановка задач исследования, описание особенностей ахроматопсии, анализ существующих методов и средств реколоризации изображений для ахроматопов, разработка алгоритма предварительной обработки изображений для их последующей реколоризации, подведение итогов, формирование структуры статьи.

# BIG DATA TECHNOLOGIES IN IMAGE PROCESSING FOR PEOPLE WITH ACHROMATOPSIA

***V.V. Sinitsyna***
*Postgraduate student of the*
*Department of Engineering*
*Psychology and Ergonomics*
*BSUIR, Master of engineering*

**Abstract.** Today, there are quite a few methods that transform the color information of images in accordance with its correct perception by people with achromatopsia. However, the problem of achromatic vision is not only the inability to distinguish colors that a normal trichromat perceives differently, but also the transformation of the entire color component of the image in these methods, while it is often necessary to transform only some areas of the image that are most important for achromatopes in terms of the value of visual information. It is to solve these problems that a method of image preprocessing for their subsequent recoloring for people with achromatopsia has been proposed, in which an important stage is the processing of big data.

**Keywords**: achromatopsia, recoloring, algorithms, preprocessing, big data.

УДК 004.522

# АРХИТЕКТУРА TRANSFORMER ДЛЯ ПРЕОБРАЗОВАНИЯ ГОЛОСА В ТЕКСТ

***М. Т. Мырадов***
*Заведующий кафедры*
*«Информационные системы»*
*Институт Телекоммуникаций и информатики Туркменистана*
maksat.myradow.92@mail.ru

***М.Т. Мырадов***
*Окончил Институт телекоммуникаций и информатики Туркменистана. Область научных интересов распознавания речи, искусственный интеллект, защита данных*

**Аннотация:** Архитектура Transformer успешно применяется для преобразования голоса в текст благодаря своей способности эффективно обрабатывать последовательности данных. Модель использует механизмы внимания (self-attention) для выявления зависимости между элементами входных данных, что повышает точность распознавания речи. Такой подход позволяет системе обучаться на больших объемах аудиоданных и добиваться высокой производительности в реальных приложениях.

**Ключевые слова:** Архитектура Transformer, преобразование речи в текст, внимание, распознавание речи.

**Введение.** Архитектура Transformer была впервые представлена компанией Google в 2017 году в статье под названием «Attention Is All You Need». Эта модель эффективно используется в различных областях языка, включая системы автоматического распознавания речи (ASR). **“Attention Is All You Need”.** Доминирующие модели последовательной трансдукции основаны на сложных рекуррентных или сверточных нейронных сетях в конфигурации кодер-декодер. Наиболее эффективные модели также связывают кодер и декодер через механизм внимания. Мы предлагаем новую простую сетевую архитектуру, Transformer, основанную исключительно на механизмах внимания, полностью обходясь без рекуррентности и сверток. Эксперименты с двумя задачами машинного перевода показывают, что эти модели превосходят по качеству, будучи более параллелизуемыми и требуя значительно меньше времени на обучение. Наша модель достигает 28,4 BLEU в задаче перевода с английского на немецкий язык WMT 2014, улучшая существующие лучшие результаты, включая ансамбли, более чем на 2 BLEU. В задаче перевода с английского на французский язык WMT 2014 наша модель устанавливает новый современный показатель BLEU для одной модели в 41,8 после обучения в течение 3,5 дней на восьми графических процессорах, что составляет небольшую часть затрат на обучение лучших моделей из литературы. Мы показываем, что Transformer хорошо обобщается и на другие задачи, успешно применяя его для анализа английской электоральной аудитории как с большими, так и с ограниченными обучающими данными**.**

Основы модели трансформатора

Модель представляет собой высокоскоростную и универсальную модель, разработанную на основе архитектуры Transformer. Трансформатор состоит из кодера и декодера. Кодер обрабатывает входные данные (мелодию голоса, дикцию и структуру слова), а декодер используется для интерпретации этих данных.

**Модель:**

1 Кодер: спектрально кодирует входные аудиоданные. Эта информация воспринимается как имеющая такие характеристики, как скорость звука, высота звука и то, как она меняется со временем.

2 Декодер: декодирует закодированную информацию и преобразует этот звук в текст. Основная функция декодера — дать интерпретацию относительно структуры языка и подлинности слов.

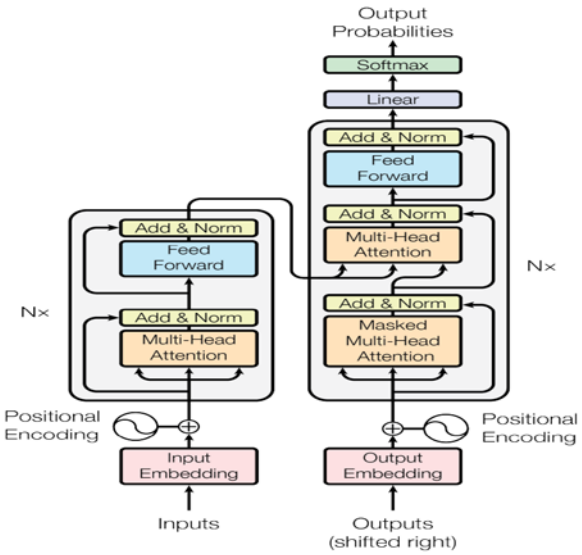



*Рисунок 1.* Как работают Transformer

**Подготовка вашего голоса для индивидуальных моделей:** Модель генерирует спектрограммы звуков, используя входные данные. Спектрограммы — это графическое представление звуковой информации, разделенной по частоте и времени. Эти данные служат исходными входными данными для кодировщика.

**Технические специалисты по архитектуре:**

1 Использование спектрограммы. Входной аудиофайл (MP3, WAV) преобразуется в спектрограмму. Эти данные обрабатываются с частотой 16 кГц. Архитектура модели позволяет быстро и с высокой точностью улавливать основную информацию о звуке при использовании спектрограммы.

2 Слои трансформатора. Число слоев преобразователя в кодере и декодере составляет от 12 до 24, что позволяет выполнять многослойные преобразования. Каждый слой:

- проверяет связь между данными;
- преобразует звуки в текст с учетом структуры языка.

**Размеры на разных компьютерах.** Производительность модели различается на разных компьютерных платформах. Ключевую роль играют такие характеристики, как мощный процессор, память графического процессора и объем оперативной памяти используемой платформы.

**Тесты и измерения.** Работоспособность модели проверена в различных конфигурациях. Результаты этих испытаний показаны ниже:

Таблица 1. Работоспособность платформ

| **Платформа** | **Время обработки** | **Использование** | **Использование** |
|---|---|---|---|

| | **данных** | **графического процессора** | **оперативной памяти** |
|---|---|---|---|
| Intel i5, 8GB RAM | 50 **секунд** | нет | 4 GB |
| NVIDIA RTX 3060, 32GB RAM | 2 секунд | да | 5 GB |

**Анализ:**

1 Стандартные процессоры (ЦП): модель работает с процессорами Intel и AMD, но время работы без ГП увеличивается.

2 При использовании графического процессора: модель работает очень быстро с графическими процессорами, такими как NVIDIA RTX 3060. Генерация спектрограммы и декодирование текста ускоряются с помощью графического процессора.

3 Влияние памяти (ОЗУ): устройства с большим объемом оперативной памяти повышают способность модели обрабатывать большие аудиофайлы.

**Практическое использование архитектуры:** Модель может использоваться во многих приложениях:

1 Офлайн-перевод: предоставление услуг перевода для людей, говорящих на разных языках.

2 Анализ голоса: позволяет оператору проверить громкость и использование голоса.

3 Регулирование информации: Управление форматом аудиозаписей.

**Оптимизация** Теперь давайте протестируем модель, изменив некоторые параметры на этапе оптимизации.

Первоначально размер пакета каждого сеанса обучения составлял 24, теперь: мы изменили размер пакета до 16, скорость обучения ранее была 0,001, теперь мы изменили ее на 1e-5, количество шагов обучения было 500, теперь мы увеличили его до 1000, и мы перешли на технологию токенизации OpenAI. Эти обновления оказались особенно важными для повышения точности обучения модели.

**Методология:** Размер партии: установите 16. Такой размер способствует обработке данных и эффективному использованию памяти графического процессора. Скорость обучения: использовалась низкая скорость 1e-5. Такая скорость снижает опасные колебания при обучении модели с более сложными данными и обеспечивает стабильную сходимость. Шаги: количество шагов увеличено до 1000. Это позволило модели взаимодействовать с большим количеством данных, что позволило ей достичь эффективного уровня обучения.

Токенизатор: переключен на систему токенизаторов OpenAI. Этот токенизатор характеризуется лучшим сохранением контекста данных.

Поскольку архитектура Lexical Analysis Transformer предназначена для обработки прямых цифровых данных, между текстом и токенами должен быть процесс перевода. Токен – целое число, представляющее символ или короткий сегмент символов. На входной стороне входной текст разбивается на последовательность токенов. Между тем, на стороне вывода выходные токены преобразуются обратно в текст. Модуль, который преобразует текст в последовательности токенов, называется «токенизатор».

Все токены образуют словарную базу токенизатора, а ее размер называется размером словаря. Для токенов, находящихся за пределами словаря, обычно используется специальный токен, который часто записывается как «[UNK]» (неизвестно).

Встраивание слов Каждый токен преобразуется в вектор встраивания путем поиска его в специальной таблице. Другими словами, однократное представление токена умножается на матрицу вложения M. Например, если токен доступа равен 3, то горячая форма будет иметь вид:

[0, 0, 0, 1, 0, 0, …]

Вектор внедрения становится следующим:

Встроить(3) = [0, 0, 0, 1, 0, 0, …] M

Векторы внедрения токенов добавляются к соответствующим им векторам кодирования позиций, и они образуют последовательность входных векторов.

Число измерений векторов внедрения называется «скрытым размером» или «измерением внедрения».

Отсоединение Слой извлечения представляет собой процесс, обратный слою внедрения. Слой внедрения преобразует токен в вектор, а слой извлечения преобразует вектор в распределение вероятностей между токенами.

Слой деинсталляции состоит из линейно-мягкого слоя:

UnEmbed(x) = softmax(xW + b)

**Заключение** Архитектура обеспечивает высокоуровневые возможности для эффективного, быстрого и точного преобразования аудиоданных в текст. Эта система в сочетании с моделью трансформатора может быть полезна не только для технологии распознавания голоса, но и в системах изучения языка и информационных системах.

При тестировании производительности Whisper на различных компьютерных системах мы обнаружили, что лучшие результаты достигаются при использовании графических процессоров. Это показывает важность графических процессоров в отрасли информационных технологий.

Для того чтобы архитектура, используемая для преобразования речи в текст, была успешной, особое внимание необходимо уделить конфигурации технологии.

## Список литературы

[1] Чарыев А, Информационные системы и технологии Научное издательство «Символ науки» 2024 13-16 с

[2] Мурадов М Назарова С, Современные технологии распознавания речи Научное издательство «Наука и мировоззрение» 2024 301-305 с

[3] Мурадов М., Распознавание речи Сборник статей Международной научно-практической конференции Научный потенциал - 2023 Петрозаводск 2023 405-409 с

## Авторский вклад

**Мырадов Максат Тачмухаммедович -** авторский вклад в архитектуру Transformer для преобразования голоса в текст, разработав, например, новый метод адаптации механизма внимания, оптимизирующий процесс обучения или интегрировав инновационные слои для улучшения обработки аудиоданных и повышения точности распознавания речи.

# TRANSFORMER ARCHITECTURE FOR VOICE-TO-TEXT CONVERSION

***M. T. Myradov***
*Head of the Department of*
*Information Systems,*
*The Institute of Telecommunications and*
*Informatics*

**Abstract:** The Transformer architecture is successfully applied to speech-to-text conversion due to its ability to efficiently process data sequences. The model uses self-attention mechanisms to identify dependencies between input data elements, which improves speech recognition accuracy. This approach allows the system to be trained on large volumes of audio data and achieve high performance in real-world applications.

**Keywords:** Transformer architecture, speech-to-text conversion, attention, speech recognition.

УДК 004.8

# BIG DATA И ПРОДВИНУТЫЙ АНАЛИЗ ДАННЫХ: СОВРЕМЕННЫЕ ТЕНДЕНЦИИ, ТЕХНОЛОГИИ И ПЕРСПЕКТИВЫ РАЗВИТИЯ

***П.С. Мырадов***
*Преподаватель*
*Государственного*
*энергетического*
*института*
*Туркменистана*
pvm87818@gmail.com

***П.С. Мырадов***
*Окончил институт Телекоммуникации и информатики Туркменистана. Область научных интересов связана с изучением информационных технологий, в Государственном энергетическом институте Туркменистана.*

**Аннотация.** В статье подробно рассматриваются ключевые концепции, методы и технологии, лежащие в основе обработки больших данных (Big Data) и продвинутого анализа данных (Advanced Analytics). Описаны современные архитектуры, алгоритмы машинного обучения, методы обработки естественного языка, статистические подходы, а также роль облачных вычислений и интернета вещей (IoT). Рассмотрены примеры практического применения в бизнесе, здравоохранении, производстве и образовании, а также обозначены основные вызовы и направления дальнейших исследований.

**Ключевые слова:** Big Data, продвинутый анализ данных, машинное обучение, искусственный интеллект, статистический анализ, облачные вычисления.

**Введение.** В современном мире наблюдается экспоненциальный рост объемов генерируемой информации благодаря цифровизации, появлению социальных сетей, сенсорных устройств и интернета вещей. Организации сталкиваются с необходимостью обработки и анализа как структурированных, так и неструктурированных данных, что стимулирует развитие новых технологий для работы с «большими данными». Продвинутый анализ данных позволяет не только извлекать закономерности и предсказывать тренды, но и принимать обоснованные решения на основе комплексного анализа информации.

**Определение и характеристики.** Термин Big Data охватывает огромные объемы информации, которые характеризуются следующими параметрами:

– объём (Volume): масштаб данных, которые могут исчисляться петабайтами и эксабайтами;

– скорость (Velocity): скорость поступления, обработки и анализа данных в режиме реального времени;

– разнообразие (Variety): широкий спектр форматов — от традиционных табличных данных до мультимедийных файлов и логов;

– достоверность (Veracity): качество и точность данных, что является критически важным для построения надежных моделей.

**Архитектуры и инструменты.** Современные системы обработки Big Data базируются на распределённых вычислительных платформах, таких как:

– Hadoop: экосистема для распределённого хранения и параллельной обработки данных;

– Apache Spark: платформа для быстрого анализа больших объемов данных с поддержкой потоковой обработки;

– NoSQL базы данных: оптимизированные для работы с неструктурированными данными, например, MongoDB, Cassandra и HBase.

**Машинное обучение и глубокое обучение.** Методы машинного обучения (ML) и глубокого обучения (DL) позволяют выявлять сложные нелинейные зависимости в данных. Основные подходы включают:

– классификацию: алгоритмы, такие как решающие деревья, метод опорных векторов (SVM) и логистическая регрессия, используются для разделения данных на категории;

– кластеризацию: методы, такие как алгоритм k-средних и иерархическая кластеризация, позволяют группировать данные по схожим характеристикам;

– нейронные сети: сверточные (CNN) и рекуррентные нейронные сети (RNN) применяются для анализа изображений, речи и временных рядов.

**Статистический анализ и предиктивное моделирование.** Статистические методы остаются важными для интерпретации данных:

– регрессионный анализ: используется для определения зависимости между переменными и построения предиктивных моделей;

– факторный и кластерный анализ: помогают выявлять скрытые структуры в данных.

– байесовские модели: позволяют учитывать неопределенности и обновлять прогнозы по мере поступления новой информации.

**Обработка естественного языка (NLP).** Методы NLP применяются для анализа текстовых данных:

– анализ тональности: выявление эмоциональной окраски текстов (например, отзывы клиентов или социальные медиа);

– тематическое моделирование: автоматическое выделение тем в больших коллекциях документов.

**Роль облачных вычислений и интернета вещей (IoT).** Облачные технологии и IoT играют ключевую роль в сборе, хранении и обработке данных:

– облачные платформы: предоставляют масштабируемую инфраструктуру для хранения и анализа данных (например, AWS, Google Cloud, Microsoft Azure);

– интернет вещей: генерирует огромные объемы данных с устройств, что требует эффективных методов обработки и анализа в реальном времени.

**Бизнес и финансы**

– маркетинг и клиентская аналитика: персонализация предложений и оптимизация рекламных кампаний;

– управление цепочками поставок: оптимизация логистических процессов и прогнозирование спроса;

– финансовый анализ: выявление мошеннических транзакций, оценка кредитного риска и прогнозирование рыночных трендов.

**Здравоохранение**

- Клинические исследования: анализ геномных данных, эпидемиологический мониторинг и персонализированная медицина.

- Диагностика и лечение: применение алгоритмов для поддержки принятия врачебных решений и прогнозирования исходов заболеваний.

**Производство и логистика**

– предиктивное обслуживание: своевременное выявление неисправностей оборудования для снижения затрат на ремонт;

– оптимизация производственных процессов: повышение эффективности производства за счет анализа данных в режиме реального времени.

**Образование и наука**

– адаптивное обучение: разработка индивидуальных образовательных траекторий на основе анализа успеваемости и активности студентов;

– научные исследования: применение методов продвинутого анализа для обработки экспериментальных данных и моделирования сложных процессов.

**Перспективы дальнейших исследований**

– гибридные модели анализа: сочетание традиционных статистических методов с алгоритмами машинного обучения для повышения точности прогнозов;

– инновационные алгоритмы: разработка новых подходов для обработки неструктурированных данных и улучшения интерпретации результатов анализа.

**Заключение.** Big Data и продвинутый анализ данных становятся ключевыми инструментами для трансформации современных экономических, социальных и технологических процессов. Интеграция распределённых вычислительных платформ, алгоритмов машинного обучения, облачных технологий и IoT позволяет не только оптимизировать существующие процессы, но и открывать новые возможности для инноваций. Несмотря на многочисленные вызовы, дальнейшие исследования в данной области обещают значительный прогресс, а разработка этических стандартов и нормативной базы станет важным фактором устойчивого развития технологий обработки данных.

## Список литературы

[1] Mayer-Schönberger, V. & Cukier, K. Big Data: A Revolution That Will Transform How We Live, Work, and Think.

[2] Chen, M., Mao, S. & Liu, Y. Big Data: A Survey.

[3] Provost, F. & Fawcett, T. Data Science for Business.

[4] Russell, S. & Norvig, P. Artificial Intelligence: A Modern Approach.

## Авторский вклад

**Мырадов Пальван Сахетмырадович** – руководство исследованием Big Data и продвинутый анализ данных: современные тенденции, технологии и перспективы развития.

# BIG DATA AND ADVANCED DATA ANALYTICS: CURRENT TRENDS, TECHNOLOGIES, AND DEVELOPMENT PROSPECTS

***P.S. Myradov***
*Teacher, The State Energy*
*Institute of Turkmenistan*

**Abstract.** This article thoroughly examines the key concepts, methods, and technologies underlying Big Data processing and Advanced Analytics. Modern architectures, machine learning algorithms, natural language processing techniques, statistical approaches, and the role of cloud computing and the Internet of Things (IoT) are described. Practical application examples in business, healthcare, manufacturing, and education are considered, and the main challenges and directions for further research are outlined.

**Keywords:** Big Data, advanced analytics, machine learning, artificial intelligence, statistical analysis, cloud computing.

УДК 004.032.26

# НЕЙРОННЫЕ СЕТИ: МАТЕМАТИЧЕСКИЕ МОДЕЛИ, ИСПОЛЬЗУЕМЫЕ ДЛЯ РЕШЕНИЯ СЛОЖНЫХ ЗАДАЧ, ТАКИХ КАК РАСПОЗНАВАНИЕ ОБРАЗОВ, КЛАССИФИКАЦИЯ И ПРОГНОЗИРОВАНИЕ И МЕТОДЫ ОБРАБОТКИ И АНАЛИЗА ЗВУКОВЫХ ДАННЫХ, ВКЛЮЧАЯ РАСПОЗНАВАНИЕ РЕЧИ, КЛАССИФИКАЦИЮ ЗВУКОВ, ВЫДЕЛЕНИЕ ПРИЗНАКОВ.

***А.Х.Оразгелдиев***
*Преподаватель кафедры информационных систем Института телекоммуникаций и информатики Туркменистана*
abdyorazow588@gmail.com

***О.Д.Чарыева***
*Преподаватель кафедры информационных систем Института телекоммуникаций и информатики Туркменистана*
caryevahadzat@gmail.com

***Д. Я.Караджаева***
*Преподаватель кафедры информационных систем Института телекоммуникаций и информатики Туркменистана*

***А.Х. Оразгелдиев***
*Окончил Туркменский государственный университет имени Махтумкули Туркменистана. Область научных интересов сосредоточены на исследовании Нейронные сети, Распознавание речи*

***О.Д. Чарыева***
*Окончила Института телекоммуникаций и информатики Туркменистана. Область научных интересов сосредоточены на исследовании методов передачи знаний и обучения, направленных на улучшение производительности и решения задач в различных областях.*

***Д.Я. Караджаева***
*Окончила Туркменский государственный университет имени Махтумкули Туркменистана. Область научных интересов сосредоточены на исследовании обработка данных, большие базы данных*

**Аннотация:** Методы обработки и анализа звуковых данных включают предварительную обработку, выделение признаков и классификацию аудиосигналов. Распознавание речи позволяет преобразовывать звук в текст с использованием акустических и языковых моделей. Классификация звуков применяется для идентификации различных типов аудиосигналов, включая окружающие шумы и эмоции в голосе. Современные технологии, такие как нейросети и трансформеры, значительно повышают точность анализа звука. Области применения включают голосовых ассистентов, системы биометрической идентификации и автоматическое создание субтитров.

**Ключевые слова:** Обработка звука, распознавание речи, классификация аудио, выделение признаков, нейронные сети, акустическое моделирование, языковая модель, преобразование Фурье, спектральный анализ, биометрическая идентификация.

**Введение.** Нейронные сети – это математические модели, вдохновленные структурой человеческого мозга и используемые для решения сложных задач, таких как

распознавание образов, классификация и прогнозирование Они состоят из искусственных нейронов, которые соединены между собой и передают информацию друг другу, имитируя работу биологических нейронов.

Каждый искусственный нейрон принимает на вход несколько значений, обрабатывает их с помощью весов и функций активации, а затем передает результат дальше Слои нейронной сети делятся на входной, скрытые и выходной слои Входной слой принимает данные, скрытые слои выполняют вычисления и выявляют сложные закономерности, а выходной слой формирует конечный результат.

Распознавание речи – это процесс преобразования аудиосигнала в текст, который включает несколько шагов, таких как предварительная обработка звука, выделение акустических признаков, акустическое моделирование, языковое моделирование и декодирование. Распознавание в реальном времени требует оптимизированных алгоритмов и аппаратного ускорения, например, использования графических процессоров GPU и специализированных процессоров TPU.

Основные типы нейронных сетей включают полносвязные сети, сверточные нейронные сети CNN, рекуррентные нейронные сети RNN и трансформеры Полносвязные сети соединяют каждый нейрон одного слоя с каждым нейроном следующего слоя и применяются для обработки числовых данных и классификации объектов Сверточные нейросети предназначены для работы с изображениями и анализируют пространственные зависимости между пикселями, что делает их эффективными для распознавания лиц, медицинской диагностики и компьютерного зрения. Для классификации звуков используются методы глубокого обучения, такие как сверточные нейросети CNN и рекуррентные нейросети RNN, которые анализируют временные и спектральные характеристики сигнала.

Рекуррентные нейронные сети используются для работы с последовательностями данных, таких как речь, текст или временные ряды, поскольку они учитывают информацию из предыдущих шагов и могут запоминать контекст. Они применяются в системах машинного перевода, распознавании речи и прогнозировании временных рядов.

Трансформеры представляют собой более современный тип нейросетей, основанный на механизме самовнимания, который позволяет анализировать всю входную последовательность сразу, выявляя важные связи между различными частями данных. Они широко используются в обработке естественного языка, например, в чат-ботах, системах машинного перевода и генерации текста.

Обучение нейронных сетей происходит с использованием метода обратного распространения ошибки, который корректирует веса связей между нейронами с учетом разницы между предсказанным и реальным результатом Для этого применяется градиентный спуск – метод оптимизации, позволяющий минимизировать ошибку модели путем пошагового обновления параметров.

Важную роль в обучении нейросетей играет функция активации, которая определяет, каким образом нейрон передает сигнал дальше Популярные функции активации включают сигмоиду, гиперболический тангенс, ReLU и его модификации, такие как Leaky ReLU и ELU Функция ReLU чаще всего используется в глубоких нейросетях, поскольку она помогает избежать проблемы исчезающего градиента и ускоряет обучение.
Глубокое обучение – это направление машинного обучения, в котором используются многослойные нейросети с большим числом параметров. Такие сети способны извлекать сложные признаки из данных, что делает их эффективными в задачах компьютерного зрения, обработки речи, биоинформатики и других областях.

Архитектуры нейросетей могут быть как статичными, так и адаптивными. Например, сверточные нейросети адаптируют свои фильтры в процессе обучения, чтобы выявлять ключевые элементы изображения, такие как края, текстуры и формы, в то время

как рекуррентные нейросети адаптируют свою память, чтобы учитывать длинные зависимости в последовательностях.

Для повышения производительности и уменьшения переобучения используются методы регуляризации, такие как дропаут, нормализация весов и пакетная нормализация Дропаут отключает случайные нейроны во время обучения, чтобы модель не запоминала данные слишком точно и могла лучше обобщать информацию Пакетная нормализация стабилизирует распределение входных данных внутри сети, что ускоряет обучение и улучшает сходимость.

Обучение нейронных сетей требует значительных вычислительных ресурсов, поэтому часто используется аппаратное ускорение, такое как графические процессоры GPU и тензорные процессоры TPU. Эти устройства способны параллельно обрабатывать большие объемы данных, что делает обучение моделей значительно быстрее и эффективнее.

Современные нейросетевые архитектуры, такие как ResNet, EfficientNet и Vision Transformers, используются для сложных задач компьютерного зрения, включая медицинскую диагностику, автономное вождение и биометрическую идентификацию.

Применение нейронных сетей охватывает множество областей, включая финансы, где они используются для предсказания фондового рынка и выявления мошенничества, медицину, где они помогают анализировать снимки, находить патологии и разрабатывать персонализированные схемы лечения, а также производство, где они оптимизируют процессы и автоматизируют контроль качества продукции.

Нейросети активно используются в искусстве и развлечениях, например, для создания музыки, живописи и анимации, а также для генерации синтетических изображений и видео с высоким уровнем реалистичности.

Они также применяются в робототехнике, помогая автономным системам ориентироваться в пространстве, принимать решения и адаптироваться к изменяющимся условиям.

Обученные модели нейросетей могут использоваться повторно, благодаря чему создаются предобученные модели, такие как BERT, GPT и CLIP, которые можно дообучить для решения конкретных задач без необходимости начинать обучение с нуля.

Нейросети способны выявлять скрытые закономерности в данных, что делает их незаменимыми для анализа больших данных, обнаружения аномалий и построения рекомендательных систем.

Развитие технологий нейросетей привело к появлению мощных языковых моделей, таких как GPT-4, которые могут генерировать тексты, отвечать на вопросы, переводить языки и выполнять широкий спектр интеллектуальных задач.

Несмотря на свою эффективность, нейросети могут быть подвержены проблемам, таким как переобучение, нехватка данных и уязвимость к атакующим воздействиям, например, к атакам с использованием специально измененных входных данных, которые могут привести к неверным предсказаниям модели.

Для улучшения работы нейросетей разрабатываются методы интерпретируемости, такие как Grad-CAM и SHAP, которые позволяют визуализировать, какие именно признаки повлияли на решение модели, что особенно важно в критически важных областях, таких как медицина и право.

Методы обработки и анализа звуковых данных включают несколько ключевых этапов, начиная с преобразования звука в цифровую форму и заканчивая его интерпретацией с использованием искусственного интеллекта и машинного обучения.

Предварительная обработка звука включает удаление шума, нормализацию громкости и преобразование сигнала в удобное представление для дальнейшего анализа.

Выделение признаков – важный этап, который позволяет преобразовать звуковой сигнал в набор значимых параметров, таких как мел-кепстральные коэффициенты MFCC,

коэффициенты линейного предсказания LPC и спектральные признаки, которые используются для дальнейшей классификации и анализа.

Акустическое моделирование основано на использовании нейросетей или статистических моделей, таких как скрытые марковские модели HMM, которые помогают установить связь между звуковыми сигналами и фонемами.

Языковое моделирование применяется для предсказания наиболее вероятных последовательностей слов и построения осмысленного текста, используя вероятностные модели или нейросетевые архитектуры, такие как трансформеры.

Классификация звуков – это процесс, в котором звуковые данные разделяются на категории, например, идентификация типов окружающих звуков, детектирование тревожных сигналов или распознавание эмоций в голосе.

Одним из ключевых методов анализа звука является преобразование Фурье, которое позволяет разложить сигнал на составляющие частоты и представить его в виде спектрограммы.

Временные признаки, такие как продолжительность звука, энергия и тональность, помогают различать различные звуковые события и анализировать речь.

Спектральные признаки, такие как форма спектра, центроид и дисперсия, используются для анализа тембра звука и идентификации его источника.

Методы автоматического определения границ фонем и слов помогают улучшить точность распознавания речи, особенно в системах диалогового взаимодействия.

Генеративные модели, такие как вариационные автоэнкодеры и генеративные состязательные сети GAN, позволяют синтезировать звуки и улучшать качество речи в системах преобразования текста в речь.

Системы шумоподавления и улучшения качества звука используют нейросетевые алгоритмы для фильтрации фоновых шумов и улучшения разборчивости речи.

Распознавание эмоций по голосу основано на анализе интонации, ритма и спектральных характеристик, что позволяет использовать эти технологии в чат-ботах и виртуальных ассистентах.

Для анализа многоканальных аудиозаписей применяются алгоритмы локализации звука, которые позволяют определить источник звука в пространстве.

Современные системы анализа речи используют архитектуры трансформеров, такие как Wav2Vec и Whisper, которые позволяют достигать высокой точности при обработке звуковых данных.

Аудиобиометрия, основанная на анализе индивидуальных особенностей голоса, применяется в системах идентификации и аутентификации пользователей.

Обучение моделей обработки звука требует больших объемов данных, поэтому используются методы увеличения данных, такие как искусственное создание вариаций голосовых записей и смешивание фоновых шумов.

Применение искусственного интеллекта в анализе звука охватывает множество сфер, включая автоматические субтитры, системы управления голосом, медицинскую диагностику и криминалистику.

**Заключение** Будущее технологий обработки звука связано с развитием более точных и энергоэффективных моделей, способных работать в условиях реального мира с высокой точностью. Нейронных сетей связано с развитием новых архитектур, уменьшением энергопотребления, увеличением интерпретируемости моделей и созданием более универсальных систем, способных решать широкий спектр задач без необходимости узкоспециализированного обучения.

## Список литературы

[1] Овездурдыева И, Мырадов М, Технологии работы с большими данными “big data”: сбор, хранение и обработка больших данных Научное издательство «Наука и мировоззрение» 2024 70-74 с

[2] Мурадов М Назарова С, Современные технологии распознавания речи Научное издательство «Наука и мировоззрение» 2024 301-305 с

[3] Мурадов М., Распознавание речи Сборник статей Международной научно-практической конференции Научный потенциал - 2023 Петрозаводск 2023 405-409 с

## Авторский вклад

**Оразгельдиев Абдымухаммет Ходжамбердиевич-**занимался разработкой и применением нейросетевых моделей для распознавания речи и классификации аудиосигналов.

**Чарыева Огулхажат Довледовна-**внесла значительный вклад в исследование методов обработки звуковых данных, включая предварительную обработку и выделение признаков.

**Караджаева Джемал Язмырадовна-**провела анализ современных технологий акустического и спектрального моделирования, а также их применения в биометрической идентификации.

# NEURAL NETWORKS: MATHEMATICAL MODELS USED TO SOLVING COMPLEX PROBLEMS SUCH AS PATTERN RECOGNITION, CLASSIFICATION AND FORECASTING AND METHODS OF PROCESSING AND ANALYZING AUDIO DATA INCLUDING SPEECH RECOGNITION, SOUND CLASSIFICATION, FEATURE EXTRACTION.

***A.Kh. Orazgeldiev***
*Lecturer, Department of Information Systems, Institute of Telecommunications and Informatics of Turkmenistan*

***O.D. Charyeva***
*Lecturer of the Department of Information Systems of the Institute of Telecommunications and Informatics of Turkmenistan*

***D.Ya. Karadjaeva***
*Lecturer, Department of Information Systems, Institute of Telecommunications and Informatics of Turkmenistan*

**Abstract:** Methods for processing and analyzing audio data include pre-processing, feature extraction, and classification of audio signals. Speech recognition converts sound into text using acoustic and language models. Sound classification is used to identify different types of audio signals, including ambient noise and emotions in the voice. Modern technologies such as neural networks and transformers significantly improve the accuracy of sound analysis. Application areas include voice assistants, biometric identification systems, and automatic subtitling.

**Keywords:** Sound processing, speech recognition, audio classification, feature extraction, neural networks, acoustic modeling, language model, Fourier transform, spectral analysis, biometric identification

УДК 658.5.012.7

# ТЕХНОЛОГИИ BIG DATA В АВТОМАТИЗАЦИИ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

***С.С. Шайымов***
*Преподаватель кафедры "Автоматизации" института Телекоммуникаций и информатики Туркменистана*
smyradow615@gmail.com

***А.Б.Чарыев***
*Преподаватель кафедры "Автоматизации" института Телекоммуникаций и информатики Туркменистана*
arslan.caryyew93@gmail.com

***С.С. Шайымов***
*Окончил Государственный энергетический институт Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.*

***А.Б. Чарыев***
*Окончил Государственный энергетический институт Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.*

**Аннотация.** Переход к «цифровому производству» и технологическое развитие на основе цифровых технологий – это основа глобальной конкурентоспособности современных предприятий. Ранее под термином «цифровое производство» понимали набор прикладных систем, которые в основном использовались на этапе технологической подготовки производства, а именно: для автоматизации процессов разработки программ для станков с ЧПУ, для автоматизации разработки технологических процессов для сборки, для автоматизации задач, связанных с планированием рабочих мест при программировании роботов, и для интеграции с системами цехового уровня (или системами MES, Manufacturing Execution System) и системами управления ресурсами ERP. В последние годы, в связи с появлением новых прорывных технологий, этот термин получил более широкую трактовку.

**Ключевые слова:** «цифровое производство», Big Data, автономные роботы, горизонтальная и вертикальная интеграция систем.

Современные роботизированные системы функционируют в средах с высокой динамикой и неопределённостью. В таких условиях использование BIG DATA обеспечивает своевременную корректировку управляющих параметров на основе анализа крупных потоков сенсорной информации. Это согласуется с выводами, указывающими на важность статистической обработки больших наборов данных для минимизации ошибок позиционирования.

Эксперты выделяют следующие основные характеристики BIG DATA: объем данных от 150 ГБ в день.

BIG DATA постоянно обновляются, и для их обработки в режиме онлайн необходимы интеллектуальные технологии.

Разнообразие типов информации. Данные могут быть структурированными, неструктурированными или полуструктурированными. Например, поток информации в

социальных сетях неструктурирован и может включать текстовые сообщения, фотографии или видео.

С помощью высокопроизводительных технологий, таких как распределенные вычисления или аналитика в оперативной памяти, компании могут использовать любые объемы BIG DATA для анализа. Иногда они структурируют их, сначала выбирая то, что необходимо для анализа. BIG DATA используются для решения задач в рамках комплексной аналитики, в том числе с применением искусственного интеллекта.

Существует 4 основных метода анализа BIG DATA:

1 Наиболее распространенным методом является описательная аналитика. Он спросил: «Что случилось?» и анализирует текущую и историческую информацию. Его главная цель — выявить причины и закономерности успеха или неудачи в определенной области и использовать их в более эффективных моделях.

2 Предиктивная аналитика - помогает прогнозировать возможные события на основе имеющихся данных. Для этого используются модели, основанные на определенных объектах или сходных явлениях. Таким образом, можно спрогнозировать возможное падение или изменение цен на рынке казначейских обязательств или, по крайней мере, определить платежеспособность заемщика по проекту.

3 Предписывающая аналитика — с помощью BIG DATA и современных технологий можно выявить проблемные области в бизнесе или любой другой деятельности и определить, что необходимо сделать, чтобы предотвратить их повторение в будущем.

4 Диагностическая аналитика — использует данные для анализа причин инцидента. Это помогает обнаружить необычные обстоятельства и случайные связи между событиями и действиями [2].

Исследование процесса автоматизации технологических процессов с использованием технологий Big Data, определение основных этапов внедрения технологий Big Data в автоматизацию технологических процессов, анализ опыта внедрения технологий «больших данных» на современных промышленных предприятиях.

**Материалы и методы исследования**. Специалисты полагают, что изменения в современной промышленности, которые «цифровое производство» подразумевает, будут происходить по следующим ключевым направлениям:

– цифровое моделирование – развитие получает концепция цифрового двойника, то есть изготовление изделия в виртуальной модели, включающей в себя оборудование, производственный процесс и персонал предприятия.

– «Большие данные» (Big Data) и бизнес-аналитика, которые возникают в процессе производства.

– автономные роботы, которые получат большую промышленную функциональность, независимость, гибкость и исполнительность по сравнению с предыдущим поколением.

– горизонтальная и вертикальная интеграция систем – большая часть из огромного количества использующихся в настоящее время информационных систем интегрировано, но необходимо наладить более тесное взаимодействие на различных уровнях внутри предприятия, а также между различными предприятиями.

– промышленный Интернет вещей, когда поступающая с производства информация с большого количества датчиков и оборудования объединяется в единую сеть.

Один из признаков «цифрового производства» – наличие системы интеллектуального управления, то есть возможность плотной интеграции имеющегося технологического оборудования и получение широкого объема информации о технологическом процессе из любой точки производственной экосистемы.

Сегодня данные – одна из важнейших составляющих жизни общества и каждого человека. Современный этап развития общества характеризуется постоянным увеличением объема данных. Данные поступают из множества различных источников, это данные GPS-

навигаторов, спутников, интернет-запросов, социальных сетей, данные, получаемые из IoT (Internet of Things, Интернет вещей). Структура и состав этих данных зачастую не определены.

Big Data (большие данные) обладают следующими свойствами: огромные размеры, разнородность и неупорядоченность, требуют быстрой обработки. Технологии Big Data – это совокупность инструментов, подходов и методов обработки как структурированных, так и неструктурированных данных огромного размера для дальнейшего их использования. К основным технологиям и инструментам Big Data относятся:

– Hadoop & MapReduce;

– NoSQL базы данных;

–углубленная аналитика (статистика, предиктивная аналитика и Data Mining, лингвистическая обработка текстов);

– инструменты класса Data Discovery.

**Результаты исследования и их обсуждение**. Технологически функционирование высокоавтоматизированного цифрового предприятия обобщенно выглядит следующим образом. С помощью технологий Интернета вещей в физическом пространстве собираются огромные массивы информации, которые отправляются в киберпространство, где анализируются с привлечением искусственного интеллекта. Результаты этого анализа возвращаются обратно в физическое измерение, и здесь уже на их основе принимаются управленческие решения.

Технологии Big Data позволяют автоматизировать технологические и бизнес-процессы, что приводит к повышению скорости реакции бизнеса на внешние и внутренние возмущения. Экономический эффект достигается за счет повышения прозрачности процессов, улучшения качества планирования, внедрения повсеместного управления по отклонениям (или по целям), повышения оперативности определения причин отклонений, постоянным нормированием и стандартизацией лучших практик. Внедрение Big Data имеет вполне конкретную цель – реализация динамической модели управления бизнесом, которая обеспечивает быструю реакцию бизнеса на внешние и внутренние возмущения. Если в системе происходят возмущения, будь то нестабильность в сбыте, вариации в производстве, отклонения в снабжении, адаптивная модель управления бизнесом позволяет быстро «перенастроиться» в соответствии с лучшей стратегией, что обеспечивает ему кардинальное повышение конкурентоспособности. Более того, адаптивная модель управления бизнесом использует нестабильности, вариации, отклонения для постоянного улучшения структуры и параметров бизнеса.

Общая схема применения технологий Big Data в процессе автоматизации технологических и бизнес – процессов на предприятии приведена на рисунке.

Технологии Big Data в процессе автоматизации технологических и бизнес-процессов
Производители устанавливают датчики на ключевые части оборудования для сбора информации в режиме реального времени. Полученные и обработанные данные отправляются во все подразделения предприятия для обеспечения взаимодействия между структурными подразделениями и принятия соответствующих управленческих решений.
Эту информацию можно использовать для улучшения сервисного обслуживания (предотвращение простоев, поломок оборудования), для создания целевых маркетинговых предложений.

Непрерывный мониторинг ключевых показателей дает возможность определить проблему и принять необходимые меры для ее решения. Современные системы позволяют наблюдать за технологическим процессом и выявлять влияющие на него факторы с помощью любого Web-браузера. Такие решения позволяют превратить производственные данные в информацию, необходимую для эффективного управления предприятием.

В качестве основных этапов внедрения технологий Big Data в автоматизацию технологических процессов можно назвать следующие:

Извлечение данных из источников (большая часть данных генерируется на уровне датчиков, АСУТП, MES, ERP).

Хранение данных (соответствующее хранилище данных, не зависящее от производственных серверов.

Обработка данных (нормализация данных для того, чтобы впоследствии можно было применять инструменты «гибкой аналитики», статистического контроля и анализа, цифрового моделирования).

Анализ данных. «Гибкая аналитика» позволяет с помощью мастеров самостоятельно просто и быстро организовать для себя клиент-ориентированную аналитику. То есть любой обученный работник настраивает для себя ту отчетность, которая ему необходима для работы: отчеты, графики, гистограммы, регрессии, карты Шухарта. Тем самым, кардинально повышается эффективность анализа данных в бизнесе.

Цифровое моделирование позволяет найти «цифровые двойники» не только технологических объектов, но и бизнес – процессов. Когда у специалиста есть «цифровой двойник», он может оперативно найти наилучшие нормы, технологические режимы, процедуры, регламенты. Кроме того, «цифровые двойники» позволяют реализовать сквозное оперативное планирование потока создания ценности в соответствии со стратегическими целями, что обеспечивает как оперативную синхронизацию объектов потока, так их и оптимизацию.

Оптимизация.

Прогнозирование.

Технологии Big Data в качестве ключевого компонента промышленного Интернета вещей уже широко используются на многих предприятиях в России и за рубежом, позволяя работникам улучшить уже имеющиеся навыки, а предприятию функционировать более эффективно.

Приведем несколько примеров использования технологий Big Data на промышленных предприятиях.

ThyssenKrupp AG, один из ведущих мировых производителей лифтов, внедрила систему MAX, которая через Интернет вещей собирает данные от датчиков, установленных в лифтах компании, и по ним строит модели на платформе Azure Machine Learning, позволяющие предупредить инцидент до его возникновения и передать технику конкретный код поломки для сокращения времени обслуживания.

Компания General Electric Oil & Gas, занимающаяся производством высокотехнологичного оборудования для нефтегазового сектора, использует Big Data для минимизации времени незапланированной остановки нефтедобычи.

Используя данные по продажам за предыдущие периоды и оптимизационные алгоритмы, компания Nestle с помощью технологий анализа больших данных осуществляет оптимизацию производственной цепочки, это позволяет снизить ошибки при прогнозировании спроса на материалы в два раза, также снижаются убытки от хранения избыточных оборотных средств, от задержек в производстве и т.д.

Каждый компьютерный компонент, производимый компанией Intel, перед выходом на рынок должен пройти примерно 19000 различных тестов. Использование технологий Big Data позволяет на основе анализа данных по всему производственному процессу оставить только часть этих проверок, соответственно, уменьшается время тестирования компонентов и затраты на него.

Платформа обрабатывает параметры плавки, а затем выдает соответствующие рекомендации. Специалисты ММК считают, что это приводит к существенной экономии, кроме того, подобная технология может найти применение в целлюлозно-бумажной и химической промышленности.

Применение Big Data имеет и ряд проблем. Основной являются затраты на обработку данных, включающие в себя дорогостоящее оборудование и расходы на заработную плату

квалифицированным специалистам, способным обслуживать огромные массивы информации. Второй проблемой является предвзятость. Если исследование дает не 2–3, а многочисленное количество результатов, очень сложно остаться объективным и выделить из общего потока данных только те, которые окажут реальное влияние на состояние какого-либо явления. Третья проблема – это защита Big Data. Методологии по защите информационных систем классической трёхзвенной архитектуры оказываются неприменимы к новым технологиям. Возникает потребность в создании и подготовке нового класса специалистов по безопасности Big Data.

**Выводы**. Технологии Big Data сейчас – это уже вполне реально работающий набор технологий, используемых практически во всех сферах человеческой деятельности и обладающих большим потенциалом для дальнейшего развития.

Для внедрения технологий Big Data требуется не только техническое сопровождение, но и организационное. Первое подразумевает организацию извлечения данных, хранилища данных, унифицированных АРМ для анализа, цифрового моделирования, оптимизации и прогнозирования. Второе направление потребует формирования соответствующих квалификаций в бизнесе по Big Data. Необходимы специалисты с новыми квалификациями «инженеры по данным», «data scientist» для моделирования, оптимизации и прогнозирования.

## Список литературы

[1] Чарыев А.Б., Шайымов С.С., Аннабердиев И.А., “Важность использования BIG DATA для оптимизации управления роботами-манипуляторами”. Международный научный журнал “ИННОВАЦИОННАЯ НАУКА” №2-1/2025.

[2 ]Шайымов С.С., Гочов А.С., Веллиева С.Т., “Значение BIG DATA в цифровую эпоху”. Международный научный журнал “ИННОВАЦИОННАЯ НАУКА” №2-2/2025.

## Авторский вклад

Авторы внести равнозначный вклад.

# BIG DATA TECHNOLOGIES IN AUTOMATION OF TECHNOLOGICAL PROCESSES

***S.S. Shayymov***
*Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan*

***A.B. Charyev***
*Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan*

**Abstract.** The transition to "digital manufacturing" and technological development based on digital technologies is the basis for the global competitiveness of modern enterprises. Previously, the term "digital manufacturing" was understood as a set of application systems that were mainly used at the stage of technological preparation of production, namely: to automate the processes of developing programs for CNC machines, to automate the development of technological processes for assembly, to automate tasks related to the planning of workplaces when programming robots, and to integrate with shop-floor systems (or MES, Manufacturing Execution System) and ERP resource management systems. In recent years, due to the emergence of new breakthrough technologies, this term has received a broader interpretation.

**Keywords**: "digital manufacturing", Big Data, autonomous robots, horizontal and vertical integration of systems.

УДК 004.021:004.622

# ПРИНЯТИЕ РЕШЕНИЙ НА ОСНОВЕ ДАННЫХ В СЕЛЬСКОМ ХОЗЯЙСТВЕ

***А.Д. Русецкий***
*Ассистент кафедры*
*экономической информатики*
*БГУИР*
ruseckiy.alexander@gmail.com

***А.Д. Русецкий***

*Окончил Минский радиотехнический колледж. Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой и оптимизацией методов и алгоритмов учета и вычисления показателей производственно-коммерческой деятельности сельскохозяйственных предприятий Беларуси.*

**Аннотация.** Рассмотрена интеграция анализа больших данных в точное земледелие с упором на разработку прогнозных моделей, которые могут оптимизировать урожайность, сократить отходы и улучшить распределение ресурсов, включая воду, удобрения и энергию.

Рассмотрена роль сбора данных в реальном времени с датчиков, дронов и спутников в сочетании с моделями машинного обучения для предоставления сельскому хозяйству практической информации.

Изучены проблемы наличия, доступности и масштабируемости данных, особенно для сельских хозяйств в отдаленных от областных центров регионах.

**Ключевые слова:** прогнозная аналитика, точное земледелие, большие данные, Интернет вещей в сельском хозяйстве, оптимизация урожайности, устойчивое сельское хозяйство, сельское хозяйство, управляемое данными.

**Введение.** Сельскохозяйственный сектор сталкивается с растущими вызовами, такими как рост населения, изменение климата, нехватка ресурсов и экологическое разрушение. Сельское хозяйство потребляет 70% пресной воды и производит почти четверть выбросов парниковых газов [1]. Изменение климата снижает урожайность, увеличивает частоту экстремальных погодных явлений и нарушает доступность воды [2]. Проблемы истощения почв, деградации земель и неэффективного использования удобрений усугубляют ситуацию, требуя инновационных подходов.

Точное сельское хозяйство, основанное на технологиях Big Data, или больших данных, IoT, или интернета вещей, и искусственного интеллекта, способно предложить решения для оптимизации ресурсов, повышения продуктивности и снижения экологического воздействия. Использование спутниковой съемки, датчиков и аналитики позволяет в реальном времени отслеживать состояние культур, почвы и экологических факторов, что способствует принятию обоснованных решений [3]. Машинное обучение и прогнозное моделирование помогают предсказать урожайность, оптимизировать посадку и бороться с вредителями [4].

Интернет вещей и большие данные могут предоставить сельским хозяйствам актуальную информацию для управления ресурсами, такими как вода, удобрения и пестициды, повышая эффективность и устойчивость [5]. Исследование направлено на

изучение интеграции этих технологий для решения ключевых проблем сельского хозяйства, включая оптимизацию урожайности, сокращение потерь и улучшение управления. Акцент делается на разработке прогнозных моделей для своевременного реагирования на вредителей, болезни и погодные условия, а также на анализе лучших практик и преимуществ внедрения точного сельского хозяйства.

**Аналитика больших данных в сельском хозяйстве.** Большие данные в сельском хозяйстве представляют собой обширные и сложные наборы информации, собираемой из множества источников, таких как датчики почвы, спутниковые изображения, метеостанции, сельхозтехника, дроны и даже геномные данные. Эти данные охватывают широкий спектр информации, включая экологические условия, состояние почвы, здоровье культур, использование ресурсов (воды, удобрений, пестицидов) и рыночные тенденции. Их анализ позволит сельскому хозяйству принимать обоснованные решения, повышая продуктивность, эффективность и устойчивость сельскохозяйственных практик [6].

Объем больших данных в сельском хозяйстве значительно превышает традиционные методы сбора информации. Современные технологии, такие как IoT-устройства, дроны и спутники, генерируют огромные массивы данных в реальном времени. Например, датчики IoT способны предоставить информацию о влажности почвы, температуре и уровне питательных веществ, а дроны с мультиспектральными камерами помогут оценивать здоровье культур и выявлять проблемные зоны. Спутниковые изображения позволят отслеживать изменения в растительности и качестве почвы на больших территориях, а метеостанции будут собирать локализованные данные о погоде, что критически важно для планирования сельскохозяйственных работ.

Для обработки таких объемов данных потребуются передовые аналитические инструменты, такие как машинное обучение и искусственный интеллект. Эти технологии позволят создавать прогнозные модели, которые помогут прогнозировать урожайность, оптимизировать графики посадки, оценивать риски, связанные с погодными условиями или вспышками вредителей, и более эффективно распределять ресурсы. Например, анализируя исторические данные и текущие условия, сельскохозяйственные предприятия могут определить оптимальное количество воды, удобрений и пестицидов, что снижает потери и минимизирует негативное воздействие на окружающую среду [7].

Прогнозная аналитика и машинное обучение играют ключевую роль в трансформации сельского хозяйства. Они позволяют не только прогнозировать будущие результаты, но и оперативно реагировать на изменения. Например, можно получать уведомления о потенциальных вспышках заболеваний или неблагоприятных погодных условиях, что позволит своевременно принимать меры для защиты урожая. Это способствует повышению продуктивности, снижению затрат и улучшению устойчивости сельхозпроизводства.

Однако внедрение больших данных в сельском хозяйстве сталкивается с рядом вызовов. Одна из основных проблем – гетерогенность источников данных. Информация поступает из разных платформ (датчики, дроны, спутники), каждая из которых имеет свои форматы и стандарты, что затрудняет интеграцию и анализ. Кроме того, огромные объемы данных требуют мощной инфраструктуры для хранения и обработки, что может быть недоступно для небольших сельскохозяйственных предприятий, особенно в развивающихся странах. Также возникают вопросы конфиденциальности и безопасности данных, поскольку сельхозпроизводители должны защищать информацию о своих операциях [8].

Несмотря на эти трудности, большие данные имеют огромный потенциал для модернизации сельского хозяйства. Они способствуют развитию точного земледелия, где решения принимаются на основе детализированных данных о каждом участке поля, вплоть до отдельного растения. Это позволяет минимизировать использование ресурсов, снизить экологический след и повысить урожайность. В условиях растущих глобальных вызовов,

таких как изменение климата и истощение ресурсов, большие данные становятся ключевым инструментом для создания устойчивых и эффективных сельскохозяйственных систем.

**Прогностические модели для оптимизации урожайности.** Прогностические модели в сельском хозяйстве являются важными для улучшения принятия решений и повышения эффективности управления урожаем. Два основных типа прогностических моделей – это моделирование урожая и прогнозирование погоды.

*Моделирование урожая*. Модели урожая симулируют рост и развитие растений в различных экологических условиях, помогая сельскому хозяйству предсказывать урожай и эффективно управлять ресурсами. Эти модели интегрируют данные о характеристиках почвы, климате, физиологии растений и методах управления для прогнозирования таких результатов, как темпы роста, потенциал урожая и потребности в питательных веществах. Например, такие модели, как DSSAT (Система поддержки принятия решений для агрономической передачи технологий) и APSIM (Симулятор сельскохозяйственных производственных систем), широко используются для оценки различных сценариев посева и информирования о решениях по посадке [9].

*Прогнозирование погоды*. Точные модели прогнозирования погоды имеют решающее значение для сельского хозяйства, так как они предоставляют важную информацию о количестве осадков, температуре, влажности и условиях ветра, которые непосредственно влияют на здоровье растений. Современные метеорологические модели используют исторические данные о погоде, спутниковые наблюдения и атмосферные симуляции для прогнозирования краткосрочных и долгосрочных погодных условий. Работники сельскохозяйственных организаций могут использовать эту информацию для принятия обоснованных решений относительно орошения, управления вредителями и сроков сбора урожая [10]. Используя эти прогнозные модели, участники сельского хозяйства могут повысить продуктивность и устойчивость в условиях все более неопределенного климата.

Хотя прогнозные модели в сельском хозяйстве предлагают значительные преимущества, они также сталкиваются с ограничениями, которые могут препятствовать их эффективности. Одним из основных ограничений является качество и доступность данных. Многие модели зависят от точных и полных наборов данных, но такие проблемы, как неполные данные, неточности и вариации в методах сбора данных, могут привести к ненадежным прогнозам. Кроме того, региональная изменчивость в сельскохозяйственных практиках, климате и типах почвы может повлиять на применимость обобщенных моделей в разных местах.

Еще одной проблемой является сложность сельскохозяйственных систем. Прогнозные модели часто испытывают трудности с учетом множества переменных, влияющих на рост растений и динамику вредителей, таких как взаимодействия между различными видами, почвенные микробиомы и непредвиденные экологические факторы.

Для улучшения прогнозных моделей можно использовать достижения в методах интеграции данных и применение данных в реальном времени, что может повысить точность. Включение более детализированных источников данных, таких как показания датчиков на месте, может предоставить более целостное представление о сельскохозяйственных условиях. Кроме того, разработка адаптивных алгоритмов, которые учатся и развиваются в условиях изменения, может помочь создать более устойчивые модели, позволяя делать более точные прогнозы в условиях климатической изменчивости и возникающих сельскохозяйственных проблем.

**Распределение ресурсов и оптимизация.** Точное орошение, как часть точного сельского хозяйства, использует современные технологии, такие как мониторинг в реальном времени и автоматизированные системы, для оптимизации использования воды в сельском хозяйстве. Этот подход основан на данных с датчиков влажности почвы, прогнозах погоды и потребностях растений в воде, что позволяет точно применять воду в нужное время и в нужном месте. Мониторинг в реальном времени с помощью датчиков

влажности почвы помогает избежать перенасыщения или недостаточного насыщения почвы водой, что снижает потери воды и повышает урожайность [11]. Автоматизированные системы, такие как капельное или спринклерное орошение, работают на основе данных в реальном времени, минимизируя человеческие ошибки и снижая затраты на труд [12]. Эти технологии способствуют значительной экономии воды, улучшению здоровья растений и повышению продуктивности, поддерживая устойчивое управление водными ресурсами.

Аналитика данных также играет ключевую роль в оптимизации применения удобрений. Используя данные с датчиков и спутниковых изображений, сельские хозяйства могут выявлять участки поля, требующие удобрений, и применять их целенаправленно, что минимизирует отходы и снижает риск загрязнения окружающей среды [13]. Прогнозное моделирование на основе стадий роста культур, погодных условий и состояния почвы позволяет оптимизировать время и нормы внесения удобрений, снижая потери питательных веществ. Это повышает урожайность и способствует экологической устойчивости.

Интернет вещей (IoT) трансформирует сельскохозяйственные операции, снижая потребление энергии. Умные системы орошения на основе IoT контролируют влажность почвы и погодные условия, что позволяет экономить энергию на перекачку воды. Автоматизированное оборудование, такое как тракторы и комбайны, оптимизирует использование техники, снижая потребление топлива. Мониторинг энергии с помощью IoT помогает выявлять области для экономии, например, оптимизируя освещение в теплицах или вентиляционные системы. Эти решения повышают операционную эффективность и способствуют устойчивости сельского хозяйства.

Оптимизация использования ресурсов приносит как экономические, так и экологические выгоды. Эффективное управление водой, удобрениями и энергией снижает затраты и повышает прибыль. Снижение чрезмерного использования ресурсов уменьшает экологические последствия, такие как загрязнение водоемов и истощение водных ресурсов, что особенно важно в условиях изменения климата. В целом, оптимизация ресурсов поддерживает экономическую жизнеспособность сельского хозяйства и способствует глобальной продовольственной безопасности, обеспечивая долгосрочную устойчивость сельскохозяйственных систем.

**Интернет вещей и сбор данных в реальном времени.** Интернет вещей (IoT) революционизирует сельское хозяйство, повышая продуктивность и устойчивость через использование устройств, таких как датчики, дроны и умная техника. Эти устройства обеспечивают сбор и анализ данных в реальном времени, что позволяет оптимизировать управление ресурсами и улучшать принятие решений.

Датчики влажности почвы, температуры и питательных веществ предоставляют актуальную информацию о состоянии полей, помогая точно регулировать орошение и внесение удобрений. Это способствует экономии ресурсов и повышению урожайности. Дроны, пусть на их использование и требуется разрешение государства, проводят аэрофотосъемку, выявляя проблемы с здоровьем растений, вредителями и потребностями в воде. Они позволяют быстро собирать данные на больших площадях, улучшая оперативность решений. Умная техника, такая как тракторы и комбайны с GPS и аналитическими системами, автоматизирует задачи (посев, внесение удобрений, сбор урожая), снижая затраты на топливо и рабочую силу.

Мониторинг в реальном времени с помощью IoT позволяет оперативно реагировать на изменения условий. Датчики влажности почвы помогают определить оптимальное время для орошения, предотвращая стресс растений и экономя воду. Анализ данных с использованием машинного обучения выявляет тенденции, такие как вспышки вредителей или потребности в питательных веществах, что позволяет заранее распределить ресурсы и снизить потери [14].

Преимущества данных в реальном времени включают повышенную реактивность и своевременное управление. Сельхозпроизводители могут вовремя реагировать на

изменения, предотвращая проблемы до их обострения. Например, прогнозирование вспышек вредителей позволяет применять меры заранее, сводя ущерб к минимуму. Кроме того, оптимизация использования воды и удобрений снижает затраты и экологическое воздействие.

В целом, IoT предоставляет сельским хозяйствам инструменты для принятия обоснованных решений, повышая продуктивность, устойчивость и жизнеспособность сельскохозяйственных практик.

**Проблемы масштабируемости и доступности.** Несмотря на трансформационный потенциал технологий Интернета вещей (IoT) в сельском хозяйстве, их внедрение сталкивается с серьезными проблемами, такими как масштабируемость и доступность.

Проблемы масштабируемости включают высокие затраты на приобретение и обслуживание IoT-устройств, что делает их недоступными для малых хозяйств или хозяйств с большими посевными площадями. Кроме того, интеграция технологий в существующие практики требует значительных ресурсов, что затрудняет их широкое применение [15]. Проблемы доступности связаны с отсутствием надежного интернета в сельских районах, что ограничивает использование данных в реальном времени. Нехватка технической экспертизы также препятствует внедрению IoT-решений.

В отдаленных от областных центров регионах ситуация усугубляется недостаточным развитием инфраструктуры, такой как интернет и стабильное электроснабжение, что делает невозможным использование облачных платформ и аналитики данных. Высокая стоимость технологий и постоянные расходы на их обслуживание также ограничивают доступ для многих хозяйств, вынуждая их полагаться на традиционные методы [16].

Цифровой разрыв создает дополнительные барьеры: отсутствие доступа к смартфонам, компьютерам и интернету, низкий уровень цифровой грамотности и финансовые ограничения. Это препятствует использованию цифровых инструментов и инноваций в сельском хозяйстве.

Безопасность данных также является критической проблемой. Сельскохозяйственные данные, включая урожайность и финансовую информацию, уязвимы для киберугроз. Для их защиты необходимы надежные протоколы кибербезопасности, шифрование и обучение работников. Четкие правила конфиденциальности и прозрачность в использовании данных также важны для построения доверия.

Для решения этих проблем предлагаются следующие меры:

*Развитие инфраструктуры*: государственные инвестиции в интернет и обеспечение стабильного электроснабжения.

*Доступные технологические решения*: разработка недорогих устройств и мобильных приложений, способных работать в оффлайн-режиме.

*Программы обучения*: повышение цифровой грамотности работников сельскохозяйственных организаций через семинары.

*Финансовая поддержка*: государственное субсидирование и варианты финансирования для приобретения IoT-устройств.

Эти шаги помогут улучшить доступность данных и масштабируемость технологий, способствуя устойчивому развитию сельского хозяйства.

**Кейс-исследования: успешные исполнения систем, основанных на данных.**

Компания John Deere, мировой лидер в производстве сельхозтехники, использует платформу «Центр операций John Deere» для сбора и анализа данных с датчиков, метеостанций и машин. Эта платформа предоставляет фермерам информацию в реальном времени о состоянии почвы, здоровье растений и погодных условиях, что позволяет оптимизировать графики посадки, орошения и внесения удобрений. Например, фермеры, выращивающие кукурузу, смогли увеличить урожайность на 15% за счет корректировки норм удобрений на основе данных о питательных веществах в почве. Предиктивная

аналитика также помогает предсказывать вспышки вредителей и заболеваний, минимизируя потери и снижая экологическое воздействие.

Проект Digital Green в Индии помогает мелким фермерам использовать мобильные приложения для сбора данных в реальном времени о состоянии почвы, росте культур и погодных условиях. Фермеры получают рекомендации по орошению, борьбе с вредителями и срокам сбора урожая. Например, в Мадхья-Прадеш урожайность увеличилась на 20% благодаря этим технологиям. Проект также использует подход «равный к равному», где фермеры обмениваются опытом через видеодемонстрации, что улучшает передачу знаний и устойчивость к изменениям климата [17].

Внедрение точного сельского хозяйства ведет к экономическим, социальным и экологическим последствиям.

*Экономические*: технологии, такие как IoT и предиктивная аналитика, сокращают затраты на ресурсы и повышают урожайность. Например, точное орошение снижает потребление воды на 30%, увеличивая прибыльность.

*Социальные*: доступ к данным в реальном времени помогает принимать обоснованные решения, улучшая продовольственную безопасность и устойчивость аграрного сектора в целом.

*Экологические*: оптимизация использования удобрений и воды снижает экологический след, предотвращая деградацию почвы и улучшая биоразнообразие.

Можно выделить следующие лучшие практики внедрения точного сельского хозяйства:

*Оценка местных потребностей*: учет региональных особенностей и доступности ресурсов.

*Инвестиции в инфраструктуру*: развитие интернета и стабильного электроснабжения, особенно в отдаленных от областных центров районах.

*Обучение и поддержка*: повышение цифровой грамотности фермеров через семинары и онлайн-уроки.

*Сотрудничество*: партнерство между сельским хозяйством, поставщиками технологий и государственными органами.

*Масштабируемость*: использование гибких решений, которые можно адаптировать как для мелких, так и крупных хозяйств.

Эти подходы способствуют устойчивому развитию сельского хозяйства, повышая продуктивность, экономическую жизнеспособность и экологическую устойчивость.

**Будущее сельского хозяйства, основанного на данных.**

Появляющиеся технологии, такие как искусственный интеллект (ИИ), блокчейн и продвинутая робототехника, революционизируют сельское хозяйство, повышая его эффективность и устойчивость.

Искусственный интеллект способен анализировать большие объемы данных для оптимизации решений, предсказывать урожайность, выявлять вредителей и определять оптимальные сроки посадки.

Например, дроны с ИИ способны оценивать здоровье растений с помощью аэрофотосъемки, повышая продуктивность.

Технология блокчейн повышает прозрачность и отслеживаемость в логистической цепочке поставок. Его можно использовать для проверки органических сертификатов, укрепляя доверие потребителей.

Продвинутая робототехника, а именно автоматизированные тракторы и комбайны снижают затраты на труд и повышают точность.

Интеграция с IoT позволяет собирать и анализировать данные в реальном времени, улучшая управление урожаем. Эти технологии повышают эффективность, устойчивость и адаптивность сельского хозяйства, решая проблемы глобального спроса на продовольствие.

**Заключение.** Интеграция аналитики больших данных в сельское хозяйство открывает возможности для повышения продуктивности, устойчивости и продовольственной безопасности. Ключевые выводы из исследования подчеркивают следующее:

*Улучшение принятия решений*: подходы, основанные на данных, позволяют сельским хозяйствам принимать обоснованные решения на основе актуальной информации, оптимизируя использование ресурсов и улучшая управление урожаем. Прогнозная аналитика может значительно увеличить урожайность при минимизации потерь.

*Экономические выгоды*: внедрение аналитики данных приводит к увеличению доходов сельских хозяйств за счет снижения затрат на ресурсы и повышения операционной эффективности. Технологии точного земледелия, такие как оптимизированное орошение и применение удобрений, показали снижение расходов и улучшение прибыльности.

*Социальная поддержка*: доступ к технологиям, основанным на данных, дает возможность малым сельским хозяйствам преодолевать пробелы в знаниях и способствует сотрудничеству в сообществе. Инициативы демонстрируют положительное влияние технологий на продовольственную безопасность и устойчивость хозяйств.

*Экологическая устойчивость*: аналитика больших данных способствует устойчивым сельскохозяйственным практикам, сводя к минимуму потери ресурсов и снижая негативное воздействие на окружающую среду, например, за счет уменьшения потребления воды и оптимизации управления питательными веществами.

*Поддержка политики*: эффективные политики и регуляторные рамки необходимы для стимулирования инноваций. Устранение проблем конфиденциальности данных, предоставление стимулов и повышение цифровой грамотности являются критически важными для создания среды для сельского хозяйства, основанного на данных.

В целом, аналитика больших данных имеет значительный потенциал для решения глобальных сельскохозяйственных проблем и обеспечения устойчивого будущего.

Сельское хозяйство, основанное на данных, способно оптимизировать использование ресурсов и значительно повысить урожайность. Используя передовые технологии, такие как аналитика больших данных, Интернет вещей (IoT) и искусственный интеллект (AI), сельские хозяйства могут получать практические рекомендации по всем аспектам своей деятельности.

*Оптимизация ресурсов*: технологии точного земледелия позволяют хозяйствам в реальном времени контролировать состояние почвы, уровень влажности и доступность питательных веществ. Этот целенаправленный подход позволяет им более эффективно применять воду, удобрения и пестициды, снижая потери и сокращая затраты. Например, системы точного орошения могут сократить потребление воды до 30%, обеспечивая растения необходимым количеством для оптимального роста.

*Повышение урожайности*: прогнозная аналитика играет ключевую роль в повышении урожайности, заранее предупреждая потенциальные проблемы, такие как нашествия вредителей или неблагоприятные погодные условия. Более того, принятие решений на основе данных способствует постоянному улучшению, позволяя хозяйствам анализировать показатели эффективности и совершенствовать практики со временем.

В заключение, сельское хозяйство, основанное на данных, не только повышает эффективность использования ресурсов, но и способствует увеличению урожайности, что вносит вклад в продовольственную безопасность и устойчивые методы ведения сельского хозяйства в условиях ограниченных ресурсов.

Чтобы малые сельские хозяйства могли извлечь выгоду из сельского хозяйства, основанного на данных, крайне важно разработать масштабируемые и доступные решения, адаптированные к их конкретным потребностям. Следующие стратегии могут помочь в достижении этой цели:

*Упрощенные технологии*: разработка удобных для пользователя приложений и инструментов, которые требуют минимальной технической подготовки. Интуитивно

понятные интерфейсы помогут сельским хозяйствам легко получать доступ к данным и интерпретировать их, делая информацию более практичной и актуальной для их повседневной деятельности.

*Доступные решения*: экономически эффективные технологии необходимы для малых сельских хозяйств, которые часто работают с ограниченными ресурсами. Предоставление субсидированных или многоуровневых ценовых моделей может повысить доступность и способствовать внедрению.

*Вовлечение сообщества*: участие работников сельскохозяйственных организаций в проектировании и внедрении решений гарантирует, что технологии будут адаптированы к местным условиям и практикам, что повысит их эффективность.

*Обучение и поддержка*: реализация комплексных программ обучения, которые наделяют работников сельскохозяйственных организаций необходимыми навыками для использования технологий, основанных на данных. Постоянная поддержка через семинары, онлайн-ресурсы и местные службы расширения может повысить уверенность и компетентность.

*Развитие инфраструктуры*: инвестиции в развитие инфраструктуры, а именно стабильный интернет, чтобы обеспечить развертывание решений, основанных на данных [18].

Сосредоточив внимание на этих стратегиях, возможно создать поддерживающую экосистему, которая позволит сельскохозяйственным организациям процветать в условиях сельского хозяйства, основанного на данных.

## Список литературы

[1] ФАО. Использование воды в сельском хозяйстве. – Рим: Издательство ФАО, 2021. – 25 с.

[2] МГЭИК. Изменение климата и земля. – Женева: Секретариат МГЭИК, 2019. – 21 с.

[3] Фунтас, С. и др. Внедрение и преимущества точного земледелия в европейских странах. – Берлин: Springer, 2020. – 12 с.

[4] Ван, Х. и др. Применение искусственного интеллекта и машинного обучения в точном земледелии. – Нью-Йорк: Academic Press, 2021. – 18 с.

[5] Джаяшанкар, П. и др. Интернет вещей в сельском хозяйстве: устойчивые решения для точного земледелия. – Лондон: CRC Press, 2018. – 11 с.

[6] Вольферт, С. и др. Большие данные в умном сельском хозяйстве – обзор. – Вагенинген: Wageningen Academic Publishers, 2017. – 26 с.

[7] Камиларис, А. и др. Обзор применения искусственного интеллекта в сельском хозяйстве. – Амстердам: Elsevier, 2018. – 28 с.

[8] Басco, Б., Антл, Дж. Цифровое сельское хозяйство для проектирования устойчивых агросистем. – Кембридж: Cambridge University Press, 2020. – 12 с.

[9] Шарма, Р. и др. Роль прогнозной аналитики в сельском хозяйстве: обзор. – Нью-Дели: Springer, 2020. – 24 с.

[10] Кумар, П. и др. Проблемы и возможности больших данных в сельском хозяйстве. – Берлин: Springer, 2020. – 17 с.

[11] Китинг, Б. А. и др. Обзор APSIM. – Брисбен: Инициатива APSIM, 2003. – 38 с.

[12] Холландер, А. и др. Роль прогнозирования погоды в сельском хозяйстве. – Нью-Йорк: Wiley, 2021. – 10 с.

[13] Феререс, Э., Сориано, М. А. Дефицитное орошение для сокращения использования воды в растениеводстве. – Уоллингфорд: CABI Publishing, 2007. – 18 с.

[14] Чжан, Ю. и др. Умные системы орошения: обзор. – Берлин: Springer, 2018. – 28 с.

[15] Смит, Р. и др. Точное земледелие: возможности и вызовы. – Чикаго: University of Chicago Press, 2020. – 17 с.

[16] Вольферт, С. и др. Большие данные в умном сельском хозяйстве – обзор. – Вагенинген: Wageningen Academic Publishers, 2017. – 19 с.

[17] Смит, Р. и др. Использование больших данных для оптимизации урожайности в сельском хозяйстве. – Чикаго: University of Chicago Press, 2020. – 25 с.

[18] Кумар, П. и др. Расширение возможностей мелких фермеров: влияние Digital Green. – Берлин: Springer, 2020. – 23 с.

**Авторский вклад**

**Русецкий Александр Дмитриевич** – постановка задачи исследования, исследование по оценке перспектив внедрения технологий на основе больших данных и интернета вещей в сельское хозяйство, формирование структуры статьи.

# DATA-BASED DECISION MAKING IN AGRICULTURE

***A.D. Rusetsky***
*Assistant of the Department of Economic Informatics,*
*BSUIR*

**Abstract.** Examines the integration of big data analytics into precision agriculture, with a focus on developing predictive models that can optimize yields, reduce waste, and improve resource allocation, including water, fertilizer, and energy.

Explores the role of real-time data collection from sensors, drones and satellites combined with machine learning models to provide actionable information to agriculture.

The problems of data availability, accessibility and scalability have been studied, especially for rural farms in regions remote from regional centers.

**Keywords**: predictive analytics, precision agriculture, big data, IoT in agriculture, yield optimization, sustainable agriculture, data-driven agriculture.

УДК 004.8+004.048

# СПОСОБ ПРЕДОСТАВЛЕНИЯ ДАННЫХ V2X ДЛЯ ДАЛЬНЕЙШЕЙ ОБРАБОТКИ СИСТЕМАМИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

***В.В. Дрозд***
*Объединенный институт машиностроения НАН Беларуси*
*Научный сотрудник*
veronika.litarovich@gmail.com

***Е.А. Холод***
*Объединенный институт машиностроения НАН Беларуси*
*Научный сотрудник*
egor930301@gmail.com

***В.В. Гоза***
*Объединенный институт машиностроения НАН Беларуси*
*Младший научный сотрудник*
mechenistalker@gmail.com

***Я.В. Кипцевич***
*Объединенный институт машиностроения НАН Беларуси*
*Техник*
yan.kipcevich@gmail.com

***В.В. Дрозд***
*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием технологий взаимосвязи между высокоавтоматизированными транспортными средствами и прилегающей инфраструктуры.*

***Е.А. Холод***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием методов и способов повышения уровня автоматизации транспортных средств.*

***В.В. Гоза***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием бортовых систем для высокоавтоматизированных транспортных средств.*

***Я.В. Кипцевич***
*Студент Белорусского государственного университета информатики и радиоэлектроники. Область интересов связана с программированием микроконтроллеров.*

**Аннотация.** Проведен анализ способов подготовки большого объема данных для последующей обработки системами искусственного интеллекта. Выделены критерии отбора и подготовки данных для их передачи системам искусственного интеллекта. Определены категории и выполнена классификация коммуникационных сообщений V2X. Реализована выборка больших данных V2X согласно установленным критериям. Получены фреймы данных, которые обеспечивают предоставление актуальной информации о дорожных условиях и участниках дорожного движения.

**Ключевые слова:** высокоавтоматизированное транспортное средство, искусственный интеллект, V2X, обработка данных, предоставление данных.

**Введение.** В последние годы одним из ключевых направлений в автомобильной промышленности является развитие технологий автономных транспортных средств (ТС). Все больше получают применение системы, ориентированные на повышение уровня автоматизации ТС, развивается инфраструктура технологии V2X (Vehicle-to-Everything) [1]. Вместе с этим увеличивается объем маршрутизируемых данных, как в самом

автомобиле, так и между ТС и окружающей инфраструктурой. Что в свою очередь способствует возникновению такой проблемы как необходимость сбора, фильтрации и обработки больших массивов данных в режиме реального времени.

Развитие искусственного интеллекта (ИИ) в последние годы сделало очень большой рывок. Системы ИИ внедряются во все сферы жизни начиная от примитивных систем, используемых в бытовых приборах, заканчивая транспортными [2] и другими высокоранговыми системами обладающими способностями к самообучению.

Передача больших массивов данных системам ИИ позволит значительно сократить время на их обработку. Надежность полученного результата зависит от качества информации предоставленной ИИ. Следовательно, необходимо правильно выполнять процедуры фильтрации и подготовки данных.

**Подготовка данных для предоставления системам ИИ.** Подготовка и форматирование данных являются важными этапами в предоставлении данных ИИ, так как они непосредственно влияют на качество, производительность надежность дальнейших результатов. После анализа предложенных решений от IT-компаний, специализирующихся на разработке технологий в области ИИ [3 - 6], были сформулированы ключевые рекомендации, направленные на обеспечение эффективной и корректной подготовки данных:

1 Изучить данные. Необходимо определиться с типом данных, их характеристикой, каким способом собирались данные, какие цели и ограничения. К примеру, необходимо понимать являются ли данные структурированными или неструктурированными, есть ли в них пропущенные значения и ошибки, присутствует ли маркировка или нет и т.д. Изучение дынных позволит правильно выбрать инструменты, методы и стандарты для форматирования.

2 Очистить данные. Удаление или исправление любых ошибок, несоответствий, которые в дальнейшем могут оказать влияние на конечный результат. Например, удаление дубликатов или ненужных записей, заполнение пропущенных значений, исправление опечаток и т.д. Все это поможет повысить точность, достоверность и удобство использования.

3 Преобразовать данные. Преобразование данных означает приведение данных к тому формату, который подходит для интерпретации ИИ. К примеру, кодировка значений, изменение размера, применение методов извлечения или выбора признаков и т.д. Преобразование данных приведет к уменьшению размерности и избыточности, снизит сложность.

4 Проверить данные. Проверка данных – контроль соответствия требованиям и ожидаемым результатам после обработки ИИ. Например, проверить полноту, последовательность и совместимость данных, правильный формат и т.д. Все это поможет избежать ошибок или сбоев при работе ИИ.

5 Разделить данные. Дробление на различные подмножества для разных целей. К примеру, разделить данные на обучающие, проверочные и тестовые наборы или на различные группы для перекрестной проверки. Поможет оптимизировать производительность ИИ, избежать переобучения или недообучения, а также оценить обобщенность и надежность.

6 Документировать данные. Документирование данных – создание и ведение записей о происхождении, истории, формате и качестве данных. К примеру, допускается документировать источники данных, метод сбора, процесс очистки, этапы преобразования, критерии разделения и т.д. Документирование данных поможет отслеживать изменения, обеспечит воспроизводимость и укажет на ценность и значение.

**Предоставление данных V2X.** В рамках взаимодействия с информационной дорожной средой использование V2X является оптимальным решением для сбора и отслеживания актуальных параметров, характеризующих состояние ТС, дорожной

инфраструктуры и окружающей среды. Однако полный массив данных может достигать больших объемов, что создает технические ограничения по пропускной способности каналов связи, вычислительным ресурсам и времени обработки информации. Для решения этой проблемы предложено передавать часть данных напрямую системам ИИ.

*Изучение данных V2X*. В процессе разработки коммуникационной платформы V2X, одна из решаемых задач была связана с анализом информационных потоков и определением аспектов внутрисистемного информационного взаимодействия [7 - 10]. Для работы с основным набором V2X сообщений использовался стандарт SAE J2735 «V2X Communications Message Set Dictionary» [11]. Чтобы выполнить задачу, связанную с обеспечением данных для систем ИИ, было принято решение классифицировать все перечисленные V2X-сообщения по следующим категориям:

1 Сообщения реестра: MSG_MessageFrame (FRAME).

2 Сообщения безопасности: MSG_BasicSafetyMessage (BSM), MSG_CommonSafetyRequest (CSR), MSG_RoadSideAlert (RSA).

3 Сообщения, содержащие в себе информацию о дорожном движении: MSG_IntersectionCollisionAvoidance (ICA), MSG_MapData (MAP), MSG_SignalPhaseAndTiming (SPAT), MSG_SignalRequestMessage (SRM), MSG_SignalStatusMessage (SSM), MSG_TravelerInformation Message (TIM).

4 Сообщения, содержащие информацию о работе оперативных служб: MSG_EmergencyVehicleAlert (EVA).

5 Сообщения, участвующие в обмене информации между участниками дорожного движения: MSG_CooperativeControlMessage (CCM), MSG_ManeuverSharingAndCoordinatingMessage (MSCM), MSG_PersonalSafetyMessage (PSM), MSG_PersonalSafetyMessage2 (PSM2).

6 Информационно-технические сообщения: MSG_NMEAcorrections (NMEA), MSG_ProbeDataConfigMessage (PDC), MSG_ProbeDataManagement (PDM), MSG_ProbeDataReportMessage (PDR), MSG_ProbeVehicleData (PVD), MSG_RTCMcorrections (RTCM), MSG_TollAdvertisementMessage (TAM), MSG_TollUsageAckMessage (TUMack), MSG_TollUsageMessage (TUM), MSG_TestMessages, MSG_RoadUserChargingConfigMessage (RUCCM), MSG_ RoadUserChargingReportMessage (RUCRM).

7 Сообщения, содержащие информацию о погодных и дорожных условиях: MSG_RoadWeatherMessage (RWM).

8 Зарезервированные сообщения: MSG_RoadGeometryAndAttributes (RGA), MSG_RoadSafetyMessage (RSM), MSG_SensorDataSharingMessage (SDSM), MSG_ SignalControlAndPrioritizationRequest (SCPR), MSG_SignalControlAndPrioritizationStatus (SCPS), MSG_TrafficSignalPhaseAndTiming (TSPaT).

После проведения анализа представленных сообщений, были определены и отобраны фреймы данных, подходящие под следующие критерии:

1 Данные передающиеся по CAN-шине от встроенных систем ADAS.

2 Данные, принимаемые от прилегающей инфраструктуры и имеющие отношение непосредственно к определенному ТС.

3 Данные, которые относятся к определенным регионам с известной топологией, геометрией дорог и перекрестков.

Полученные фреймы данных представлены в таблице 1.

Таблица 1. Фреймы данных

| Наименование | Сообщение | Состав |
|---|---|---|
| AccelerationSet4Way | BSM | Набор значений ускорений ТС в трех ортогональных направлениях (продольном, поперечном и вертикальном) вместе с угловыми скоростями вращения вокруг оси крена. |

| | | |
|---|---|---|
| AccelSteerYawRate_Confidence | BSM | Надежность измерений: ускорения, угла поворота рулевого колеса и положения рулевой системы, скорости вращения вокруг вертикальной оси. |
| AdvisorySpeed | SPAT MAP | Рекомендуемая скорость приближения к перекрестку: значение скорости или диапазон скоростей, точные координаты перекрестка, тип ТС, рекомендации по времени и выполняемым действиям. |
| ApproachOrLane | BSM MAP | Определение местоположения ТС относительно системы индексации, используемой в V2X – картах. |
| BrakeSystemStatus | SPAT BSM | Текущая активность тормозной системы и систем управления ТС. Состоит из флагов состояния для различных компонентов тормозной системы. |
| BSMcoreData | BSM | Минимальный набор данных, необходимых для корректной работы системы безопасности. |
| Circle | RSA MAP | Координаты центра круга, координаты радиуса действия знака и дополнительные параметры, например временные ограничения. |
| ConfidenceSet | SPAT BSM | Значения показывают степень достоверности различных параметров, таких как положение, скорость, направление движения и ускорение. |
| ConnectionManeuver_Assist | SPAT | Информация о динамическом потоке транспортного движения для полос и маневров. |
| DDateTime | RWM MAP | Единый стандартный формат представления даты и времени. |
| EmergencyDetails | EVA | Данные о ТС и его состоянии: тип инцидента, состояние водителя, пассажиров, наличие пострадавших, тип ТС, приоритет реагирования, наличие опасных грузов. |
| EventDescription | EVA BSM | Тип события, описание события, состояние маячков, скорость и направление движения, приоритет. |
| FullPositionVector | PVD | Координаты от ГНСС приёмника, датчики скорости, гироскопы и магнитометры. |
| IntersectionAccessPoint | SSM | Указывает индекс подхода или полосы для проезда. |
| IntersectionGeometry | SPAT MAP | Содержит информацию о всех возможных траекториях движения через перекресток. |
| J1939-Data Items | PDM | Позволяет отправлять информацию о состоянии транспортного средства, такую как уровень топлива, давление шин, температура двигателя и другие параметры, которые обычно доступны через CAN. |
| ObstacleDetection | BSM | Информация о: местоположение препятствия, тип препятствия, размер препятствия, дистанция до препятствия и др. |
| PathHistory | SPAT BSM | Информация о пути движения ТС. |
| PathPrediction | ISA PSM PSM2 | Прогнозируемая траектория пути. |
| Position3D | PSM NMEA | Точное местоположение в системе координат WGS-84. |
| RegulatorySpeedLimit | SPAT MAP | Информация о регулируемой скорости, касающейся одной или нескольких полос движения или сегмента дороги. |
| RequestorDescription RequestorType | SRM PVD | Идентификация выбранного ТС или пользователей. |
| Snapshot SnapshotTime | SRM BSM | Отчет о состоянии одной или нескольких бортовых систем ТС, вместе с набором данных о положении и направлении |

| | | |
|---|---|---|
| | | движения, указывающих место создания отчета. |
| SpecialVehicle_Extensions | SRM<br>BSM | Информация об экстренных службах: состояние маячков (включены/выключены), тип выполненной операции (например, реагирование на аварию или патрулирование), немедленные намерения (например, поворот направо или движение прямо).<br>Для прочих ТС: информация о текущем маршруте, планируемые остановки, параметры груза и др. |
| TransmissionAndSpeed | SRM<br>BSM | Скорость транспортного средства и состояние коробки передач. |
| VehicleClassification<br>VehicleData<br>Vehicle ID | SRM<br>BSM | Тип, габаритные размеры, вес транспортного средства.<br>Роль ТС в дорожной обстановке. Дополнительные параметры ТС. |
| VehicleStatus | BSM | Состояние тормозной системы, двигателя, систем безопасности, световых сигналов. Информация о рулевом управлении. Дополнительные данные о состоянии автомобиля. |
| WeatherReport | RWM<br>BSM | Текущие погодные условия: дождь, снег, град, туман, ветер. Степень интенсивности погодного явления.<br>Значение температуры воздуха в градусах Цельсия или по Фаренгейту. Температура дорожного покрытия. Видимость. Скорость и направление ветра. Дополнительные параметры. |

*Очистка данных V2X*. Очистка данных осуществляется исходя из условий, представленных в стандарте J2735, а также параметров, установленных под конкретную задачу. Например: набор значений ускорений должен удовлетворять следующим требованиям – данные не нулевые и соответствуют заданным значениям; время и дата – значения совпадают с NTP сервером и соответствуют формату предоставления даты и времени; координаты – значения отличные от нулевых и соответствующие системе геодезических параметров WGS-84; скорость – в диапазоне заданных минимальных и максимальных значений; флаги состояний – отсутствие пустых и недостоверных флагов.

*Преобразование данных V2X*. Преобразование данных осуществляется путем кодирования и шифрования. IEEE 1609.2 определяет криптографические службы для обеспечения безопасности данных. Службы включают подписание и проверку данных для аутентификации и целостности, а также шифрование и дешифрование с использованием асимметричных публичных ключей. IEEE 1609.2 также указывает, имеет ли пользователь V2X-приложения право отправлять определенные объекты данных через коммуникационный интерфейс. Рекомендуемый подход к проектированию определен в SAE J2945/5. Элементы данных кодируются с использованием ASN.1.

При развертывании V2X инфраструктуры в новых регионах возможно использовать другие стандарты шифрования и кодирования, при соответствующем запросе и предъявляемых требований от местных телекоммуникационных служб, провайдеров и операторов. Использование других типов кодирования допустимо, но требует тщательной оценки последствий для безопасности, эффективности и совместимости.

*Проверка данных V2X*. Проверка данных осуществляется для того, чтобы исключить возможность использования недостоверных данных и гарантировать, что все участники экосистемы получают актуальную информацию. Некоторые примеры способов проверки информации:

1 Бортовые устройства. Для осуществления проверки каждое бортовое устройство имеет свои уникальные данные (сертификат, ID, подпись и т.д.), которые используются для подтверждения его легитимности.

2 Координаты транспортного средства. Система сопоставляет координаты с известными картами дорог. Алгоритмы проверки используют данные из архивов, чтобы выявить аномалии и несоответствия, отличия от статистики.

3 Дата и время. Проверяется актуальность данных, чтобы исключить использование устаревшей информации.

4 Данные от участников V2X. Сравниваются данные от разных устройств для выявления противоречий.

5 Общие данные. Выполняется проверка на соответствие данных логике. Пример: от ТС оборудованного двигателем внутреннего сгорания пришло сообщение о том, что зарядка завершена.

*Группировка данных V2X*. Группы данных по типу (например, тестовая, проверочная, обучающая и др.) и вариант проверки определяются исходя из поставленной задачи и конечных результатов, которые необходимо получить после обработки системами искусственного интеллекта.

*Документирование данных V2X*. Этапом документирования можно считать все документы, которые оформляются в процессе всей подготовки данных к передачи их системам искусственного интеллекта.

**Заключение.** В результате проделанной работы можно сделать следующие выводы:

1 Большие объемы данных V2X требуют тщательного отбора ключевых параметров, которые имеют наибольшую значимость для задач ИИ. Это позволяет минимизировать нагрузку на каналы связи и вычислительные ресурсы, сохраняя при этом необходимую информативность.

2 Использование стандартов обеспечивает совместимость и унификацию процесса передачи данных между различными участниками дорожного движения в режиме реального времени. Применение методов фильтрации и преобразования данных повышают уровень информационной безопасности систем ИИ.

3 Классификация V2X-сообщений по категориям помогает определить наиболее важные типы данных для каждой задачи. Фреймы данных, полученные в результате выборки и классификации, обеспечивают актуальную информацию о дорожной среде и состоянии участников дорожного движения.

4 Подготовка данных является одним из самых важных этапов в разработке систем ИИ для транспортной среды. Она напрямую влияет на точность прогнозирования и производительность системы.

Интеграция систем ИИ с V2X-технологиями открывает широкие возможности для создания интеллектуальных решений в транспорте.

## Список литературы

[1] Automotive V2X Market: Global Industry Analysis And Forecast (2023-2029) // maximizemarketresearch [Electronic resource]. – Mode of access: https://www.maximizemarketresearch.com/market-report/global-v2x-market/9764/. – Date of access: 24.02.2025.

[2] Artificial Intelligence in Automotive Market Insights // zionmarketresearch [Electronic resource]. – Mode of access: https://www.zionmarketresearch.com/report/artificial-intelligence-automotive-market/. – Date of access: 24.02.2025.

[3] Data Preparation for Machine Learning: The Ultimate Guide to Doing It Right // PecanAI [Electronic resource]. – Mode of access: https://www.pecan.ai/blog/data-preparation-for-machine-learning/. – Date of access: 24.02.2025.

[4] How to Prepare Data for AI: Essential Steps and Tips // Datastax [Electronic resource]. – Mode of access: https://www.datastax.com/guides/how-to-prepare-data-for-ai/. – Date of access: 24.02.2025.

[5] Data Preparation for AI Initiatives: The Essential Steps // Addepto [Electronic resource]. – Mode of access: https://addepto.com/blog/data-preparation-for-ai-initiatives-the-essential-steps/. – Date of access: 24.02.2025.

[6] A Guide to Preparing Organizational Data for AI // CoreBTS [Electronic resource]. – Mode of access: https://corebts.com/blog/a-guide-to-preparing-organizational-data-for-ai/. – Date of access: 24.02.2025.

[7] Информационное взаимодействие центрального коммутационного блока высокоавтоматизированного транспортного средства с коммуникационной платформой V2X. / Е.А. Холод, А.Д. Крупок, М.В. Исаевич. // Актуальные вопросы машиноведения / ВЫПУСК 10 Минск, 2021 – С. 41 - 44.

[8] Анализ информационных потоков в коммуникационной платформе C-V2X / В.В. Литарович, Е.В. Мыльников, В.В. Савченко, М.А. Чернин // Актуальные вопросы машиноведения: сб. науч. тр. / Объедин. ин-т машиностроения НАН Беларуси; редкол.: С.Н. Поддубко [и др.]. – 2019. – Вып. 8. – С. 145-147.

[9] Анализ информационных потоков в DSRC для высокоавтоматизированных транспортных средств / В.В. Литарович, В.В. Савченко, Е.В. Мыльников, С.О. Копыток // Актуальные вопросы машиноведения: сб. науч. тр. / Объедин. ин-т машиностроения НАН Беларуси; редкол.: С.Н. Поддубко [и др.]. – 2021. – Вып. 10. – С. 54-57.

[10] Консолидация данных в коммуникационной платформе для высокоавтоматизированных транспортных средств / В.В. Савченко, В.В. Литарович // Труды НГТУ им. Р.Е. Алексеева. 2022. №2. С. 105-114. DOI: 10.46960/1816-210X_2022_2_105

[11] SAE standard "V2X Communications Message Set Dictionary" № J2735_202309 // SAE International. V2X Core Technical Committee. 2023-09-22 г. № https://doi.org/10.4271/J2735_202309.

## Авторский вклад

**Дрозд Вероника Витальевна** – анализ способов подготовки данных для дальнейшей обработки системами искусственного интеллекта, выделение критериев для подбора и подготовки данных.

**Холод Егор Александрович** – анализ данных, представленных в стандарте SAE J2735 «V2X Communications Message Set Dictionary», группировка сообщений.

**Гоза Виталий Валентинович** – постановка критериев и выборка по ним данных для дальнейшей обработки системами искусственного интеллекта.

**Кипцевич Ян Витальевич –** анализ способов подготовки данных в сфере кодирования и шифрования.

# METHOD FOR PROVIDING V2X DATA FOR FURTHER PROCESSING BY ARTIFICIAL INTELLIGENCE SYSTEMS

***V.V. Drozd***
*Joint Institute of Mechanical Engineering of the National Academy of Sciences of Belarus*
*Researcher*

***E.A. Holod***
*Joint Institute of Mechanical Engineering of the National Academy of Sciences of Belarus*
*Researcher*

***V.V. Hoza***
*Joint Institute of Mechanical Engineering of the National Academy of Sciences of Belarus*
*Junior Researcher*

***Y.V. Kiptsevich***
*Joint Institute of Mechanical Engineering of the National Academy of Sciences of Belarus*
*Technician*

**Abstract.** An analysis of methods for preparing data for subsequent processing by artificial intelligence systems was conducted. Criteria for selecting and preparing data for their transmission to AI systems have been identified. Categories have been defined, and V2X communication messages were classified. V2X data sampling was implemented according to the established criteria. Received data frames that provide up-to-date information on road conditions and users.

**Keywords**: highly automated vehicle, artificial intelligence, V2X, data processing, data provision.

УДК 004.8:004.9

# КЛАССИФИКАЦИЯ ГЛАЗНЫХ ЗАБОЛЕВАНИЙ ПО СНИМКАМ ГЛАЗНОГО ДНА С ПОМОЩЬЮ НЕЙРОННОЙ СЕТИ EFFICIENTNET

***К.И. Свистунова***
*Студентка кафедры информационных систем управления ФПМИ БГУ*
kseniyasvistunova@icloud.com

***С.В. Абламейко***
*Профессор БГУ*
*Доктор технических наук,*
*Академик НАН Беларуси,*
*Гл. научный сотрудник ОИПИ*
ablameyko@bsu.by

***К.И. Свистунова***

*Является студенткой четвертого курса Белорусского государственного университета. Область научных интересов связана с решением практических задач в области машинного обучения.*

***С.В. Абламейко***

*Является профессором БГУ, гл. научным сотрудником ОИПИ. Область научных интересов связана с разработкой методов и алгоритмов для решения задач обработки изображений и распознавания образов.*

**Аннотация.** Предложен алгоритм распознавания заболеваний глаз человека по снимкам глазного дна. В результате проведенного анализа была выбрана сверточная нейронная сеть EfficientNet и на ее основе разработан алгоритм, использующий архитектуры EfficientNetB6 и EfficientNetB0. Проведены экспериментальные исследования, которые показали высокую эффективность данного подхода для диагностики заболеваний глаз.

**Ключевые слова:** нейронные сети, EfficientNet, аугментация, анализ изображений, распознавание образов, классификация, изображения глазного дна.

**Введение.** Практически каждый человек в своей жизни сталкивается с проблемой болезни глаз – это любое состояние или расстройство, которое нарушает способность глаза правильно работать или оказывает вредное воздействие на остроту зрения. Болезни глазного дна являются основной причиной слепоты у людей, поэтому своевременная диагностика на начальных стадиях крайне важна для их лечения, так как позволяет предотвратить дальнейшее прогрессирование заболеваний.

Ручная диагностика глазных болезней по снимкам глазного дна на ранних этапах может представлять собой значительную трудность. В связи с этим целесообразно разработать автоматизированную систему для обнаружения заболеваний глаз с помощью компьютерных технологий, которая будет способна выявлять различные нарушения, основываясь на изображениях глазного дна.

Алгоритмы, основанные на глубоком обучении, становятся все более распространенными в анализе медицинских изображений. Исследования показали, что такие модели эффективно справляются с задачами, связанными с обнаружением объектов, классификацией медицинских изображений и выявлением заболеваний.

К настоящему времени опубликовано достаточно много работ по анализу глазных (офтальмологических) изображений. В [2] предложен метод автоматизации анализа

флюоресцентных ангиограмм глазного дна человека, а в [3] – методы для оценки состояния липидного слоя в интермаргинальном пространстве века глаза человека, анализа степени плотности клеточными структурами (целлюлярности) роговичной ткани глаза человека, оценки состояния ретинального кровотока при анализе флюоресцентных ангиограмм глазного дна человека.

В докладе приведен результат анализа эффективности нейронной сети EfficientNet для распознавания различных заболеваний глаз.

**Выбор сверточной нейронной сети.** EfficientNet - это семейство сверточных нейронных сетей (CNN), целью которого является достижение высокой производительности с меньшими вычислительными ресурсами по сравнению с предыдущими архитектурами. Оно было представлено Mingxing Tan и Quoc V. Ле из Google Research в своей статье 2019 года «EfficientNet: переосмысление масштабирования модели для сверточных нейронных сетей». Основной идеей EfficientNet является новый метод масштабирования, называемый составным коэффициентом, для масштабирования моделей простым, но эффективным способом. Вместо случайного масштабирования ширины, глубины или разрешения, составное масштабирование равномерно масштабирует каждое измерение с определенным фиксированным набором коэффициентов масштабирования. Этот подход гарантирует, что модель легко адаптируется к различным вычислительным ограничениям, сохраняя при этом свою производительность в различных масштабах и задачах. Используя данный метод масштабирования, авторы разработали семь моделей различных размеров, которые превосходили самую современную точность большинства сверточных нейронных сетей и с гораздо лучшей эффективностью [8].

**Исходный набор данных.** Выбранный датасет представляет собой набор данных, доступный на Kaggle, для выявления глазных заболеваний. Он содержит снимки глаз более пяти тысяч пациентов. Набор данных представляет собой цветные снимки глазного дна левого и правого глаза, которые размечены ключевыми диагнозами, поставленными врачами. Врачи классифицируют болезни глаз по восьми категориям: нормальный(здоровый) (Н), диабет (Д), глаукома (Г), катаракта (С), гипертония (Н), патологическая близорукость (М), возрастная макулярная дегенерация (А), другие заболевания/аномалии (О).

**Работа с данными.** В поставленной задаче было принято решение рассматривать пары классов (здоровые и заболевание) для выявления классификации заболевания.

Для начала было рассмотрено количество снимков, соответствующие каждому классу, для анализа того, какую подготовку данных следуют совершить для решения задачи. В таблице 1 представлено количество снимков, содержащих определенное заболевание.

В таблице 1 показано, что при рассмотрении любого класса заболевания в сравнении со здоровым, у нас будет несбалансированный набор данных. Для решения этой проблемы было решено применять аугментацию на лету. Данный метод был использован для того, чтобы каждый раз на обучение поступали изображения, которые модель еще не видела. Это помогает улучшить обобщающую способность модели, снизить вероятность переобучения и делает её более устойчивой к изменениям в данных. Изучив изображения, на которых модель ошибается, было принято решение использовать способы аугментации, такие как поворот изображений на случайный угол, смещение изображения по ширине, смещение изображения по высоте. Для заполнения новых пикселей, которые могут возникнуть в результате трансформаций (например, при повороте или изменении размера) используется "nearest" (ближайший сосед). Это означает, что новые пиксели будут заполнены значениями ближайших пикселей.

Таблица 1. Количество снимков согласно заболеваниям

| № п/п | Заболевание | Количество |
|---|---|---|
| 1 | Здоровый | 5501 |
| 2 | Катаракта | 594 |
| 3 | Диабетическая ретинопатия | 165 |
| 4 | Миопия | 457 |
| 5 | Гипертензия | 382 |
| 6 | Возрастная макулярная дегенерация | 551 |
| 7 | Глаукома | 616 |

**Обучение нейронной сети.** Для решения задачи были выбраны EfficientNetB6 и EfficientNetB0 для выявления того, как изменение размера изображения повлияет на точность определения глаза с заболеванием. Для изображений размера 224x224 применим EfficientNetB0, для изображений 512x512 (исходный размер изображений) - EfficientNetB6.

Данные из набора преобразуем в массивы изображений и меток. Разделяем их на обучающую и тестовую выборки, где 20 % данных выделяем для тестирования. На следующем этапе создаем генератор данных для выполнения аугментации, что помогает улучшить обобщающую способность модели за счет разнообразия обучающих данных.

Создаем экземпляры моделей, использующих предварительно обученные веса на наборе данных ImageNet. Указание параметра include_top=False исключает верхние полностью связанные слои, input_shape=(image_size, image_size, 3) задает форму входных изображений. Создаем последовательную модель с помощью Sequential из TensorFlow, которая позволяет добавлять слои последовательно друг за другом. Добавление предварительно обученной модели к последовательной модели позволяет использовать архитектуру и веса модели как часть нашей конечной модели. К модели EfficientNetB6 добавляем слой GlobalAveragePooling2D, который выполняет глобальное усреднение по каждому из пространственных размеров входных данных. Он уменьшает пространство признаков, создавая вектор признаков фиксированной длины из выходных данных EfficientNetB6, и устраняет избыточность, что может помочь предотвратить переобучение. Также добавляем слой Dense - полносвязный слой с 256 нейронами. Активация ReLU (Rectified Linear Unit) применяется для введения нелинейности в модель, что позволяет ей обучаться более сложным функциям. Добавляем полносвязный слой Dense с одним нейроном и функцией активации sigmoid. В данном случае используется один нейрон с активацией sigmoid, так как задача является бинарной классификацией. Модель компилируется с использованием оптимизатора Adam, функции потерь binary_crossentropy и метрики точности.

Модель обучается с использованием data_gen.flow(X_train, y_train, batch_size=16), который создает генератор данных для подачи изображений и меток в модель по партиям (batch). Это полезно для обработки больших наборов данных, так как позволяет загружать данные по частям в память. Обучение происходит в течение 15 эпох. Для предотвращения переобучения используются механизмы ранней остановки и сохранения лучшей модели.

**Результаты классификации.** Для оценки эффективности предложенного алгоритма были проведены экспериментальные исследования на наборе данных изображений глазного дна. Нейронная сеть EfficientNet была обучена и протестирована. В таблице 2 представлены результаты классификации для разных типов заболеваний.

Таблица 2. Результаты классификации с помощью EfficientNet для разных типов заболеваний

| Заболевание | Точность EfficientNetB0 (224x224) | Точность EfficientNetB6 (512x512, size of images) |
|---|---|---|
| Миопия | 0.9790576100349426 | 0.9947643876075745 |
| Катаракта | 0.9663865566253662 | 0.9729729890823364 |
| Глаукома | 0.9211618304252625 | 0.9300000071525574 |
| Возрастная макулярная дегенерация | 0.9497717022895813 | 0.9674418568611145 |
| Диабетическая ретинопатия | 0.9305555820465088 | 0.9166666865348816 |
| Гипертензия | 0.9044585824012756 | 0.9426751732826233 |

Лучшие результаты показывает сеть EfficientNetB6 на изображениях исходного размера. Это показывает, что сохранение размера изображений в решении данной задачи важно и необходимо, так как происходит меньший процент потери данных. Например, в диагностике возрастной макулярной дегенерации точность EfficientNet значительно увеличилась с 0.9498 (224x224) до 0.9674 (512x512), показывая, что размер изображения важен для повышения точности.

Лучшие результаты наблюдаются в выявлении миопии и катаракты, что может быть связано с большим отличием в изображениях здорового глаза и глаза, имеющего данные заболевания.

Для оценки работы были использованы следующие метрики:

*Precision (точность)* — это доля объектов, названных классификатором положительными и при этом действительно являющимися положительными. Высокие показатели Precision говорят:

− о высокой чувствительности модели − она хорошо обнаруживает случаи с положительным диагнозом;

− о низком уровне ложноотрицательных результатов.

*Recall (полнота)* — это то, какую долю объектов положительного класса из всех объектов положительного класса нашёл алгоритм. Высокие показатели Recall говорят:

− о высокой чувствительности модели - она хорошо обнаруживает случаи с положительным диагнозом;

− о низком уровне ложноотрицательных результатов.

*F1-score* — одна из метрик, используемых для оценки качества бинарных классификаторов, которая учитывает, как точность (Precision), так и полноту (Recall). Эта метрика особенно полезна в случаях, когда важно находить баланс между этими двумя показателями, и, соответственно, когда классы имеют неравное количество примеров или, когда неравномерные ошибки имеют разные последствия. Например, в диагностике заболеваний, пропуск положительного случая (ложный отрицательный) может иметь серьезные последствия. В этом случае высокая полнота более критична, и F1-score позволяет учитывать это.

В Таблице 3 представлены показатели recall, precision и F1-score для каждого класса.

Таблица 3. Оценки работы EfficientNetB6 для различных классов

| | Класс | Precision | Recall | F1 score |
|---|---|---|---|---|
| Здоровый и миопия | 0 | 1.00 | 0,99 | 0,99 |
| | 1 | 0,99 | 1.00 | 0,99 |
| Здоровый и катаракта | 0 | 0.98 | 0.96 | 0.97 |
| | 1 | 0.97 | 0.98 | 0.97 |

| | | | | |
|---|---|---|---|---|
| Здоровый и диабетическая ретинопатия | 0 | 0.92 | 0.92 | 0.92 |
| | 1 | 0.91 | 0.91 | 0.91 |
| Здоровый и глаукома | 0 | 0.92 | 0.94 | 0.93 |
| | 1 | 0.94 | 0.92 | 0.93 |
| Здоровый и ВМД | 0 | 0.97 | 0.96 | 0.96 |
| | 1 | 0.97 | 0.97 | 0.97 |
| Здоровый и гипертензия | 0 | 0.95 | 0.94 | 0.95 |
| | 1 | 0.93 | 0.94 | 0.94 |

Для анализа того, какие изображения были распознаны неверно, была прописана функция вывода изображений, на которых метка классов не совпадает с предсказанной меткой. Результат работы такой функции для класса Диабетической ретинопатии представлен на рисунке 1.

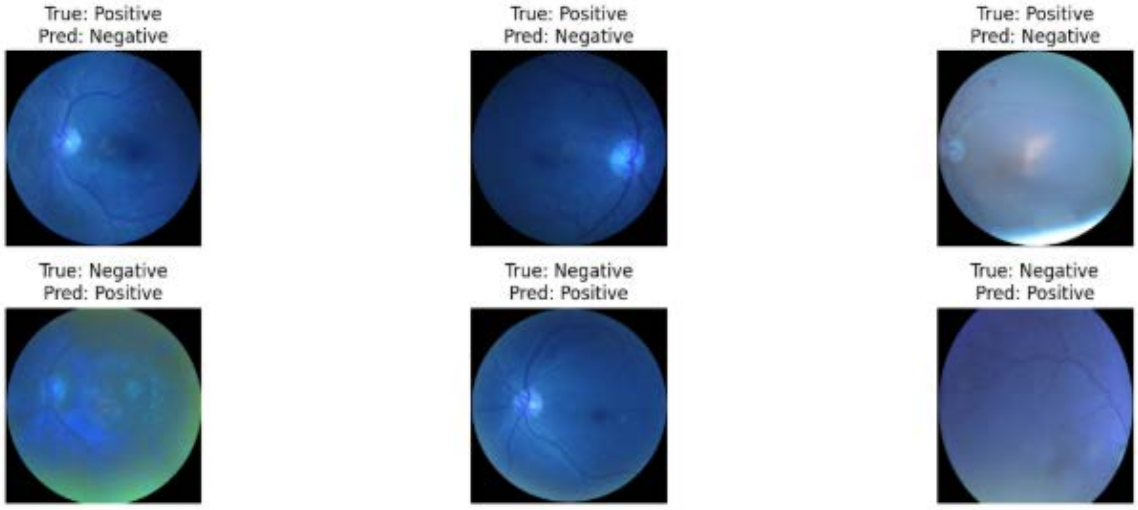



*Рисунок 1.* Изображения для класса Диабетическая ретинопатия, для которых истинные метки не совпали с предсказанными

Визуальный анализ ошибочно классифицированных изображений может помочь выявить общие паттерны среди неверно классифицированных изображений, такие как нечеткость снимков, присутствие шумов и артефактов. Анализируя результаты, можно понять, какие виды аугментации данных наиболее эффективны для улучшения производительности модели и уменьшения количества ошибок. Результаты визуального анализа могут быть использованы для разработки стратегии дальнейшего улучшения модели, включая создание новых наборов данных, улучшение предобработки и аугментации.

**Заключение.** В докладе рассмотрена нейронная сеть EfficientNet для анализа изображений глазного дна. Предложен алгоритм распознавания заболеваний глаза человека по снимкам глазного дна. Также приведены результаты экспериментальных исследований.

## Список литературы

[1] Волков, Е. Н. Gradient-based методы объяснительного искусственного интеллекта для классификации заболеваний глаза / Е. Н. Волков, А. Н. Аверкин // IV Международная конференция по нейронным сетям и нейротехнологиям (NeuroNT'2023) : сб. докладов, СПб., 16 июня 2023 г. – СПб. : СПбГЭТУ «ЛЭТИ», 2023. – С. 7–10.

[2] Gurevich I. B, Yashina V. V., Ablameyko S. V., Nedzved A. M., Ospanov A. M., Tleubaev A. T., Fedorov A. A., Fedoruk N. A. Development and experimental investigation of mathematical methods for automating the

diagnostics and analysis of ophthalmological images [Electronic resource]. – Pattern Recognit Image Anal. – 2018. – № 28(4). – P. 612–636. – Mode of access: https://doi.org/10.1134/S1054661818040120. – Date of access: 01.08.2024.

[3] Гуревич И. Б., Яшина В. В., Абламейко С. В., Недзведь А. М., Оспанов А. М., Тлеубаев А. Т., Федорук Н. А. Новый математический метод автоматизации анализа флюоресцентных ангиограмм глазного дна человека [Электронный ресурс]. – Вестник офтальмологии. – 2021. – № 137(3). – P. 49–57. – Режим доступа: https://doi.org/ 10.17116/oftalma202113703149. – Дата доступа: 01.08.2024.

[4] Sarki, R. Automatic detection of diabetic eye disease through deep learning using fundus images: a survey / R. Sarki [et al.]. – IEEE Access. – 2020. – Vol. 8. – P. 151133–151149. – DOI: 10.1109/ACCESS.2020.3015258.

[5] Ting, D. Artificial intelligence and deep learning in ophthalmology / D. Ting [et al.]. – Br J Ophthalmol. – 2019. – Vol. 103. – P. 67–75. – DOI: 10.1136/bjophthalmol-2018-313173.

[6] Rahimy, E. Deep learning applications in ophthalmology / E. Rahimy. – Curr Opin Ophthalmol. – 2018. – Vol. 29. – P. 54–60. – DOI: 10.1097/ICU.0000000000000470.

[7] Deep learning in ophthalmology: a review / P. S. Grewal. – Can J Ophthalmol. – 2018. – Vol. 53(4). – P. 309–313. – DOI: 10.1016/j.jcjo.2018.04.019.

[8] EfficientNet paper — Tan, M., & Le, Q. v. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. 36th International Conference on Machine Learning, ICML 2019, 2019-June.

# CLASSIFICATION OF EYE DISIASES ON FUNDUS IMAGES USING EFFICIENTNET NEURAL NETWORK

***K.I. Svistunova***
*Student of the Department of Information Management Systems of BSU*

***S.V. Ablameyko***
*Professor of BSU, DSc of Technical sciences, Academician of the National Academy of Sciences of Belarus, Senior Researcher of the UIIP*

**Abstarct.** The paper presents an algorithm for recognizing human eye diseases from fundus images. Based on the analysis, the convolutional neural network EfficientNet is selected and the algorithm based on EfficientNetB6 and EfficientNetB0 is proposed. Experimental studies have shown high efficiency for diagnosing eye diseases.

**Keywords:** neural networks, EfficientNet, augmentation, image analysis, pattern recognition, classification, fundus images.

УДК 004.021:004.75

# АППАРАТНАЯ ПОДДЕРЖКА ОБРАБОТКИ ДАННЫХ В ИТ-ДИАГНОСТИКЕ

***В.А. Вишняков***
*Доктор технических наук, профессор кафедры инфокоммуникационных технологий БГУИР*
vish@bsuir.by

***И. Ся***
*Аспирант кафедры инфокоммуникационных технологий БГУИР*
Xiayiwei4@gmail.com

***В.А. Вишняков***
*Область профессиональных интересов/исследований: управление и безопасность в инфокоммуникациях, Интернет вещей,блокчейн, электронный маркетинг, ИТ в образовании и медицине, интеллектуальные системы управления.*

***И. Ся***
*Аспирант кафедры инфокоммуникационных технологий Белорусского государственного университета информатики и радиоэлектроники.*

**Аннотация.** Болезнь Паркинсона – это распространённое неврологическое заболевание. Для его диагностики и мониторинга используют много данных: голос, походку, нейроизображения. Обычные вычисления на CPU не всегда подходят, так как они медленные и потребляют много энергии. Поэтому используют ускорители, такие как FPGA, Vitis HLS, GRU. В докладе рассматривается архитектура обработки данных движения на FPGA. Использован Vitis HLS для аппаратной реализации обработки движений при болезни Паркинсона, также применено вейвлет-преобразование для выделения признаков. Для анализа данных используемся сеть GRU. Предложенный метод даёт большую скорость, меньшие задержки и меньшее энергопотребление по сравнению с обычными CPU.

**Ключевые слова: болезнь** Паркинсона, FPGA, Vitis HLS, GRU, вейвлет-преобразование.

**Введение.** Болезнь Паркинсона [1] – одно из распространённых неврологических заболеваний, сопровождающееся нарушениями движений. Анализ таких данных медленен из-за их большого объёма, необходимости высокой скорости обработки и ограничений по энергопотреблению. Для решения этих проблем в работе предложено использовать FPGA [2], что обеспечивает параллельную обработку при низком энергопотреблении, что является идеальным для носимых медицинских устройств.

Разработка упрощается с помощью Vitis HLS, который преобразует код C/C++ в аппаратную реализацию и поддерживает интерфейсы AXI. Для анализа временных данных применяется нейронная сеть GRU [3] – облегчённая версия LSTM, обеспечивающая эффективную обработку движений. Дополнительно используется вейвлет-преобразование [4], позволяющее выявлять особенности походки и тремора. В результате разработана FPGA-система, объединяющая Vitis HLS, GRU и вейвлет-преобразование, что даёт высокую скорость, низкое энергопотребление и точную диагностику.

**Методологическая основа разработки.** Необработанные данные с датчиков движения пациента представляют собой временные ряды с нестабильностью, что

затрудняет прямое выделение признаков. Для их обработки применяется метод скользящего окна (размер – 32, шаг – 16, перекрытие – 50 %), позволяющий сгладить флуктуации и унифицировать формат данных.

Выделение признаков данных осуществляется с помощью дискретного вейвлет-преобразования (DWT) на основе вейвлетов Daubechies 6 (db6). В связи с отсутствием встроенных функций в стандартной библиотеке C++, алгоритм программируется вручную. Обработка включает три этапа: получение коэффициентов фильтров, одноуровневое DWT и многоуровневую декомпозицию.

После разложения из каждой из четырёх ступеней выделяются пять ключевых признаков[5]:

1 Вейвлет-энтропия – характеризует степень неопределённости сигнала;

2 Вейвлет-энергия – отражает концентрацию мощности сигнала;

3 Длина вейвлет-сигнала – оценивает сложность формы сигнала;

4 Дисперсия вейвлет-коэффициентов – характеризует разброс значений сигнала;

5 Стандартное отклонение вейвлет-коэффициентов– отражает уровень стабильности сигнала.

Поскольку каждый сенсор фиксирует данные по трём осям (X, Y, Z), итоговое количество признаков для одного окна составляет 75 (5 признаков × 5 уровней разложения × 3 оси).

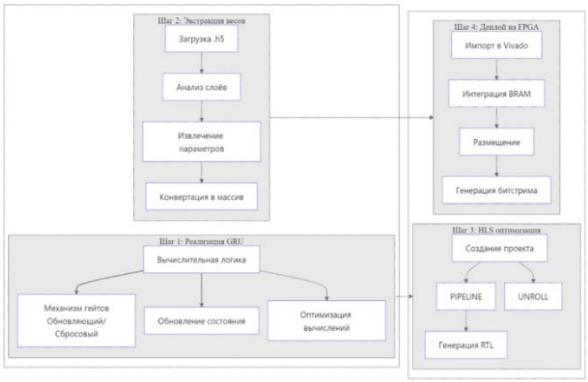



*Рисунок 1.* Оптимизация и развертывание GRU на FPGA

Поскольку FPGA не может напрямую нейросеть, обученную в Python, модель сначала переводится в C++, затем выполняется высокоуровневый синтез (HLS) с помощью Vitis HLS, после чего генерируется код Verilog/VHDL, Как показано на рисунке 1, предложенный метод позволяет адаптировать GRU для аппаратного выполнения на FPGA.

Для повышения производительности и уменьшения использования ресурсов FPGA применяются следующие элементы оптимизацииоптимизации.

1 Хранение весов в статических массивах – поскольку FPGA не поддерживает файловый ввод-вывод (I/O), все веса сохраняются в BRAM или LUT.

2 Оптимизация вычислений с фиксированной запятой вместо арифметики с плавающей запятой используется ap_fixed, что снижает нагрузку на аппаратные ресурсы.

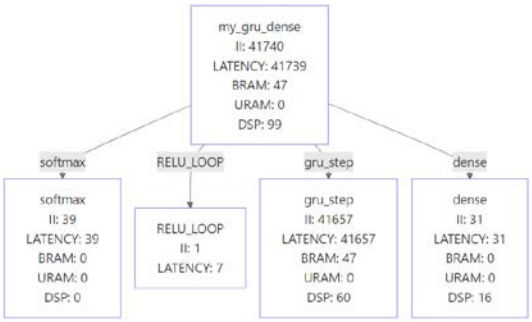



*Рисунок 2.* Иерархическая структура HLS для вычислений GRU-инференса

3 Разворачивание циклов (Loop Unrolling) – ускоряет матричное умножение за счёт увеличения параллелизма.

4 Конвейеризация (Pipeline) – улучшает поток обработки данных, повышая пропускную способность и эффективность вычислений.

**Экспериментальная верификация.** Представлена верификация разработанных вычислительных модулей, реализованных с использованием HLS, (структура приведена на рисунке 2) для FPGA. Оценка использования аппаратных ресурсов для инференса глубокого обучения (GRU + Dense + Softmax) на FPGA с целью оптимизации вычислительной эффективности и распределения ресурсов. Эксперименты были проведены для проверки корректности работы модулей вычисления признаков на основе вейвлет-преобразования и вычислений GR. Верификация включала функциональное моделирование, анализ аппаратно-программных интерфейсов и проверку соответствия вычислительных результатов ожидаемым значениям.

Для проверки правильности работы модуля вейвлет-признаков была выполнена программная симуляция (CSIM) кода HLS. На рисунке 3 представлен журнал вывода симуляции, в котором фиксируются значения выходных признаков feature_out, вычисленных для различных временных окон (window). Каждое окно соответствует определённому сегменту входного сигнала, а результирующие признаки позволяют анализировать его частотно-временные характеристики. На рисунке 4 представлен результат проверки вычислительного модуля GRU. В ходе симуляции CSIM успешно завершился без ошибок, что подтверждает корректность функциональной реализации алгоритма в HLS.

Полученные результаты демонстрируют, что разработанный модуль успешно синтезирован и может быть далее оптимизирован для аппаратного развертывания. Обе программные модели прошли функциональную проверку на этапе CSIM и показали отсутствие ошибок. Это свидетельствует о следующем.

1 Алгоритмы вейвлет-признаков и GRU корректно функционируют на уровне HLS.

2 Программные интерфейсы успешно сопоставлены с аппаратными ресурсами FPGA.

3 Данные корректно передаются между программной и аппаратной частью через AP_MEMORY и управляющие сигналы.

**Заключение.** В данной работе представлена аппаратная поддержка обработки сигналов на базе FPGA для анализа движений пациентов с болезнью Паркинсона, включающая вейвлет-преобразование для выделения признаков и GRU для обработки временных рядов. Проведённая верификация показала, что предложенные вычислительные модули успешно функционируют на уровне HLS, обеспечивая высокую скорость, низкое энергопотребление и корректность аппаратной реализации.

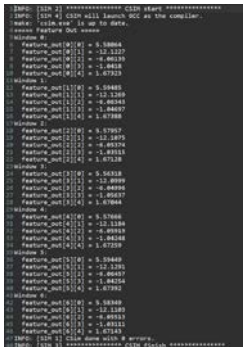


*Рисунок 3.* Проверка вычислительного модуля для извлечения вейвлет-признаков

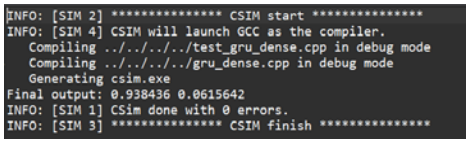



*Рисунок 4.* Проверка вычислительного модуля GRU

## Список литературы

[1] Bloem B.R. Parkinson's disease / B.R. Bloem, M.S. Okun, C. Klein // The Lancet. – 2021. – Vol.397, No.10291. – P. 2284–2303.

[2] Liu F. Review of neural network model acceleration techniques based on FPGA platforms / F. Liu, H. Li, W. Hu, Y. He // Neurocomputing. – 2024. – P. 128511.

[3] Rajmohan R. Exploration of deep RNN architectures: LSTM and gru in medical diagnostics of cardiovascular and neuro diseases / R. Rajmohan, M. Pavithra, T.A. Kumar, P. Manjubala // Handbook of Deep Learning in Biomedical Engineering and Health Informatics. – 2021. – P. 167–202. – Apple Academic Press.

[4] Vimalajeewa D. Parkinson’s disease diagnosis with gait characteristics extracted using wavelet transforms / D. Vimalajeewa, E. McDonald, M. Tung, B. Vidakovic // IEEE Journal of Translational Engineering in Health and Medicine. – 2023. – Vol.11. – P. 271–281.

[5] Vishniakou U.A. It diagnostics of Parkinson's disease based on voice markers and decreased motor activity / U.A. Vishniakou, X. Yiwei // SA&AI. 2023. – Vol. 21. N 2. – P.102–110.

## Авторский вклад

Вишняков В.А. предложил концепцию метода, Ся Ивей реализовал его.

# HARDWARE SUPPORT FOR DATA PROCESSING IN IT DIAGNOSTICS

***V.A. Vishnyakov***
*Doctor of Technical Sciences, Professor of the Department of Infocommunication Technologies, BSUIR*

***Y. Xia***
*PhD student in the Department of infocommunication technologies at BSUIR*

**Abstract.** Parkinson's disease is a common neurological disease. A lot of data is used for its diagnosis and monitoring: voice, gait, and neuroimages. Regular CPU calculations are not always suitable, as they are slow and consume a lot of energy. Therefore, accelerators such as FPGA, Vitis HLS, and GRU are used. The report examines the architecture of motion data processing on an FPGA. Vitis HLS was used for the hardware implementation of motion processing in Parkinson's disease, and a wavelet transform was also used to isolate features. The GRU network is used for data analysis. The proposed method provides higher speed, lower delays and lower power consumption compared to conventional CPUs..

**Keywords**: Parkinson's, FPGA, Vitis HLS, GRU, wavelet transformation.

УДК 004.056, 004.3

# ПОДХОД К АППАРАТНОЙ ПОДДЕРЖКЕ ДАННЫХ ТЕКСТА С ИСПОЛЬЗОВАНИЕМ ПЛИС

***В.А. Вишняков***
*Доктор технических наук, профессор кафедры инфокоммуникационных технологий БГУИР*
vish@bsuir.by

***Ч. Юй***
*Аспирантка кафедры инфокоммуникационных технологий БГУИР*
1259017720@qq.com

***В.А. Вишняков***
*Область профессиональных интересов/исследований: управление и безопасность в инфокоммуникациях, Интернет вещей, блокчейн, электронный маркетинг, ИТ в образовании и медицине, интеллектуальные системы управления.*

***Ч. Юй***
*Аспирантка кафедры инфокоммуникационных технологий Белорусского государственного университета информатики и радиоэлектроники.*

**Аннотация.** Цель данного исследования – разработка аппаратно-ориентированного метода ускорения классификации текста на основе метода случайного леса в условиях ограниченных вычислительных ресурсов. Оптимизирована модель, реализована облегченная архитектура, с использованием арифметики с фиксированной запятой и вычислительного модуля на ПЛИС. Создана сквозная конвейерная обработка от исходного текстового ввода до извлечения признаков методом TF-IDF и классификации с использованием случайного леса. Предложенный подход продемонстрировал числовую согласованность с эталонной реализацией на языке Python. Хотя обработка TF-IDF выполняется на центральном процессоре, её взаимодействие с модулем ПЛИС осуществляется через шину AXI-Lite, что закладывает основу для полной аппаратной интеграции в дальнейших исследованиях.

**Ключевые слова:** ПЛИС, случайный лес, анализ текста, TF-IDF, аппаратное ускорение.

**Введение.** Технологии анализа текста играют важную роль в различных областях, таких как медицина, мониторинг общественного мнения и обслуживание клиентов [1-2]. В этих сценариях применения зачастую требуется эффективная обработка больших объемов текстовых данных, извлечение ключевой информации и выполнение классификации с высокой точностью. Традиционные задачи обработки естественного языка (NLP) обычно выполняются с использованием вычислений на CPU или GPU, что часто оказывается вычислительно затратным и энергоемким процессом. В сценариях граничных вычислений, где требуется низкая задержка и высокая энергоэффективность, такие среды зачастую обладают ограниченными вычислительными ресурсами [3], что затрудняет выполнение задач в режиме реального времени или при ограниченном энергопотреблении.

Для решения этих проблем могут быть применены несколько стратегий: сжатие моделей, использование легковесных алгоритмов, совместное использование ресурсов на границе и в облаке, а также применение специализированных аппаратных ускорителей, таких как GPU и ПЛИС, для повышения эффективности обработки. Методы интеллектуального анализа текста на основе ПЛИС (программируемые логические

интегральные схемы) стали актуальной областью исследований, поскольку предоставляют высокоэффективные аппаратные ускорители. Однако при применении аппаратного ускорения на базе ПЛИС в задачах NLP необходимо учитывать ключевые проблемы, такие как эффективная обработка крупномасштабных словарей, предобработка текста и интеграция традиционных программных компонентов с аппаратными модулями.

В области машинного обучения сверточные нейронные сети (CNN) [4] и модели глубокого обучения, такие как BERT [5], уже были развернуты на ПЛИС и показали улучшение пропускной способности и энергоэффективности по сравнению с чисто CPU-ориентированными решениями. Хотя матричные операции SVM и BERT естественным образом подходят для параллельных вычислений на ПЛИС, структура случайного леса (Random Forest) характеризуется нерегулярными ветвлениями и разрывами в памяти, что требует индивидуальной оптимизации.

Кроме того, в типичном процессе обработки NLP извлечение признаков TF-IDF и связанные этапы (например, фильтрация стоп-слов и токенизация) по-прежнему выполняются на CPU. Это связано с тем, что задачи динамического выделения памяти и работы со строками затрудняют их непосредственное аппаратное ускорение с помощью инструментов высокоуровневого синтеза (HLS). Разделение между CPU-ориентированным извлечением признаков NLP и классификацией на ПЛИС не только влечет за собой накладные расходы на интерфейс (например, передачу векторных признаков), но и создает основу для гибридных решений на базе CPU-ПЛИС.

Данное исследование направлено на разработку архитектуры аппаратного ускорения классификации текста на основе случайного леса (Random Forest) с применением ПЛИС, что позволит обеспечить высокую производительность и низкую задержку интеллектуального анализа текста в условиях ограниченных вычислительных ресурсов.

**Методологическая основа.** Для обеспечения эффективного развертывания на ПЛИС рабочий процесс разделен на несколько этапов. Сначала алгоритм был оптимизирован и адаптирован для аппаратной реализации, чтобы гарантировать, что размер модели и структуры данных соответствуют ресурсам и характеристикам ПЛИС. Затем основные алгоритмические модули были реализованы на языке C++ с фиксированной запятой для достижения более высокой производительности и эффективности использования аппаратных ресурсов. Затем выполнена аппаратная симуляции с использованием инструментов HLS для верификации функциональности и предварительных показателей производительности. Алгоритмическая реализация извлечения текстовых признаков и использование случайного леса соответствует подходу, описанному в авторской разработке [6].

Архитектура системы гибридного конвейера CPU-ПЛИС представлена на рисунке 1. Модуль на стороне CPU включает в себя подмодули для предварительной обработки текста, извлечения признаков TF-IDF и снижения размерности методом SVD, а также преобразования и сохранения данных. Подмодуль предварительной обработки текста выполняет токенизацию входного тестового текста, очистку символов, приведение к нижнему регистру и фильтрацию стоп-слов. Модуль извлечения признаков TF-IDF и снижения размерности SVD преобразует обработанный текст в 108-мерный вектор признаков с фиксированной точкой. Модуль преобразования и сохранения данных одновременно сохраняет вектор признаков в двоичном и текстовом форматах для сравнения с реализацией на Python, а также передает извлеченный вектор признаков в модуль ПЛИС.

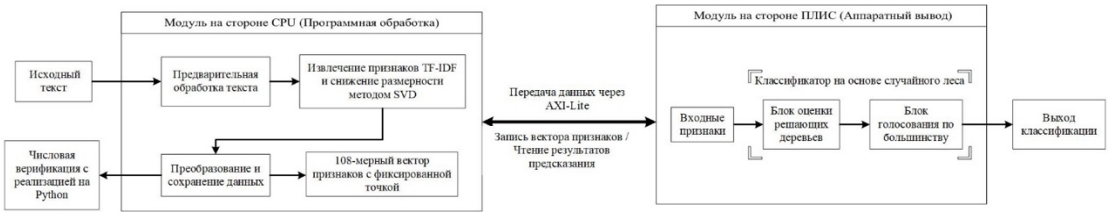



*Рисунок 1.* Схема архитектуры системы

Интерфейс AXI-Lite обеспечивает передачу 108-мерного вектора признаков, извлеченного на стороне CPU, в модуль классификатора случайного леса на ПЛИС, а вычисленное на аппаратной стороне значение «result» отображается в доступный для чтения и записи регистр AXI-Lite, позволяя пользователю считывать его со стороны CPU. Модуль на стороне ПЛИС включает классификатор на основе случайного леса, состоящий из блока оценки, содержащего решающие деревья, в которых каждый узел хранится с использованием структуры TreeNode. Выходные данные нескольких деревьев обрабатываются с помощью механизма голосования по большинству для получения финального предсказания. После завершения обработки на аппаратной стороне данные передаются из этого модуля на следующий этап системы или могут быть считаны извне.

Для проверки проведены функциональные и производительные тесты сквозного (Python-C++, ПЛИС) процесса классификации текста для обеспечения согласованности результатов на разных платформах. Предварительная обработка текста, извлечение признаков TF-IDF и снижение размерности SVD были выполнены с использованием Python для генерации векторов признаков, совместимых с модулем C++. Соответствующие параметры были экспортированы, а выходные данные реализации на C++ сравнивались с эталонной реализацией на Python. Результаты тестирования показали, что косинусное сходство между векторами признаков, сгенерированными из одного и того же тестового текста на разных платформах, составило 0,9990, что свидетельствует о высокой степени направленного совпадения выходных векторов C++ и Python, подтверждая корректность вычислений TF-IDF и SVD.

**Экспериментальная верификация.** В модуле ПЛИС классификатор случайного леса был реализован с использованием инструментов HLS, при этом аппаратная реализация деревьев решений через распределённое хранение узлов в LUTRAM с конвейерной обработкой на фиксированной точке. Согласно отчету о синтезе, общая задержка функции инференса случайного леса составила примерно 11 193 такта, что соответствует времени выполнения предсказания 0,112 миллисекунды при тактовой частоте 100 МГц. В отношении использования ресурсов ПЛИС-реализация составила 5 % – 30 % от общего объема аппаратных ресурсов, при этом использование LUT, FF и DSP соответствовало заданным проектным целям и не превышало ограничений ресурсоемкости целевого устройства Xilinx XCVU11P. Дальнейшая оптимизация может дополнительно повысить вычислительный параллелизм и увеличить пропускную способность.

**Заключение.** Авторы, в рамках гибридной архитектуры CPU + ПЛИС, выполнили извлечение признаков TF-IDF + SVD на стороне C++ (или предварительную обработку на стороне Python), затем передали полученный 108-мерный вектор признаков с фиксированной точкой в IP-ядро случайного леса на ПЛИС и подтвердили числовую согласованность с эталонной реализацией на Python, достигнув косинусного сходства 0,9990, а также согласованности предсказанных результатов. Задержка аппаратного инференса и использование ресурсов указывают на то, что данный метод обладает потенциалом масштабируемости на крупномасштабных ПЛИС-устройствах.

## Список литературы

[1] Janowski A. Natural Language Processing Techniques for Clinical Text Analysis in Healthcare. Journal of Advanced Analytics in Healthcare Management. 2023; 7(1):51–76.

[2] Hu Y. Text Mining and Data Information Analysis for Network Public Opinion. Data Science Journal. 2019; 18(1):7.

[3] Wisniewski M., Bec J.-M., Boguszewski G., Gamatié A. Hardware Solutions for Low-Power Smart Edge Computing. Journal of Low Power Electronics and Applications. 2022; 12(4):61.

[4] Mouri Zadeh Khaki A., Choi A. Optimizing Deep Learning Acceleration on FPGA for Real-Time and Resource-Efficient Image Classification. *Applied Sciences.* 2025; 15(1):422.

[5] Khan H., Khan A., Khan Z., Huang L.B., Wang K., He L. NPE: An FPGA-based Overlay Processor for Natural Language Processing. *The 2021 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays (FPGA '21).* Association for Computing Machinery, New York, NY, USA, 2021; 227.

[6] Вишняков В.А., Юй Ч. Использование машинного обучения для распознавания болезни Альцгеймера на основе транскрипционной информации. *Доклады БГУИР.* 2023; 21(6):106–112.

**Авторский вклад**

**Вишняков Владимир Анатольевич** – предложение концепции метода.

**Юй Чуюэ** – реализация предложенного метода.

# APPROACH TO HARDWARE-ASSISTED TEXT ANALYSIS USING FPGA

***V.A. Vishnyakov***
*Doctor of Technical Sciences, Professor of the Department of Infocommunication Technologies, BSUIR*

***C. Yu***
*PhD student in the Department of infocommunication technologies at BSUIR*

**Abstract.** The purpose of this study is to develop a hardware-oriented method for accelerating text classification based on a random forest in conditions of limited computing resources. The model has been optimized, a lightweight architecture has been implemented, using fixed-point arithmetic and FPGA computing modules. End-to-end pipelining has been created from the source text input to feature extraction using the TF-IDF method and classification using a random forest. The proposed approach demonstrated numerical consistency with the reference implementation in Python. Although TF-IDF processing is performed on the central processor, its interaction with the FPGA module is carried out via the AXI-Lite bus, which lays the foundation for full hardware integration in further research.

**Keywords**: FPGA, Random Forest, Text Analytics, TF-IDF, Hardware Acceleration.

УДК 347

# БОЛЬШИЕ ДАННЫЕ (BIG DATA) КАК СОВОКУПНОСТЬ ОБЪЕКТОВ ГРАЖДАНСКИХ ПРАВ

***В.К. Раюшкин***
*Аспирант*
*Национального центра законодательства и правовой информации Республики Беларусь*
rayushkin.vladislav@mail.ru

***В.К. Раюшкин***

*Обучается в аспирантуре Национального центра законодательства и правовой информации Республики Беларусь. Магистр права. Область научных интересов связана с исследованием правовой сущности больших данных (Big Data).*

**Аннотация.** В статье проведена гражданско-правовая характеристика применимых к большим данным режимов объектов гражданских прав. Проанализировано законодательство об инновационной деятельности, а положения некоторых актов в области технического нормирования и стандартизации. Отражены мнения ученых по исследуемой проблематике.

**Ключевые слова:** большие данные, массивы информации, технологии, результаты интеллектуальной деятельности, интеллектуальная собственность, исключительное право.

**Введение.** Цифровизация всех сфер общественной жизни и государственного управления как закономерный процесс развития информатизации основывается на обширном использовании больших данных (Big Data) (далее – большие данные), искусственного интеллекта, нейросетей, облачных вычислений и других передовых феноменов цифрового общества. Помимо определения основных целей, направлений и способов развития таких явлений цифровизации, огромное значение и актуальность приобретает вопрос определения правового режима вышеуказанных объектов. Поскольку именно выбранный законодательным органом каждой страны путь к установлению применимых к цифровым объектам правовых норм может как способствовать их введению в гражданский оборот, так и осложнить такой процесс.

**Основная часть.** Большие данные обычно называют информационной технологией, ставя ее в один ряд с технологией искусственного интеллекта, нейросетевыми технологиями, технологиями облачных вычислений, интернета вещей и другими передовыми технологиями, основанными на обработке данных. Однако, по нашему мнению, большие данные представляют собой сложносоставной феномен информационного общества, а с юридической точки зрения комплексную юридическую модель, существенными признаками которой являются «существование массивов данных, а также использование в отношении массивов данных технологии по их обработке, синкретично составляющей такую категорию объективной реальности, как большие данные» [1, с. 105]. Узкое представление больших данных только в качестве больших массивов информации или только как технологий по обработке этих массивов ведет к недостаточному раскрытию гражданско-правовой сущности больших данных и

невозможности надлежащей правовой квалификации больших данных для целей их внедрения в гражданский оборот Республики Беларусь. [2, с. 352]. Исходя из выработанного нами подхода к определению юридической сущности больших данных как синкретичной связи массивов данных и технологий по их обработке, для формирования единого гражданско-правового режима больших данных в данной работе следует проанализировать гражданско-правовой режим элементов, составляющих большие данные.

Схожую концепцию к определению сущности больших данных можно усмотреть и в таком техническом нормативном правовом акте, как СТБ 2583-2020 «Цифровая трансформация. Термины и определения», утвержденном Постановлением Государственного комитета по стандартизации Республики Беларусь от 8 декабря 2020 г. № 95. Согласно содержанию данного акта, большие данные представляют собой «обширный наборы данных, характеризующиеся значительными объемами, разнообразием, скоростью обработки, требующие масштабируемой технологии для эффективного хранения, управления и анализа» [3]. Это определение разработано в национальном техническом стандарте на основе формулировки, выработанной Международной организацией по стандартизации (ISO) в международном стандарте ISO/IEC 20546:2019 Information technology – Big data – Overview and vocabulary (Информационные технологии. Большие данные. Обзор словарь) [4].

Также отметим, что исследованием сущности больших данных занимались ряд ученых, изучающих большие данные с различных сторон, при этом пользуясь разнообразным теоретическим инструментарием, ввиду этого по-прежнему существует дискуссия о сущности больших данных как феномена цифровизации. В ходе аналитики исследований ученых представляется возможным разделить их точки зрения по критерию представления больших данных исключительно информацией (данными) больших объемов (А. И. Савельев, К. А. Соснин, Л. В. Санникова, Ю С. Харитонова, М. А. Кольздорф, А. М. Лаптева, Д. С. Карлаш и др.), в первую очередь технологией по обработке информации (данных) (Е. А. Кириллова, С. С. Кривушева, Ю. Н. Толстова, К. Г. Антипова и др.) или синкретичной связью информации (данных) огромных массивов и технологии по их обработке (М. А. Рожкова, В. Н. Глонина, Л. Ю. Василевская, А. П. Рабец, А. А. Гусева и др.).

По нашему мнению, наиболее применимым для подобного рода объектов правом является право интеллектуальной собственности, поскольку именно оно весьма тесно связано с инновациями, современными технологиями и различного рода программными компонентами, которые, бесспорно, являются плодами интеллектуальной деятельности, и составляющими такие технологии.

Как справедливо отмечает Л. Ю. Василевская «big data – условное понятие. Речь, на наш взгляд, должна идти о результате интеллектуальной деятельности человека, о своеобразном инновационном продукте…» [5, с. 14]. В подтверждение того можно привести норму белорусского законодательства об инновационной деятельности: «введение в гражданский оборот продукции, технологии, услуги, организационно-технического решения – совершение любых сделок, предметом которых являются оказание услуг, предоставление права пользования продукцией, технологией, переход к другому лицу имущественных прав на продукцию, технологию, организационно-техническое решение, в том числе имущественных прав на результаты интеллектуальной деятельности, используемые при создании этих продукции, технологии, организационно-технического решения» (абз. 2 ст. 1 Закона Республики Беларусь от 10 июля 2012 г. №425-З «О государственной инновационной политике и инновационной деятельности») [6]. Как видим из изложенного, в качестве одной из форм введения в гражданский оборот продукции, технологии, услуги, организационно-технического решения является совершение любых сделок, предметом которых будет переход к другому лицу имущественных прав на результаты интеллектуальной деятельности, используемые при создании этих продукции,

технологии, организационно-технического решения. На основании этого можно сделать вывод о том, что законодательный орган рассматривает создание технологии как деятельность по использованию результатов интеллектуальной деятельности.

Помимо этого, как пишет И. З. Аюшеева «…осуществление и защита прав на большие данные возможны в рамках отношений, складывающихся по поводу результатов интеллектуальной деятельности» [7, с. 132]. Схожей позиции придерживаются Е. А. Вониканис, указывающая, что «по аналогии с правом собственности в отношении вещей, в сфере нематериальных объектов, к которым относят данные, самую сильную защиту предоставляет право интеллектуальной собственности. Соответственно именно интеллектуальная собственность рассматривается бизнесом как наиболее привлекательный правовой инструмент для обеспечения контроля над данными» [8, с. 142]. Данное мнение автора корреспондирует с позицией А. А. Карцхия, который в своем диссертационном исследовании пришел к выводу, что институт права интеллектуальной собственности со своим правовым инструментарием может быть использованы как концептуальная основа для регулирования цифровых правоотношений в общем [9, с. 248]. Е. А. Свадковская в своем исследовании указывает на большой потенциал, который присущ исключительному праву отмечая, что «возможности исключительного права далеко не исчерпаны, данный правовой механизм при определенных условиях может быть востребован для предоставления правовой охраны отдельным цифровым результатам интеллектуальной деятельности, в том числе и тем, которые на сегодняшний день не являются объектами права интеллектуальной собственности» [10, с. 101].

На основании приведенных позиций исследователей, следует проанализировать гражданско-правовой статус как информации (данных), так и технологий в структуре больших данных.

Для определения правового режима совокупности информации (данных) в структуре больших данных рассмотрим базы данных и коммерческую тайну по причине того, что среди всех объектов гражданских прав данные объекты представляются нам наиболее подходящими для охраны совокупности данных. Так, весь массив информации, который необходим для аналитики больших данных, представляется возможным, при соблюдении всех установленных соответствующим законодательством требований, охранять посредством придания данным сведениям режима коммерческой тайны.

Таким же образом отметим, что развитие отечественного законодательства об интеллектуальной собственности также способно обеспечить массивы больших данных надлежащей правовой охраной через адаптацию правовой регламентации базы данных, или появление новых объектов права интеллектуальной собственности, специально для этих целей созданных. Однако базы данных в современной их регламентации, по нашему мнению, не подходят для этих целей в связи с тем, что существующее законодательство о базах данных не устанавливает правовой охраны для информации в своей структуре, что, с одной стороны, исключает возможность возникновения конфликта правового режима баз данных как объекта гражданских прав и правовых режимов иной охраняемой законом информации, а с другой – выводит информацию в структуре баз данных за предмет правовой охраны данного объекта интеллектуальной собственности. Кроме этого, при формировании баз данных как структурированной и обработанной информации для последующей ее аналитики технологиями больших данных встает проблема того, что, как правило, такая систематизация информации осуществляется программными алгоритмами, в том числе искусственным интеллектом, что исключает творческий характер этой структуры ввиду отсутствия основного элемента – человека как субъекта, реализующего творческий замысел.

Также отметим, что в отношении совокупности разного рода ограниченных в распространении сведений, которые могут попасть в массивы данных, обрабатываемые технологиями больших данных, данная потенциальная возможность не

должна рассматриваться как преграда для развития оборота массивов данных, как части больших данных.

Далее проанализируем технологии больших данных, а также обеспечивающие их функционирование программные средства в контексте определения их гражданско-правового режима.

Технологии больших данных как процессы и методы обработки больших объемов информации в структуре больших данных при соответствии их требованиям к новизне, изобретательскому уровню, промышленной применимости и иным критериям патентоспособности, могут получить правовую охрану как изобретения. Исходя из этого, процессы и способы обработки больших объемов данных могут быть запатентованы в качестве изобретений в случае, когда они не известны из уровня техники, неочевидны для специалистов в данной области и являются промышленно применимыми при условии решения ими технических задач, что приводит к получению полезного технического результата. В ином случае процессы и методы обработки больших массивов информации могут получить охрану как секреты производства (ноу-хау).

Программно-технические средства, обеспечивающие функционирование технологий больших данных, также являются разного рода результатами интеллектуальной деятельности, охраняемыми в Республике Беларусь в качестве объектов авторского права или права промышленной собственности: программы и операционные системы как компьютерная программа; алгоритм функционирования программ может быть зарегистрирован в качестве изобретения или получить правовую охрану в качестве секрета производства (ноу-хау); интерфейс программ может быть объектом патентования в качестве промышленного образца; технические средства могут получить охрану в качестве изобретений, а также может быть получен патент на программно-аппаратный комплекс, являющийся изобретением.

**Заключение.** Таким образом большие данные следует рассматривать как сложносоставной объект гражданских прав, состоящий из массивов информации (данных), подлежащих обработке, и технологий по их обработке вместе с программно-техническими средствами, обеспечивающими такие процессы. Массивы информации при соблюдении определенных условий и развитии отечественного законодательства могут получить охрану в качестве коммерческой тайны и баз данных. Технологии больших данных вместе с программно-техническими средствами при их соответствии установленным соответствующим законодательством требованиям могут охраняться в качестве компьютерной программы как объекта авторского права, а также изобретений, промышленных образцов и секретов производства (ноу-хау) как объектов права промышленной собственности.

Исходя из описанного, в целом, большие данные могут рассматриваться как охраноспособный объект, состоящий из отдельных результатов интеллектуальной деятельности и коммерческой тайны. Однако, на наш взгляд, возникла необходимость в развитии отечественного законодательства в области интеллектуальной собственности в аспекте унификации правового режима данных объектов. Видится возможным как гражданско-правовая регламентация сложносоставных объектов, состоящих из нескольких результатов интеллектуальной деятельности, так и развитие института исключительных прав в аспекте формирования правового регулирования больших данных, искусственного интеллекта, нейронных сетей, облачных вычислений и других передовых феноменов цифрового развития общества.

## Список литературы

[1] Раюшкин, В. К. Понятие больших данных: основные теоретические подходы / В. К. Раюшкин // Право.by. – 2024. – № 5 (91). – С. 100–107.

[2] Раюшкин, В. К. Соотношение информации и технологий в структуре больших данных / В. К. Раюшкин // Правовая политика, наука, практика – 2024 : сб. материалов Респ. науч.-конф., Минск, 10 окт. 2024

г. / Национальный центр законодательства и правовой информации Республики Беларусь ; редкол.: Е. В. Семашко [и др.]. – Минск : НЦЗПИ, 2024. – С. 349–352.

[3] Цифровая трансформация. Термины и определения: СТБ 2583–2020. – Минск: Гос. комитет по стандартизации, 2020. – 16 с.

[4] ISO/IEC 2382:2015 Information technology – Vocabulary. – URL: https://www.iso.org/standard/63598.html (date access: 11.02.2024).

[5] Василевская, Л. Ю. Цифровизация гражданского оборота: big data в механизме гражданско-правового регулирования (цивилистическое исследование) : монография : в 5 т. / Л. Ю. Василевская, Е. Б. Подузова, Ф. А. Тасалов. – М. : Проспект, 2023. – Т. 5 – 360 с.

[6] О государственной инновационной политике и инновационной деятельности : Закон Респ. Беларусь от 10 июля 2012 г. № 425-З : в ред от 6 янв. 2023 г. № 152-З // КонсультантПлюс. Беларусь : справ. правовая система (дата обращения: 11.02.2024).

[7] Аюшеева, И. З. Большие данные: проблемы определения гражданско-правового режима / И. З. Аюшеева // Lex Russica (русский Закон). – 2023. – Т. 76, № 10 (203). – С. 125–134.

[8] Войниканис, Е. А. Регулирование больших данных и право интеллектуальной собственности: общие подходы, проблемы и перспективы развития / Е. А. Войниканис // Закон. – 2020. – № 7. – С. 135–156.

[9] Карцхия, А. А. Гражданско-правовая модель регулирования цифровых технологий : дис. … доктора юрид. наук : 12.00.03 / Карцхия Александр Амиранович; ФГБОУ ВО «Российская государственная академия интеллектуальной собственности». – Москва, 2019. – 394 л.

[10] Свадковская, Е. А. Юридическая наука и правосудие в сфере охраны новых невещественных объектов права / Е. А. Свадковская // Белорусское правосудие: исторический опыт и современность : материалы междунар. науч.-практ. конф., посвящ. 100-летию Верхов. Суда Респ. Беларусь, Минск, 15 февр. 2023 г. / Белорус. гос. ун-т ; редкол.: Л. Л. Голубева (гл. ред.), С. А. Калинин, О. Н. Романова. – Минск : БГУ, 2023. – С. 95 – 101.

## Авторский вклад

**Раюшкин Владислав Кириллович** – проведение исследования и написание статьи.

УДК 004.67

# ВОРОНКА ПРОДАЖ КАК ИНСТРУМЕНТ ВЕБ-АНАЛИТИКИ

***И.В. Марахина***
*Доцент кафедры экономики, БГУИР,*
*кандидат экономических наук, доцент*
marahina@bsuir.by

***И.В. Марахина***
*Окончила Белорусский государственный университет информатики и радиоэлектроники. Более 15 лет работает в отраслях науки и образования. Сферы интересов: инновационное развитие, теория организации, маркетинг, поисковое продвижение и веб-аналитика.*

**Аннотация.** В статье описаны теоретико-методологические основы веб-анализа с использованием воронки продаж. Проанализированы этапы, виды и ключевые характеристики такого инструмента, сгруппированы показатели воронки продаж. Проведен анализ возможностей ее построения используя такие системы аналитики, как Google Analytics и Яндекс Метрика.

**Ключевые слова:** воронка продаж, конверсия, веб-аналитика, Google Analytics, Яндекс Метрика

**Введение.** Современные процессы цифровизации экономики определяют все более активное использование веб-ресурсов в коммерции. Анализ показывает, что в Беларуси роль сайтов продолжает возрастать. Так согласно статистическим данным в 2022 году доля организаций, имевшие веб-сайт составила 71,6 % [1, с. 35]. При этом по сравнению с 2016 годом значение этого показателя увеличилось более чем на 9 % [1, с. 35]. Следует отметить и качественные изменения сайтов, в том числе значительное усложнение функционала и реализацию все большего потенциала коммерческих возможностей. Растет число организаций, использующих сайт для публикации каталогов товаров, а также расчетов посредством электронной платежной системы [1, с. 39]. Рост значимости сайта в электронной коммерции накладывает большие требования к аналитике сайтов или веб-аналитике.

Большинство исследователей под веб-аналитикой понимают систему измерения, сбора, анализа, представления и интерпретации информации о посетителях веб-сайтов с целью их улучшения и оптимизации [2, 3]. Система веб-аналитики имеет значительный набор метрик и инструментов, среди которых большое значение имеет воронка продаж. Такой инструмент позволяет оценить экономический эффект работы интернет-ресурса по привлечению клиентов. Как отмечается в [4], исследование воронки позволяет визуализировать шаги пользователей на пути к выполнению определенного действия и быстро узнать, насколько успешно клиенты проходят каждый из этапов. Например, как посетители становятся потенциальными клиентами, а затем – покупателями или как однократные покупатели становятся постоянными. Такая информации позволяет оптимизировать неэффективные или прерванные пути клиентов к покупке.

**Основная часть.** Классическая воронка продаж предполагает анализ таких этапов взаимодействия компании с клиентом, как привлечение, интерес, рассмотрение, намерение, действие, лояльность и рекомендации. Следует отметить, что в веб-аналитике эти этапы трансформируются с учетом специфики веб-сайтов. Например, согласно [3] воронка может

содержать следующие этапы: показы информации, переходы на сайт, выполнение целевого действия и подтверждения по телефону. С учетом перечисленных выше этапов классической воронки продаж для веб-сайтов можно добавить также этапы: онлайн-оплата, регистрация на сайте (заведение личного кабинета), совершение повторных покупок, написание отзывов и рекомендаций и т.д.

Для целей веб-аналитики можно использовать открытые и закрытые воронки [4]:

– в открытую воронку пользователи могут входить на любом шаге;
– в закрытую воронку пользователи могут входить только на начальном шаге.

Еще одной важной характеристикой воронки продаж, которую нужно учитывать, является время, за которое пользователь проходит этапы воронки продаж [5].

Для оценки эффективности воронки продаж можно использовать абсолютные показатели и относительные показатели. Абсолютные – отражают количество потенциальных или реальных клиентов на каждом этапе воронки продаж. Относительными показателями в первую очередь выступают показатели конверсии. Например, в Яндекс Метрике предлагают рассчитывать следующие два вида конверсии [5].

1 Конверсия от первого шага — все шаги воронки будут учитываться как отношение просматриваемого шага к первому шагу.

2 Конверсия от предыдущего шага — все шаги воронки будут учитываться как отношение просматриваемого шага к предыдущему.

Анализ и сравнение возможностей воронки продаж двух популярных систем аналитики, как Google Analytics, Яндекс Метрика представлены в таблице, в которой выделен ряд сущностных особенностей. Такие особенности необходимо учитывать при выборе системы аналитики.

Таблица. Характеристики воронки продаж в системах аналитики

| Характеристики | Google Analytics | Яндекс Метрика |
|---|---|---|
| Шаги воронки | На каждом шаге можно выбрать события и параметр. У воронки может быть до 10 шагов. | При создании этапов предлагается выбрать условие достижения шага: достижение JS-цели просмотр страницы параметры события |
| Период воронки | При выборе шага можно установить временное ограничение для каждого шага | Можно указать время, за которое пользователь должен пройти все шаги (ограниченное окно) |
| Группировка, сегментирования | Можно использовать до четырех сегментов, чтобы выбрать определенные группы пользователей | С помощью поля группировки можно выбрать, в каких срезах посмотреть воронку. Можно выбрать до 7 группировок |
| Конверсия | Автоматически показывается конверсия от предыдущего шага | В поле управления конверсиями можно выбрать один из вариантов: конверсия от первого шага и конверсия от предыдущего шага |
| Визуализация | График и таблица. Есть возможность выбора стандартной воронки | График и таблица |

| | | |
|---|---|---|
| | (дискретная диаграмма) или воронки с учетом тенденций (линейная диаграмма). Воронка с учетом тенденций позволяет посмотреть все шаги одновременно и подробные сведения о каждом из них | |
| Настройки | Переключатель «Создать открытую воронку»<br>Переключатель «Показать время между шагами» (сумма времени в микросекундах между предыдущим и текущим шагом, разделенная на количество пользователей, которые достигли текущего шага.<br>Просмотр наиболее частых действий пользователя на каждом шаге воронки.<br>Широкий спектр фильтров | |

Источник: составлено с использованием [4, 5]

**Заключение.** Проведенное исследование показало возрастающую роль веб-аналитики и значимость в ней такого инструмента, как воронка продаж. В работе выделены ключевые характеристики и показатели воронки продаж, а также рассмотрены возможности ее автоматической реализации. Проведенное сравнение реализации воронки продаж в Google Analytics и Яндекс Метрика позволило выделить ряд сущностных особенностей, которые необходимо учитывать при выборе системы аналитики.

## Список литературы

[1] Информационное общество в Республике Беларусь. Национальный статистический комитет Республики Беларусь. Мн., 2023.

[2] Википедия [Электронный ресурс]: https://ru.wikipedia.org/.

[3] Осипенков Я. Google Analytics для googлят: практическое руководство по веб-аналитике. М: Graph Analytics; 2018.

[4] Google [Электронный ресурс]: https://support.google.com/analytics/answer/9327974?hl=ru

[5] Yandex. https://yandex.ru/support/metrica/reports/funnels.html

# PURCHASE FUNNEL AS A WEB ANALYTICS TOOL

***I.V. Marahina***

*Associate professor of the Department of Economics of BSUIR,*
*Candidate of Economic Sciences, Associate Professor*

**Annotation.** The article describes the theoretical and methodological foundations of web analysis using a purchase funnel. The stages, types and key characteristics of such a tool are analyzed, and the indicators of the purchase funnel are grouped. An analysis of the possibilities of its construction was carried out using such analytics systems as Google Analytics and Yandex Metrica.

**Keywords:** purchase funnel, conversion, web analytics, Google Analytics, Yandex Metrica.

УДК 004.65

# НЕЙРОТЕХНОЛОГИИ И BIG DATA В ОБРАЗОВАНИИ

***С.С. Шайымов***
*Преподаватель кафедры "Автоматизации" института Телекоммуникаций и информатики Туркменистана*
smyradow615@gmail.com

***А.Б.Чарыев***
*Преподаватель кафедры "Автоматизации" института Телекоммуникаций и информатики Туркменистана*
arslan.caryyew93@gmail.com

***Б.Н. Атанепесов***
*Преподаватель института Телекоммуникаций и информатики Туркменистана*
bayramhhd@mail.ru

***С.С. Шайымов***
*Окончил Государственный энергетический институт Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.*

***А.Б.Чарыев***
*Окончил Государственный энергетический институт Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.*

***Б.Н.Атанепесов***
*Окончил институт Телекоммуникаций и информатики Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.*

**Аннотация.** Современное образование сталкивается с вызовами, требующими инновационных подходов к учебному процессу. В эпоху цифровых технологий нейротехнологии и Big Data играют все более важную роль, переопределяя способы обучения и обогащая педагогическую практику. В этой статье мы рассмотрим, как нейротехнологии и анализ Big Data влияют на образовательную сферу, определяя новые горизонты персонализированного обучения и инновационного подхода к развитию учащихся.

**Ключевые слова:** нейротехнология, нейрофидбек, Big Data, образования.

Современные роботизированные системы функционируют в средах с высокой динамикой и неопределённостью. В таких условиях использование BIG DATA обеспечивает своевременную корректировку управляющих параметров на основе анализа крупных потоков сенсорной информации. Это согласуется с выводами, указывающими на важность статистической обработки больших наборов данных для минимизации ошибок позиционирования.[1]

Существует 4 основных метода анализа BIG DATA:

1 Наиболее распространенным методом является описательная аналитика. Он спросил: «Что случилось?» и анализирует текущую и историческую информацию. Его главная цель — выявить причины и закономерности успеха или неудачи в определенной области и использовать их в более эффективных моделях.

2 Предиктивная аналитика - помогает прогнозировать возможные события на основе имеющихся данных. Для этого используются модели, основанные на определенных объектах или сходных явлениях. Таким образом, можно спрогнозировать возможное падение или изменение цен на рынке казначейских обязательств или, по крайней мере, определить платежеспособность заемщика по проекту.

3 Предписывающая аналитика — с помощью BIG DATA и современных технологий можно выявить проблемные области в бизнесе или любой другой деятельности и определить, что необходимо сделать, чтобы предотвратить их повторение в будущем.

4 Диагностическая аналитика — использует данные для анализа причин инцидента. Это помогает обнаружить необычные обстоятельства и случайные связи между событиями и действиями [2].

**Нейротехнологии**, такие как анализ мозговой активности и нейрофидбек, предоставляют уникальную возможность исследовать когнитивные процессы студентов в реальном времени, расширяя понимание о том, как мы учимся и какие методы обучения наиболее эффективны. С другой стороны, Big Data и аналитика обучения позволяют собирать и анализировать большие объемы данных о прогрессе студентов, их предпочтениях и уровне усвоения материала, что открывает новые возможности для персонализации образовательного опыта и повышения его эффективности.

Цель данной статьи — проанализировать влияние нейротехнологий и Big Data на образование, выявить их преимущества и вызовы, а также представить практические примеры успешной реализации этих технологий в учебных заведениях. В результате мы сможем понять, как интеграция этих инновационных подходов в учебный процесс способствует созданию более эффективной, персонализированной и индивидуализированной образовательной среды.

**Нейротехнологии в образовании**

***Анализ мозговой активности.*** В сфере образования нейротехнологии, например, анализ мозговой активности, играют ключевую роль в понимании когнитивных процессов, происходящих во время обучения. Методы, такие как функциональная магнитно-резонансная томография (fMRI) и электроэнцефалография (EEG), предоставляют возможность регистрировать активность различных участков мозга в реальном времени. Это открывает перед нами новые горизонты для адаптации учебных программ в соответствии с индивидуальными потребностями студентов.

Анализ мозговой активности позволяет исследовать, какие типы задач вызывают наибольшую активацию учебных центров в мозгу студента. Эта информация может быть использована для оптимизации материалов и методов обучения, сделав учебный процесс более эффективным и привлекательным.

***Нейрофидбек.*** Технология нейрофидбека предоставляет студентам инструмент саморегуляции и самопонимания своих когнитивных процессов. С использованием датчиков, измеряющих мозговую активность, студенты могут получать обратную связь о своем уровне внимания, релаксации или стресса в реальном времени. Это создает возможность для студентов лучше понимать свои сильные и слабые стороны в процессе обучения, что способствует их лучшему адаптивному обучению.

Такие технологии также могут быть использованы преподавателями для оценки эффективности своих методов преподавания. Путем анализа реакции мозга студентов на конкретные задания или лекции, преподаватели могут оптимизировать свои подходы, делая обучение более эффективным.

**Big Data в образовании**

***Аналитика обучения.*** В сфере образования Big Data и аналитика обучения играют решающую роль в сборе и анализе больших объемов данных о прогрессе студентов. Аналитика обучения позволяет отслеживать успехи и вызовы студентов, выявлять

тенденции в их обучении, и предоставлять ценную информацию для адаптации учебных программ.

Применение аналитики обучения позволяет создавать персонализированные образовательные планы, учитывая уровень подготовки и предпочтения каждого студента. Это способствует более эффективному усвоению материала и повышению мотивации студентов.

***Прогностическая аналитика.*** Прогностическая аналитика на базе Big Data позволяет предсказывать успех студентов и выявлять факторы, которые могут привести к дропауту. Алгоритмы машинного обучения анализируют исторические данные и выявляют паттерны, помогая индивидуализировать подход к каждому студенту.

Использование прогностической аналитики дает возможность своевременно выявлять студентов, находящихся в зоне риска, и предоставлять им дополнительную поддержку. Это способствует улучшению общего уровня успеваемости и снижению процента дропаутов.

Вместе эти технологии создают основу для образовательной среды, ориентированной на индивида, и способствуют достижению более высоких стандартов обучения и развития студентов.

**Преимущества использования нейротехнологий и Big Data**

Внедрение нейротехнологий и Big Data в образование приносит множество преимуществ.

1 *Индивидуализированное обучение:* Нейротехнологии позволяют создавать персонализированные программы обучения, учитывая индивидуальные особенности каждого студента. Аналитика обучения и прогностическая аналитика на основе Big Data дополняют этот подход, делая учебный процесс более адаптивным.

2 *Улучшение эффективности образования:* Анализ мозговой активности и данных обучения позволяет выявлять оптимальные методы преподавания и материалы, что приводит к повышению эффективности образовательного процесса.

3 *Снижение дропаутов:* Прогностическая аналитика помогает выявлять студентов с риском дропаута, предоставляя возможность для оперативной поддержки и предотвращения преждевременного завершения обучения.

***Вызовы внедрения технологий в образование***

Существуют также вызовы, которые следует учитывать при внедрении нейротехнологий и Big Data в образование.

1 *Проблемы конфиденциальности:* Сбор и анализ больших объемов данных о студентах может вызвать вопросы по поводу конфиденциальности. Важно разработать строгие политики безопасности данных и обеспечить их соблюдение.

2 *Обучение персонала:* Преподавателям и администрации требуется подготовка для эффективного использования новых технологий. Обучение персонала является важным аспектом успешной интеграции нейротехнологий и Big Data в учебные программы.

**Практические примеры исследований и реализаций**

***Исследования в области нейротехнологий***

– *применение fMRI в изучении воздействия различных методов обучения на мозговую активность студентов;*

– *эксперименты с использованием нейрофидбека для повышения концентрации студентов в классе.*

***Проекты с применением Big Data в образовании***

– *использование аналитики обучения для оптимизации учебных планов в высших учебных заведениях;*

– *внедрение прогностической аналитики в школьные системы для предотвращения дропаутов.*

Эти практические примеры подчеркивают успешное внедрение нейротехнологий и Big Data в реальные образовательные сценарии, подтверждая их потенциал для улучшения образовательного процесса.

Объединение нейротехнологий и Big Data в образовании открывает перед нами новые возможности для трансформации учебного процесса. Преодоление вызовов, связанных с конфиденциальностью и обучением персонала, позволит максимально использовать потенциал этих инноваций. Их успешная интеграция способствует созданию образовательной среды, способной адаптироваться к потребностям каждого учащегося, поднимая качество образования на новый уровень.

## Список литературы

[1]Чарыев А.Б., Шайымов С.С., Аннабердиев И.А., “Важность использования BIG DATA для оптимизации управления роботами-манипуляторами”. Международный научный журнал “ИННОВАЦИОННАЯ НАУКА” №2-1/2025.

[2]Шайымов С.С., Гочов А.С., Веллиева С.Т., “Значение BIG DATA в цифровую эпоху”. Международный научный журнал “ИННОВАЦИОННАЯ НАУКА” №2-2/2025.

## Авторский вклад

Авторы внести равнозначный вклад.

# NEUROTECHNOLOGIES AND BIG DATA IN EDUCATION

***S.S. Shayymov***
*Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan*

***A.B. Charyev***
*Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan*

***B.N. Atanepesov***
*Lecturer at the Institute of Telecommunications and Informatics of Turkmenistan*

**Abstract.** Modern education faces challenges that require innovative approaches to the learning process. In the digital age, neurotechnology and Big Data play an increasingly important role, redefining learning methods and enriching pedagogical practice. In this article, we will consider how neurotechnology and Big Data analysis influence the educational sphere, defining new horizons for personalized learning and an innovative approach to student development.

**Keywords:** neurotechnology, neurofeedback, Big Data, education.

UDC 004.041:004.5:655.26:004.09:004.8.

# THE ROLE OF BIG DATA IN GRAPHIC DESIGN AND USER INTERFACES A DATA-DRIVEN APPROACH FROM IDEA TO IMPLEMENTATION

***B.A. Al-Nami***
*Associate Professor, Department of Informatics and Computer Design, SPBSUT, Ph.D. (in technical sciences),*
alnomibasheer@gmail.com

***B.A. Al-Nami***

*Graduated from Saint Petersburg Electrotechnical University "LETI".Research interests are connected with Designing of virtual means of information technologies, 3d polygonal modeling and Ergonomics, standardization and design (usability) of software products ,Web-projecting and design, informatics and computer design, Visual interface design of digital products, Information Technologies, Information Technologies of data visualization and research processes in The Bonch-Bruevich Saint Petersburg State University of Telecommunications.*

**Abstract.** This article explores the disruptive potential of big data in graphic design and user interface, charting its effects from inception to deployment. It describes how deep awareness of user behavior is made possible by contemporary monitoring technologies and machine learning, which feed back into highly customized, adaptable interfaces that go beyond conventional designs. Important strategies include scale A/B testing, real-time optimization through reinforcement learning, and integrating neurological data to balance cognitive load and decision-making. In addition to infrastructure expenses and design homogeneity difficulties, the study addresses organizational changes including hybrid roles and data-driven education improvements. The industry is moving toward automation and intelligence, as seen by future frontiers AI-generated interfaces, culturally appropriate designs, and interactive VR/AR systems. Ultimately, the article positions data-driven design as a creative blend of machine analysis and human creativity, emphasizing the need for a balanced approach to foster future innovations.

**Key words:** Big Data Analytics, User Interface, Machine Learning, A/B Testing, Personalization, Reinforcement Learning (RL), Data-Driven Design, Cognitive Load Optimization, Eye-Tracking Studies, Context-Aware Interfaces, Cultural Adaptation, VR/AR Interfaces, Design Education Reform, Real-Time Visualization

**Introduction.** The use of big data in graphic design and user interface (UI) creation is revolutionizing the possibilities of creativity and precision. In more competitive and complex digital marketplace conditions, demand for more efficient data-driven means has led to a revolutionary transition in the field of designers at work. Rather than relying on intuition or user input only, designers now utilize vast datasets and advanced analytical tools to direct each design step, from initial conception to final deployment. This revolution is not a peripheral extension of existing practice but a paradigmatic shift in thinking, producing, and iterating the visual communication system. The wealth of user behavior data now collected has generated previously unimaginable insights into human nature, tastes, and needs insights increasingly being employed to inform design decisions across the board.

Ever wondered how their "Discover Weekly" playlist appears to have an intuition about what you're looking for? It's big data working its magic. Spotify sifts through billions of listening events songs skipped, playlists chosen, even pause button presses to make recommendations better. But it's not just music; it's design. The app's layout, with its minimalist design and prominent playlist tiles, was influenced by data showing users crave speedy, personalized access. A white paper by

Spotify's design team, released in 2021, revealed that UI changes driven by data improved user retention by 12% during a single quarter. That isn't coincidence that's success. Let's home in on Netflix, the poster child for data-driven design. (fig.1).

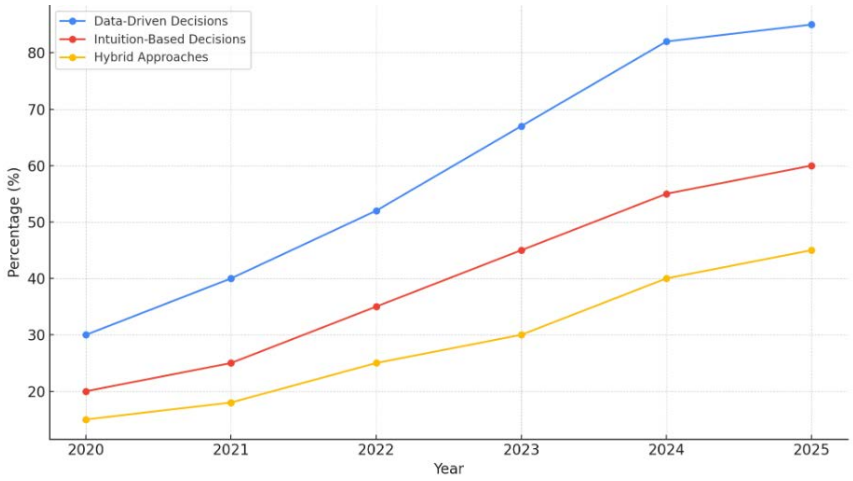



*Figure 1.* Big Data Impact on Design Decision-Making

Every time you browse their catalog, you're employing an interface sculpted by big data. They don't take a guess at which thumbnail to make you click on Stranger Things; they know. How? Through ongoing A/B testing and user behavior studies. Netflix knows everything about everything what you've watched, what you pause, even the time of day you stream [2]. A 2024 study from the International Journal of Human-Computer Studies found that their data-driven thumbnail technique observed a 20% boost in click-through rate over static images (fig.2).

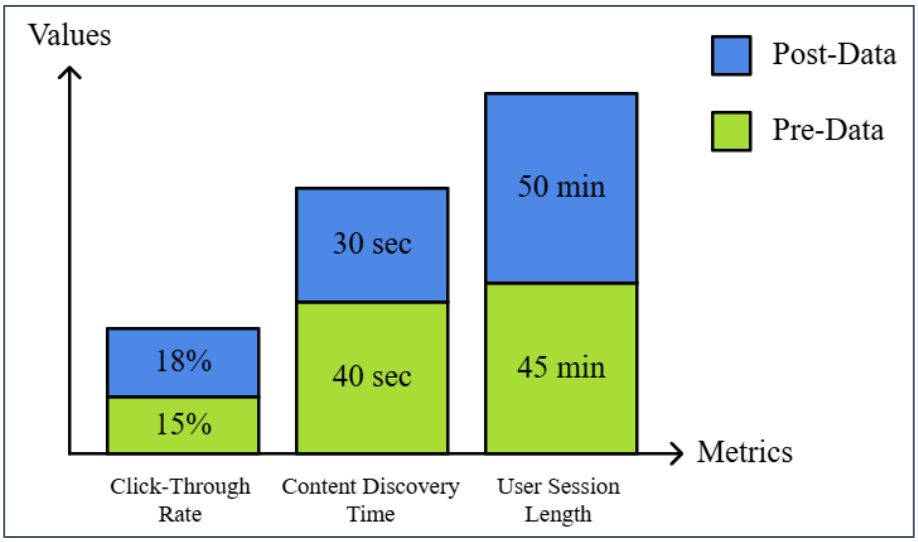



*Figure 2.* Impact of Data-Driven Design on Netflix UI Metrics

But it’s more than thumbnails. Netflix’s entire UI the way shows are grouped, the auto-play previews, the minimalist navigation is a product of data. They’ve tested countless layouts, using machine learning to predict which designs keep users hooked. One internal report (shared at the 2024 ACM CHI Conference) highlighted a redesign of their homepage that cut content discovery time by 25%, all thanks to data showing users wanted faster access to trending titles. That’s the kind of win that turns a good interface into a great one.

A study by Nielsen and Pernice (2023) showed that businesses using data-informed design processes had an impact on their key performance indicators improving them by 38% on average compared to those using old-school approaches. This big difference in performance has sped up

the use of these methods across many fields, from online shopping to healthcare, and from education to entertainment (fig.3).

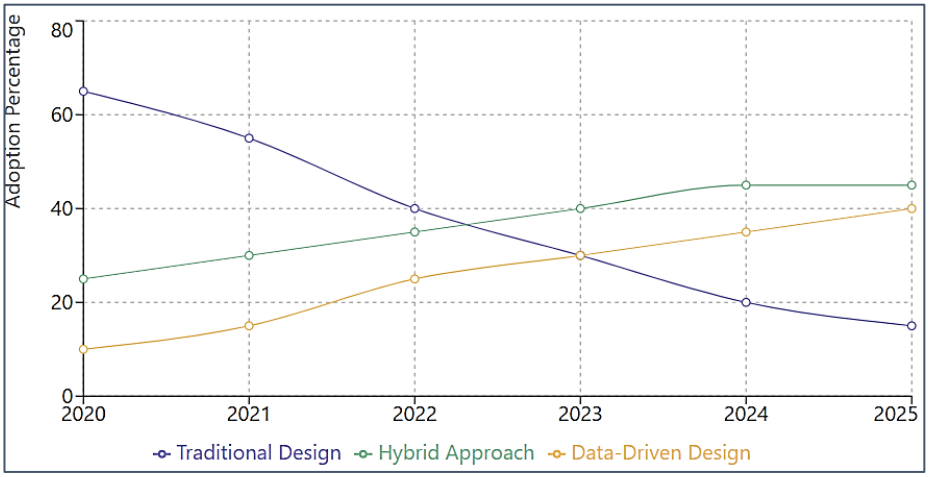



*Figure 3.* Design Methodology Adoption Trends

At the heart of this change is the ability to gather, handle, and make sense of huge amounts of data that we couldn't access or manage before. New tracking tools now record tiny details of how people use websites where their mouse goes, what they look at, what they click on how long they stay on a page how far they scroll, and many other things. This creates detailed pictures of how people behave showing not just what they do, but maybe why they do it [3]. These tracking systems have gotten much better over time. Zhang and his team (2024) found that new ways of collecting data without bothering users can guess what a person wants to do with 95% accuracy, while not slowing down the website or making it hard to use (fig.4).

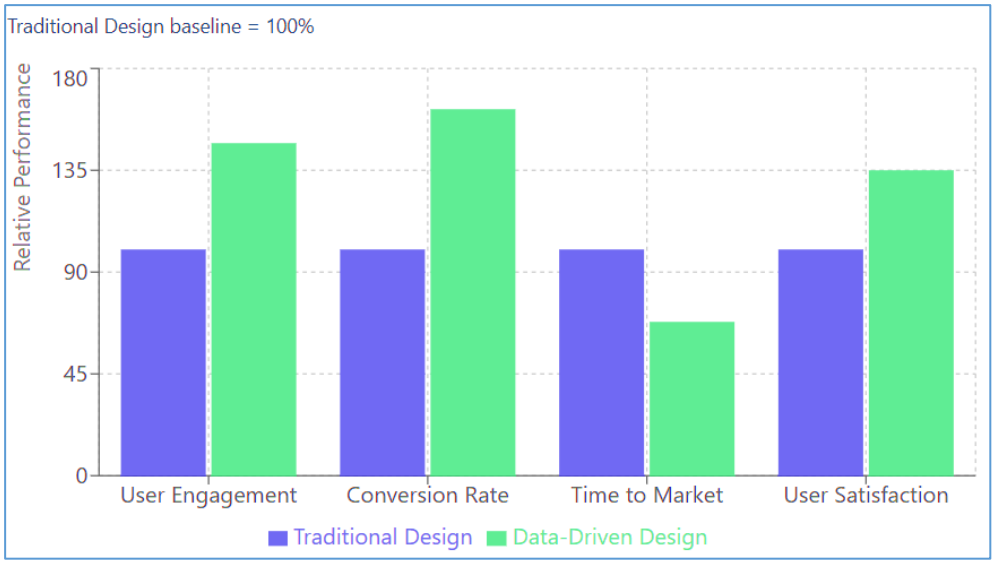



*Figure 4.* Performance Metrics Comparison (Traditional vs Data-Driven Design)

The application of this data begins at the earliest stages of the design process. Where designers once relied primarily on intuition, experience, and limited user testing to guide initial concepts, they now frequently begin with deep analysis of existing interaction patterns. A comprehensive study by Martínez and Johnson (2022) examined the design processes at 150 leading digital product companies and found that 78% now begin new projects with extensive data review sessions, analyzing both their own historical user data and industry benchmarks. This approach has demonstrably reduced initial development cycles, with the same study showing an

average 42% reduction in time-to-market for new features when data-informed ideation processes were employed [4].

The impact of big data extends well beyond the initial conceptualization phase. Throughout development, continuous data collection and analysis enable real-time optimization that was impossible in traditional workflows. Designers now commonly implement A/B testing at unprecedented scales, with some major platforms running hundreds of simultaneous experiments. According to research published in the Journal of Digital Experience Design, companies in the top quartile of A/B test volume showed user engagement metrics 53% higher than those in the bottom quartile (Ramirez et al., 2023). This correlation between testing volume and performance has driven the development of increasingly sophisticated testing infrastructures, with many organizations building custom platforms to manage their experimental design processes.

The advancement in modern user data allows for efficient tailoring of user experiences. Nowadays, websites and applications no longer offer the same interface to every user. It is able to change in real time according to user actions, interests, location, and other factors. A groundbreaking piece of work by Kovalenko and Parks (2023) showed that unique interfaces managed to achieve an average increase of 35% in conversion rates across a wide range of industries. The E-commerce industry saw the highest improvement of 49%, followed by subscription services at 41%. These personalization systems have become more advanced over time.

Some of them now incorporate machine learning technologies capable of adapting and perfecting their models independently as new data is generated through increased user activity [5].

The shift to data driven techniques have also transformed the interfaces' visual assets. The fundamental concepts of design still hold true, however their implementation is becoming increasingly supported by data instead of just artistic intent. Goldberg and Kotval's eye tracking study (2022) showed that visual hierarchy design principles do not always match user's expected actions which made many layout design best practices become much looser than before.

Their research showed that the interfaces redesigned post analysis of gaze tracking data exhibited a 28% higher rate in completing the given tasks than those who relied purely on conventional design principles (fig.5).

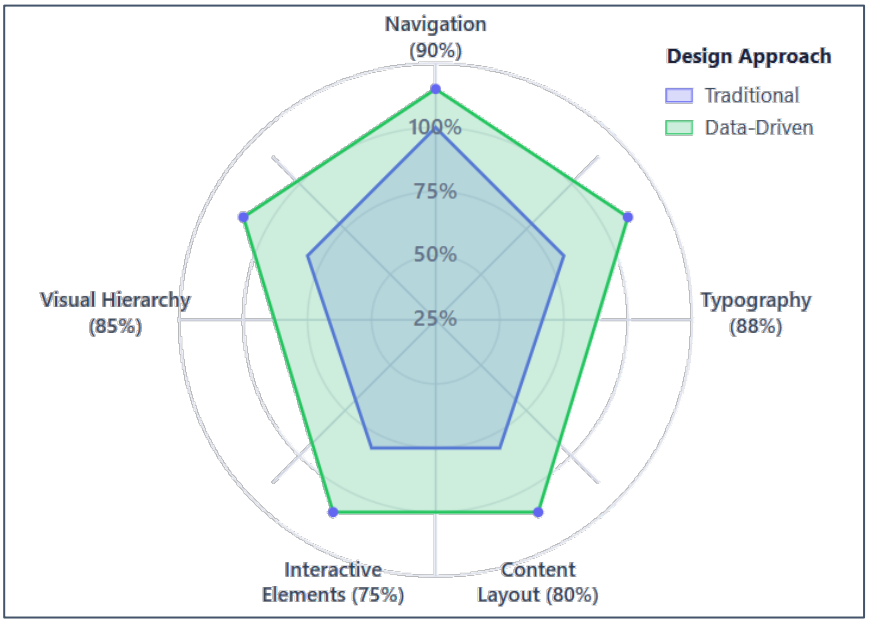



*Figure 5.* Impact of Data-Driven Design on UI Components

Color theory, typography, and information architecture are some of the design disciplines which in the same way have been optimized through data analysis. A series of user-based studies over several years done by Sharma et al. (2024) to compare user response to different color

combinations in various application types they result published a scientific evidence-supported framework for color selection which today is used by the main design systems. The data gave evidence of the fact that data-oriented color schemes were 23% better in understanding and 18% easier in cognitive processing in comparison with traditional schemes made by using the fairytale method. Readability has undergone the same treatment through the use of new readability algorithms developed by Tann and Garcia (2023) and in this way the right choice of font and text layout are now being implemented in many big digital platforms (fig.6).

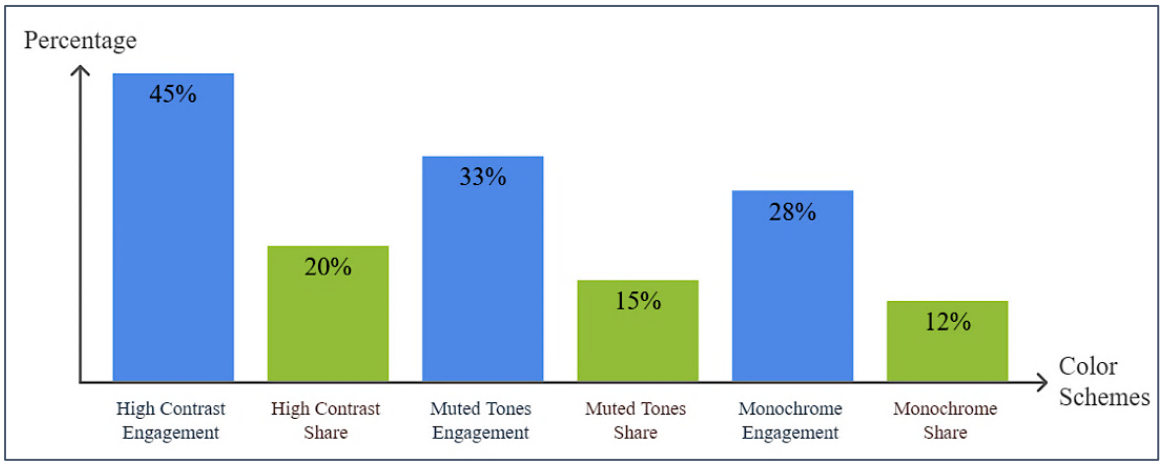



*Figure 6.* Performance of Different Color Schemes on Social Media

The advent of mobile computing has brought about new opportunities for the data-driven design world but at the same time posed some challenges. The combination of all the different types of devices, environments, and interaction modes has made the organizations take a turn to the more adaptive based on local data input also becoming more and more important with incredibly rich data of context that can be used to improve the compositions. In their investigation, Okonkwo et al. (2023) presented how location data, device orientation, ambient light intensity, and motion patterns could be integrated and able to dynamically change the user interface according to the user condition that eventually led to usability metrics' increase of 31% in comparison with the static arrangements. Contrast-aware interfaces that are this occupy now represent the inclination to create designs which are triggered not only by the direct user but all the factors of the environmental nature which impact the user needs and talents [6].

Following the application of data-driven design, the introduction of the very first of its kind, has not been devoid of certain problems from the side of a technical point of view. The technical infrastructure that is needed to gather, process, and analyze the users’ actions, which are the main components of the system they are operating, is very extensive hence the need to provide a large sum of cash for the technology and skill. One survey of McKinsey Digital (2023) found that those companies who succeeded in transferring to a more data-driven design mode invested 15% of the entire design budget on average into analytics capabilities, while best results were achieved by the ones with the biggest investment (24%). Fully comprehended, the monetary capital that was raised went to just the devices that were needed for development. Moreover, the skill in the process was acquired as the traditional design fields were closely related and bridged with the mathematical ability previously owned by data analysts only.

The effect of big data on design education has been as big. The best design schools have taken considerable efforts in modifying the curricula to bring data analysis into it, while some institutions are offering new degree programs specifically on the design aspect of data science as well. According to a global DESIGN education survey by the Design Management Institute (2023), out of the top 83% of the programs required coursework in data analysis, 65% had specialized data-driven design tracks. These changes in education are the response to the industry changes with the same survey stating that 76% of employers now indicate data literacy as a fundamental skill for design roles [7].

Even though data-driven design advantages are extensively studied, big discrepancies in mastery of the discipline are still operating differently among organizations and sectors hence research in the areas is also variant. A study from Deloitte Digital (2024) outlined five stages of maturity in data-driven design, namely, "Basic" (using basic analytics to track some parameters), "Transformative" (using predictive modeling and machine learning to forecast user demands and automatically adjust designs). Their study indicated that only 12% of organizations could be considered to the highest maturity level of data practice, while the majority of organizations (57%) were still in the middle of development stage. This difference implies that there are a lot of new opportunities for growth despite the development of the field.

The approaches to computer-driven design are most likely to go in the direction of higher automation and intelligence. The prototypes built by teams from MIT and Stanford (Venkatesan et al., 2024) proved that artificial intelligence can generate feasible designs of the interface based on the functional necessities and user information, with the designers rather being curatorial and strategic. Even though this idea of totally computerized design is more of an idea than practicability in very intricate areas, still the system has shown very good results for particular design components and simple interfaces [8].

Applying neurological data has another place the evidence-based design. Neuroscience and design combined with advanced research have made the first steps in advancing the design process by not only measuring the conventional metrics such as clicks and page views but also tracking the physical and psychological feedbacks to the design elements. Neumeier and Casillas' (2023) study used EEG and fMRI data to find interface patterns that yielded the least cognitive load and the most information retention, thus they set new standards for the information-based applications like financial dashboards and medical interfaces. Their study provided a 28% increase in precision and a 35% decrease in decision time, if the design of interfaces was then bettered with the neurological data. Stitching of big data with new technologies like augmented and virtual reality introduces a myriad of fresh chances and problems for the data-driven design. Those supporting the notion of creating immersive environments can now generate new forms of interaction data such as the movement of three-dimensional objects, physiological responses, and environmental adaptation. This additional information can be used to design more natural and intuitive interfaces. The work by Oculus Research Labs (Patel et al., 2024) has pioneered data-driven interfaces to VR that reveal a motion sickness reduction of 67% and an increase in spatial task performance by 43% in comparison to orthodoxly designed interfaces available in the market.

Not only voice interfaces but conversational UI has gained from this data-driven method. By gathering natural language interactions, machines were able to accomplish the development of more conversational patterns and users' intent prediction with increased accuracy. Amazon's conversational interfaces achieved the greater satisfaction ratings and task completion rates up to 31% and 28% respectively, in comparison with those designed by traditional methods through analyzing millions of real-world interactions. These upgrades are aimed mainly at showing the value of large-scale data analysis that can capture the small subtleties of human communication that might be missed in limited user testing scenarios [9].

For the arts and design which extend far beyond the interfaces, such as brand identity, marketing materials, and the visual communication, to enable big data has granted them more precision targeting and customization. To measure the power of the picture, graphic designers can use artificial intelligence, which nowadays can calculate the emotional attachment and prediction ability of the image for a specific audience. Visual AI Labs (Krishnamurthy et al., 2024) discovered how the machine learning models trained on consumer response data could predict the performance of design variants with 83% accuracy, and thus, letting the customer know what would be the best, even before the product is launched. These skills have designer logos evolve from the creative process to a mix of aesthetic innovation and empirical validation [10].

Despite these tensions, the trajectory of big data in design appears to be one of continued growth and deepening integration. As artificial intelligence and machine learning techniques

become more sophisticated, the ability to derive meaningful insights from complex, multivariate datasets continues to improve. Research by Google's Design Lab (Suryono et al., 2024) demonstrated how advanced AI techniques could identify subtle interaction patterns that human analysts had missed, leading to counter-intuitive but highly effective design solutions that improved key metrics by up to 26%. These capabilities suggest that the relationship between data analysis and design innovation will continue to evolve in ways that complement rather than replace human creativity.

For individual designers, the rise of data-driven methodologies has necessitated significant adaptation and skill development. A longitudinal study by the AIGA (2023) tracked the career trajectories of 500 designers over five years and found that those who developed data literacy alongside their creative skills showed 37% higher salary growth and 42% more rapid career advancement than those who remained focused exclusively on traditional design skills. This economic reality has driven growing demand for continuing education in data analysis among practicing designers, with major platforms like Coursera and LinkedIn Learning reporting triple-digit growth in enrollment for data-focused design courses [11].

**Conclusion.** The integration of big data in design practices shows no signs of slowing. As computational power continues to increase, as new data collection methods emerge, and as analytical techniques grow more sophisticated, the relationship between data and design will likely continue to deepen and evolve. What remains constant is the fundamental purpose of design: to facilitate meaningful human experiences. While the methods for achieving this purpose are being transformed by data, the ultimate measure of success remains the quality of those experiences something that combines quantitative performance with qualitative values that may always require human judgment to fully assess.

What is clear is that the future of design will neither abandon data in favor of pure intuition, nor reduce design to mechanical optimization of metrics. Rather, the most successful approaches will continue to find ways to integrate these complementary forces using data to inform and enhance human creativity, while applying creative intelligence to determine which questions to ask of the data and how to translate its answers into meaningful human experiences. This balanced approach represents not just the current state of the field, but its most promising path forward.

## References

[1] Al-Nami, B.A., Musaeva, T. Modern Methods to Solve the Problems of Adapting Images in Web Pages and the Possibility of Using Artificial Intelligence to Do So in the Future // Current Issues in Infotelecommunications in Science and Education (APINO 2023). Collection of Scientific Articles of the XII International Scientific and Technical and Scientific-Methodological Conference. In 4 volumes. Edited by S.I. Makarenk; compiled by V.S. Yelagin, E.A. Anikevich. St. Petersburg, 2023. pp. 180–185.

[2] Al-Nami, B.A. AI – The Future Entity // Informatics, Control Systems, Mathematical and Computer Modeling (IUSMCM-2024). XV International Scientific and Technical Conference within the framework of the X International Scientific Forum of the Donetsk People's Republic. Donetsk, 2024. pp. 230–234.

[3] Al-Nami, B.A., Pigareva, Y.R. Development of the User Interface of a Logistics Company's Website // Current Issues in Infotelecommunications in Science and Education. Proceedings of the XIII International Scientific and Technical and Scientific-Methodological Conference. St. Petersburg, 2024. pp. 41–44.

[4] Al-Nami, B.A., Popova, E.V. Innovative Trends in Web Design: The Future of User Experience in a Digital Environment // Current Issues in Infotelecommunications in Science and Education. Collection of Scientific Articles of the XIII International Scientific and Technical and Scientific-Methodological Conference. St. Petersburg, 2024. pp. 506–510.

[5] Al-Nami, B.A. Can Big Data and Advanced Analytics Address Problems of Ethical, Privacy, and Equity Concerns in University Education? // BIG DATA and Advanced Analytics. Collection of Scientific Articles of the X International Scientific and Practical Conference in Two Parts. Minsk, 2024. pp. 55–66.

[6] Al-Nami, B.A. The Problems and Obstacles Facing the Designers of the Virtual World // Informatics, Control Systems, Mathematical and Computer Modeling (IUSMCM-2023). Proceedings of the XIV International Scientific and Technical Conference within the framework of the IX International Scientific Forum of the Donetsk People's Republic. Donetsk, 2023. pp. 367–371.

[7] Al-Nami, B.A. Converting Images from One Extension to Another Without Losing Quality: Problems and Solutions // In: BIG DATA and High-Level Analysis. Collection of Scientific Articles of the IX International Scientific and Practical Conference: in 2 Parts. Minsk, 2023. pp. 51–61.

[8] Al-Nami, B.A., Grigorieva, S.D. The Impact of Website Design on Online Business Sales // Current Issues in Infotelecommunications in Science and Education (APINO 2023). Collection of Scientific Articles of the XII International Scientific and Technical and Scientific-Methodological Conference. In 4 volumes. Edited by S.I. Makarenk; compiled by V.S. Yelagin, E.A. Anikevich. St. Petersburg, 2023. pp. 113–116.

[9] Pigareva, Y.R., Al-Nami, B.A., Kalinina, S.S. The Influence of User Design on the Website of a Logistics Company // Informatics, Control Systems, Mathematical and Computer Modeling (IUSMCM-2024). XV International Scientific and Technical Conference within the framework of the X International Scientific Forum of the Donetsk People's Republic. Donetsk, 2024. pp. 225–228.

[10] Al-Nami, B.A., Rstakyan, A.A. The Interconnection Between Design and Marketing: How Design Influences Sales // Current Issues in Infotelecommunications in Science and Education (APINO 2023). Collection of Scientific Articles of the XII International Scientific and Technical and Scientific-Methodological Conference. In 4 volumes. Edited by S.I. Makarenk; compiled by V.S. Yelagin, E.A. Anikevich. St. Petersburg, 2023. pp. 206–209.

[11] Rui-Qiang He, Wang-Sen Lan, Fang Liu, "MRWM: A Multiple Residual Wasserstein Driven Model for Image Denoising", IEEE Access, vol.10, pp.127397-127411, 2022.

**Авторский вклад**

**Аль-Нами Башер Али** – Окончил Санкт-Петербургский электротехнический университет «ЛЭТИ». Научные интересы связаны с проектированием виртуальных средств информационных технологий, 3d полигональным моделированием и эргономикой, стандартизацией и проектированием (юзабилити) программных продуктов, Web-проектированием и дизайном, информатикой и компьютерным дизайном. Визуальное проектирование интерфейсов цифровых продуктов, информационные технологии, Информационные технологии визуализации данных и исследовательских процессов в Санкт-Петербургском государственном университете телекоммуникаций им. профессора М. А. Бонч-Бруевича.

# РОЛЬ БОЛЬШИХ ДАННЫХ В ГРАФИЧЕСКОМ ДИЗАЙНЕ И ПОЛЬЗОВАТЕЛЬСКИХ ИНТЕРФЕЙСАХ ПОДХОД ОРИЕНТИРОВАННЫЙ НА ДАННЫЕ ОТ ИДЕИ ДО РЕАЛИЗАЦИИ

***Б.А. Аль-Нами***
*Доцент кафедры информатики и компьютерного дизайна СПбГУТ,*
*кандидат технических наук, доцент*

**Аннотация.** В этой статье рассматривается разрушительный потенциал больших данных в графическом дизайне и пользовательском интерфейсе, а также описываются его эффекты от начала до развертывания. В ней описывается, как глубокое понимание поведения пользователя становится возможным благодаря современным технологиям мониторинга и машинному обучению, которые возвращаются в высоконастраиваемые, адаптируемые интерфейсы, выходящие за рамки традиционных дизайнов. Важные стратегии включают масштабное A/B-тестирование, оптимизацию в реальном времени с помощью обучения с подкреплением и интеграцию неврологических данных для балансировки когнитивной нагрузки и принятия решений. Помимо расходов на инфраструктуру и трудностей однородности дизайна, в исследовании рассматриваются организационные изменения, включая гибридные роли и улучшения в образовании на основе данных. Отрасль движется к автоматизации и интеллекту, как это видно на будущих рубежах интерфейсов, созданных с помощью ИИ, культурно приемлемых дизайнов и интерактивных систем VR/AR. В конечном счете, статья позиционирует дизайн на основе данных как творческое сочетание машинного анализа и человеческого творчества, подчеркивая необходимость сбалансированного подхода для содействия будущим инновациям.

**Ключевые слова:** аналитика больших данных, пользовательский интерфейс, машинное обучение, A/B-тестирование, персонализация, обучение с подкреплением (RL), проектирование на основе данных, оптимизация когнитивной нагрузки, исследования слежения за движениями глаз, контекстно-зависимые интерфейсы, культурная адаптация, интерфейсы виртуальной и дополненной реальности, реформа образования в области дизайна, визуализация в реальном времени.

УДК 338.4

# БИЗНЕС-ЭКОСИСТЕМЫ И ИХ РОЛЬ В ЦИФРОВОЙ ТРАНСФОРМАЦИИ

***Л.И. Архипова***
*Доцент кафедры экономики*
*Белорусского государственного университета*
*информатики и радиоэлектроники,*
*кандидат экономических наук, доцент*
l.arkhipova@gmail.com

***Л.И. Архипова***

*Окончила Белорусский государственный университет и Академию управления при Президенте Республики Беларусь. Имеет многолетний опыт работы в реальном секторе экономики (НПО «Интеграл»). Работает в Белорусском государственном университете информатики и радиоэлектроники. Проводит научные исследования в области современных технологий менеджмента и цифровой трансформации бизнеса.*

**Аннотация.** В статье рассматривается сущность понятия и роль бизнес-экосистем, как концепции современного бизнеса, дается характеристика ее структуры и ключевых компонентов. Реальная потребность создания бизнес-экосистем сегодня диктуется самим бизнесом. В современной бизнес-среде ни одна организация не может создавать все продукты и процессы абсолютно самостоятельно. Именно поэтому большинство предприятий ищут различные формы сотрудничества с другими организациями, особенно при создании новых цифровых бизнес-моделей.

**Ключевые слова:** бизнес-экосистема, цели и функции бизнес-экосистем**,** цифровой бизнес, модель управления, интеграция партерского и клиентского сервиса, цифровые технологии, клиентский опыт, омниканальность.

**Введение.** Бизнес-экосистема представляет собой сеть взаимосвязанных компаний, платформ, технологий, клиентов и партнёров, которые совместно создают ценность, обмениваются ресурсами и данными, усиливая конкурентные преимущества друг друга. Примерами наиболее успешных бизнес-экосистем являются: Amazon (маркетплейс/логистика/облачные сервисы AWS) и Alibaba (электронная коммерция/платежные системы/аналитика) [1-4].

Основными ключевыми признаками бизнес-экосистем являются [5, 6]:

– взаимозависимость участников, когда каждый элемент экосистемы получает выгоду от взаимодействия;

– общая инфраструктура, где платформы, API и облачные технологии объединяют участников;

– совместное использование данных, когда данные клиентов, транзакций и поведения анализируются для улучшения продуктов и сервисов.

Отчет консалтинговой компании (Gartner Hype Cycle) в сфере цифровых бизнес-возможностей включает четыре направления развития бизнеса – это концепции, которые, в том числе, определяют содержание современного менеджмента управления данными: Индустрия 4.0; бизнес-экосистемы; цифровые двойники, цифровые агенты или машинные клиенты (рисунок 1).

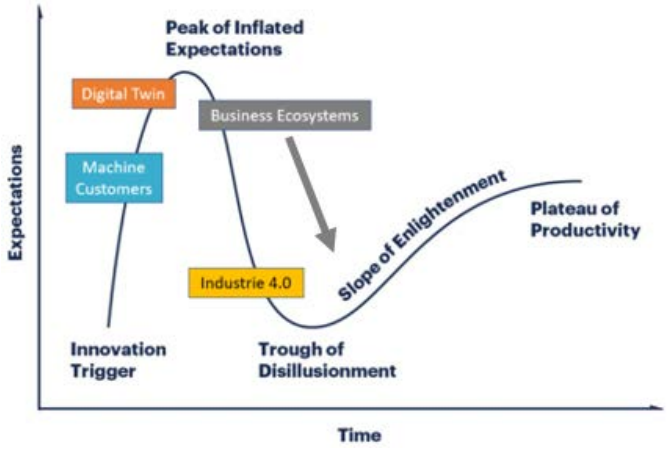



*Рисунок 1.* Gartner Hype Cycle: четыре концепции развития цифрового бизнеса [7]

Достаточно зрелой и реально движущейся к плато продуктивности является концепция построения *бизнес-экосистем,* которая прошла пик завышенных ожиданий – на рынке уже представлены цифровые платформы и инструменты, находящиеся на стадии зрелости и готовые к их практическому использованию. Бизнес-экосистема дает возможность частично или полностью решать проблемы бизнеса, как на стадии внедрения, так и масштабирования, за счет цифровых технологий и инструментов «коллективного пользования».

**Анализ структуры и основных компонентов бизнес-экосистемы.** На рисунке 2 представлены основные цели, задачи и функции, которые реализуются в рамках бизнес-экосистем [5-7]

| Цели бизнес-экосистем | Задачи бизнес-экосистем | Функции бизнес-экосистем |
|---|---|---|
| • *Создание ценности*: цель- создание дополнительной ценности для клиентов через интеграцию услуг и продуктов.<br>• *Увеличение инноваций*: объединение ресурсов способствует более быстрому внедрению инноваций.<br>• *Расширение рынка*: совместный выход на новые рынки и сегменты.<br>• *Устойчивое развитие*: бизнес-экосистемы объединяют усилия для решения социальных и экологических проблем | • *Синергия ресурсов*: цель-объединение ресурсов и компетенций для достижения целей.<br>• *Управление взаимодействиями:* управление взаимодействиями между участниками экосистемы для минимизации конфликтов и максимизации выгод.<br>• *Адаптация к изменениям*: быстрая адаптация к изменениям рынка и технологий.<br>• *Создание доверия*: обеспечение долгосрочного сотрудничества | • *Инновационная платформа*: цель - создание экосистемы, как платформы для совместного создания новых продуктов.<br>• *Обмен знаниями*: обмен знаниями и опытом способствует росту бизнеса участников.<br>• *Поддержка клиентов*: комплексное обслуживание клиентов через интеграцию услуг и продуктов.<br>• *Аналитика и данные*: сбор и анализ данных о предпочтениях потребителей для улучшения предложений |

*Рисунок 2*. Цели, задачи и функции бизнес-экосистемы

Описанные цифровые цели, задачи и функции играют ключевую роль в реализации системного подхода концепции бизнес-экосистемы, обеспечивая эффективное взаимодействие участников, оптимизацию процессов и создание ценности для клиентов (рисунок 3) [5-8].

*Рисунок 3.* Структура бизнес-экосистемы

Построение бизнес-экосистемы требует фокусирования на таких аспектах, как:

1Определение ядра экосистемы – это выбор ключевого продукта или сервиса, вокруг которого будет строиться экосистема.

2Привлечение партнеров – это разработчики, поставщики, логистические компании и др.

3Создание открытых API – это возможность третьим сторонам интегрироваться с вашей платформой.

4Инвестирование в данные – это значит создать аналитику и машинное обучение для оптимизации взаимодействий. Здесь следует особое внимание уделить таким важным методологиям, как Master Data Management (MDM) и Data Driven Business, которые являются взаимодополняющими и имеют ряд сходств и различий.

Обе методологии подчеркивают важность данных, как основного актива бизнеса, при этом MDM обеспечивает качественное управление данными, а Data Driven Business использует данные и аналитические инструменты (бизнес-аналитику, машинное обучение и визуализация данных) для принятия обоснованных решений.

Как подчеркивалось ранее, цифровая экосистема представляет собой бесшовную цифровую среду, в которой представлены собственные и партнерские сервисы компании. Экосистемное направление развития бизнеса характеризуется рядом особенностей [7-9]:

- обеспечивает высокую конкурентоспособность всех участников экосистемы;
- генерирует прибыль от нетрадиционных видов деятельности;
- увеличивает пользовательскую базу и снижает затраты на привлечение клиентов.

Интеграция сервисов в рамках цифровой экосистемы обеспечивает появление «бесшовного пользовательского опыта», который подразумевает переключение клиента между различными сервисами экосистемы без барьеров. Для клиента такая бесшовная мульти-сервисная среда включает:

- возможность использовать единый логин и пароль в разных продуктах;
- возможность не вводить многократно свои данные в профилях разных сервисов;
- доступность нужных сервисов в разных интерфейсах экосистемы;
- использование единого платежного инструмента, подписки, бонусной программы;
- просмотр релевантного контента и предложений, др.

Каждый бизнес-домен экосистемы является каналом привлечения клиентской аудитории для других сервисов, а также элементом, который препятствует выходу клиента из экосистемы. Поэтому включение нового клиента в экосистему происходит по заранее спроектированному клиентскому пути, чтобы работа с одним сервисом упрощала клиенту работу с другими сервисами платформы.

Бизнес-экосистемы создают *целостный опыт для клиента* и конкурируют за внимание и время клиента. Это определяет специфические требования к *интеграции нового клиента* в экосистему. Специалисты считают, что путь клиента (Customer Journey Map) должен быть детально спроектирован, что сделает переход между сервисами «естественным» и

превратит разовые покупки в долгосрочные отношения. При таком подходе клиент находит нужные продукты внутри экосистемы, постепенно подключается к дополнительным сервисам и не сталкивается с барьерами при переходе между этапами, что повышает лояльность клиентов и монетизацию бизнеса [7-9].

Каждая экосистема стремится расширять набор решений, которые создают простой и удобный переход клиентов между продуктами и бизнес-доменами (рисунок 4).

*Рисунок 4.* Интеграция или переход к владению частью экосистемы

Реализация таких решений невозможна без внедрения «сквозных кросс-доменных и технологических» сервисов. Они являются архитектурными доминантами, обеспечивающими сохранение клиентского опыта – главного преимущества и главного отличия экосистемной бизнес-модели. *Без использования кросс-доменных сервисов экосистема останется набором разрозненных самостоятельных клиентских продуктов.*

Используя «перекрестное» обогащение знаниями о клиенте, компании создают сложные механизмы анализа клиентского поведения, которые должны предсказывать желания и потребности клиентов, а также предлагать релевантную продукцию – товары, контент, услуги. На таком подходе построены концепции Next Best Offer (NBO) и Next Best Action (NBA). Ключевыми инструментами для проектирования пути клиента являются:

- карта пути клиента (customer journey map);
- визуализация всех точек взаимодействия;
- аналитика данных – отслеживание поведения клиента в реальном времени;
- персонализация – использование ИИ для выработки рекомендаций;
- омниканальность – использование единого клиентского опыта на всех каналах платформы (онлайн, офлайн, мобильное приложение).

Важно отметить, что каждое решение в бизнес-экосистеме может быть адаптировано под конкретные запросы потребителей и брендов, что обеспечивается, в том числе, и условиями *омниканальности.*

**Омниканальность в бизнес-экосистемах** [7-9]. В данном контексте омниканальность рассматривается как подход, при котором компании взаимодействуют с клиентами через различные каналы (онлайн и офлайн) таким образом, чтобы обеспечить единый и бесшовный опыт на каждом этапе их взаимодействия с брендом. Этот путь включает в себя использование физического магазина, сайта, мобильного приложения, социальных сетей и других каналов. Основные составляющие омниканальности в экосистемах бизнеса представлены на рисунке 5.

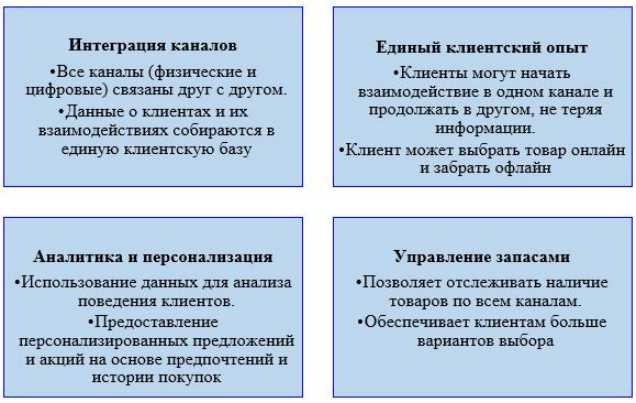



*Рисунок 5.* Ключевые составляющие омниканальности в бизнес-экосистеме

Омниканальность **рассматривается в данном контексте** не просто, как тренд, а как необходимое условие для успешного ведения бизнеса в современной конкурентной среде.

Основным преимуществом принципов омниканальности, реализуемых в бизнес-экосистемах, являются:

– **увеличение продаж – создается** возможность перекрестных продаж и высокий уровень сервиса;

– **эффективное управление ресурсами – достигается за счет** оптимизация логистики и управления запасами.

Следует отметить некоторые отличительные особенности моделей управления бизнес-экосистемами, это [9 -11]:

*Модульность* **–** компоненты в этой структуре чаще разрабатываются независимо, а затем объединяются на одной платформе, предоставляя пользователю возможность выбора требуемых комбинаций.

*Многосторонние отношения* **–** деятельность одного участника экосистемы тесно связана с функционированием остальных.

*Координация* **–** бизнес-экосистемы основаны не на иерархическом контроле, а на специфическом механизме координации, где помимо рыночных механизмов включается регулирование через стандарты, правила, процессы или платформы.

Работа с платформами бизнес-экосистем предоставляет множество возможностей для бизнеса, однако она также сопряжена с рядом рисков и недостатков, которые могут быть обусловлены такими факторами, как: безопасность данных; зависимость от третьих сторон; сложность интеграции; недостаток контроля; конфликты интересов; правовые и регуляторные риски; культурные различия; изменения в рыночной среде. Именно поэтому компании-участники бизнес-экосистем должны оценивать эти риски и разрабатывать стратегии для их минимизации, максимально эффективно используя возможности, которые предоставляют экосистемы.

**Заключение**. Создание бизнес-экосистем является актуальным подходом в современном бизнесе, где компании стремятся к сотрудничеству и интеграции с различными партнерами для достижения общих целей в условиях в цифровой трансформации. Бизнес-экосистемы создают необходимые условия для реализации успешных бизнес-проектов цифровой трансформации, а именно:

– дают ответ на вызовы цифровой эпохи, связанные динамикой и скоростью рыночных изменений;

– обеспечивают условия для персонализации, используя данные из экосистемы для предсказания потребности клиентов и предложения индивидуальных решений;

– обеспечивают возможность масштабирования и синергии за счет доступа к новым рынкам, когда партнёры экосистемы расширяют охват;

– оптимизируют и снижают затраты за счет совместного использования цифровых технологий и платформ, что уменьшает расходы на IT-инфраструктуру;

– создают условия для инноваций через открытые взаимодействия – краудсорсинг идей, когда партнёры и клиенты становятся источником инноваций;

– способствуют удержанию клиентов за счет замкнутого цикла взаимодействий, когда клиенты получают всё в одном месте (покупки, доставка, поддержка).

Таким образом, бизнес-экосистема представляет собой сеть взаимосвязанных организаций, включая поставщиков, партнеров, клиентов и (иногда) конкурентов, которые взаимодействуют друг с другом для создания и доставки ценности для клиентов. Эта концепция основана на идее, что компании могут более эффективно достигать своих целей, работая в сотрудничестве.

## Список литературы

[1] Кулапов, М. Н. Бизнес-экосистемы: определения, типологии, практики развития / М. Н. Кулапов, Е. И. Переверзева, О. Ю. Кириллова // Вопросы инновационной экономики. – 2022. – Т. 12, № 3. – С. 1597-1612 https://1economic.ru/lib/115234. – Дата доступа: 05.02.2025

[2] Бизнес-экосистемы: понятие, преимущества, прогнозы [Электронный ресурс]. – Режим доступа: https://sales-generator.ru/blog/biznes-ekosistemy/. Дата доступа: 10.01.2025

[3] Бизнес-экосистема: понятие, структура и примеры успешных решений [Электронный ресурс]. – Режим доступа: https://grandawards.ru/blog/biznes-ehkosistema-kompanii/. – Дата доступа: 16.01.2025

[4] Абрамов, В. И. Создание региональных бизнес-экосистем на основе цифровых профилей клиентов и омниканальных коммуникаций / В. И. Абрамов, В. В. Гордеев, А. Д. Столяров // Экономика, предпринимательство и право. – 2023. – Т. 13, № 5. – С. 1521-1540.

[5] Попов, Е. В. Эволюция бизнес-экосистем в промышленности - от классического типа к цифровым / Е. В. Попов, В. Л. Симонова, А. С. Зырянов // Информатизация в цифровой экономике. – 2024. – Т. 5, № 3. – С. 341-360.

[6] What business ecosystem means and why it matters [Электронный ресурс]. – Режим доступа: https://www.ey.com/en_uk/alliances/what-business-ecosystem-means-and-why-it-matters#:~:text=%22A%20business%20ecosystem%20is%20a,the%20orchestrator%20of%20the%20participan. – Дата доступа: 10.01.2025

[7] 4 Concepts in the Gartner Hype Cycle for Digital Business Capabilities that will Shape MDM [Электронный ресурс]. – Режим доступа: https://liliendahl.com/2022/02/09/4-concepts-in-the-gartner-hype-cycle-for-digital-business-capabilities-that-will-shape-mdm/.– Дата доступа: 10.01.2025

[8] The business ecosystem «ways to play» [Электронный ресурс]. – Режим доступа: https://www.pwc.com/us/en/services/consulting/business-model-reinvention/ecosystems-platform/guide-to-business-ecosystems.html. – Дата доступа: 10.02.2025

[9] Что такое бизнес-экосистемы и зачем они нужны [Электронный ресурс]. – Режим доступа: https://trends.rbc.ru/trends/innovation/6087e5899a7947ed35fdbbf3?from=copy. – Дата доступа: 16.02.2025

[10] 7 steps to building an ecosystem strategy [Электронный ресурс]. – Режим доступа: https://www.visma.com/resources/content/7-steps-to-building-an-ecosystem-strategy. – Дата доступа: 02.02.2025

[11] The Evolution of Digital Business Ecosystems [Электронный ресурс]. – Режим доступа: https://www.linkedin.com/pulse/evolution-digital-business-ecosystems-ashraf-alsadiq/.– Дата доступа: 10.02.2025

# BUSINESS ECOSYSTEMS AND THEIR ROLE IN DIGITAL TRANSFORMATION

***L.I. Arkhipova***
*Associate professor, PhD,*
*Belarussian State University of*
*Informatics and Radioelectronics*

**Abstract.** The article examines the essence of the concept and the role of business ecosystems as a concept of modern business, describes its structure and key components. The real need to create business ecosystems today is dictated by the business itself. In today's business environment, no organization can create all products and processes on its own. This is why most businesses are looking for various forms of collaboration with other organizations, especially when creating new digital business models.

**Keywords**: business ecosystem, goals and functions of business ecosystems, digital business, management model, integration of partner and customer service, digital technologies, customer experience, omnichannel.

УДК 519.7

# ОТЛАДКА ПАРАЛЛЕЛЬНЫХ ПРОГРАММ РЕШЕНИЯ ЛОГИКО-КОМБИНАТОРНЫХ ЗАДАЧ

***Д.И. Черемисинов***
*ОИПИ НАН Беларуси, ведущий научный сотрудник, кандидат технических наук, доцент*
cher@newman.bas-net.by

***Л.Д. Черемисинова***
*ОИПИ НАН Беларуси, главный научный сотрудник, доктор технических наук, профессор*
cld@newman.bas-net.by

***Д.И. Черемисинов***
*Окончил радиофизический факультет Томского государственного университета. Область научных интересов: программирование, логическое проектирование и тестирование дискретных систем управления, реализация параллельных алгоритмов управления.*

***Л.Д. Черемисинова***
*Окончила радиофизический факультет Томского государственного университета. Область научных интересов: логико-комбинаторные задачи, логическое проектирование и тестирование дискретных систем управления, реализация параллельных алгоритмов управления.*

**Аннотация.** Рассматриваются проблемы, возникающие при разработке и отладке параллельных программ для решения логико-комбинаторных задач. Обсуждаются особенности методов разработки и отладки таких параллельных программ, а также методы предотвращения ошибок при их разработке. Показано, что алгоритмы решения логико-комбинаторных задач трудно эффективно распараллелить полностью автоматически традиционными методами, а использование известных методов отладки последовательных программ не позволяет обеспечить качественную верификацию таких параллельных программ.

**Ключевые слова:** параллельная программа, кластер, MPI, отладка программ, верификация.

**Введение.** В настоящее время большой интерес проявляется к формальным средствам проектирования параллельных/распределенных систем, которые имеют сложную структурную и функциональную организацию. Такими системами, например, является аппаратура вычислительных машин и комплексов, коммуникационные протоколы, системы управления технологическими процессами и т.д. В основе методов проектирования таких систем лежит решение комбинаторно-логических задач, в основе которых лежит перебор подмножеств элементов из конечного множества большой мощности и тестирование их на обладание некоторым заданным свойством. Решение комбинаторно-логических задач состоит в переборе пространства решений, которое во многих случаях делается в процессе обхода дерева поиска. Решение комбинаторно-логических задач, возникающих на практике, представляет собой нетривиальную задачу, которая требует значительных вычислительных ресурсов. Потребность в решении комбинаторных задач и, следовательно, в разработке комбинаторно-логических алгоритмов возникает во многих областях, в том числе таких, как автоматизированное проектирование, создание систем искусственного интеллекта, разработка сетей связи и криптография.

Отличительной особенностью логико-комбинаторных задач [1] является то, что основными объектами преобразований для них являются не числа, как для задач численного анализа, а логические объекты: векторы, матрицы, графы и т.д. Характерно для задач рассматриваемого класса также и то, что объемы перерабатываемой информации зачастую сопряжены с необходимостью перебора и анализа значительного числа вариантов промежуточных решений, которое растет экспоненциально с увеличением размерности параметров решаемых задач. Одна из возможностей уменьшения времени решения комбинаторно-логические задач состоит в том, чтобы разрабатывать параллельные алгоритмы их решения и выполнять их на нескольких процессорах вычислительной системы. Параллельная обработка при решении задач проектирования получает признание как средство повышения вычислительных возможностей новых инструментов автоматизированного проектирования.

Тривиальные параллельные алгоритмы, в которых распараллеливание достигается использованием для просмотра дерева поиска решений нужного количества процессоров, описаны в литературе практически для всех задач проектирования СБИС. Например, в книге [2] описаны параллельные алгоритмы для размещения и трассировки, проверки соблюдения правил проектирования, логического синтеза, генерации тестов, моделирования ошибок в схемах и моделирования на поведенческом уровне. Но, так же как и в последовательном случае, тривиальные параллельные алгоритмы оказались не эффективными.

Цикл разработки программ, реализующих параллельные алгоритмы, значительно более сложен, чем для их последовательных аналогов. Это имеет два важных аспекта. Прежде всего, разработка и отладка программ для параллельных вычислений являются более дорогостоящими, чем для их последовательных аналогов. Во-вторых, последовательные алгоритмы, которые имеют длительную историю развития и усовершенствования, часто выигрывают у параллельных программ из-за трудности разработки последних. Параллельные алгоритмы, также как и последовательные, разрабатываются на основе моделей решения задачи. Для некоторых задач модель решения изначально параллельна, а для большинства других задач чрезвычайно трудно найти такую модель. Сложно создавать и отлаживать даже последовательные программы решения комбинаторных задач, а параллелизм повышает этот уровень сложности в разы. Даже в самых, казалось бы, простых параллельных программах обнаруживаются фатальные ошибки.

В настоящей работе рассматриваются вопросы, связанные с разработкой и отладкой параллельных программ для решения логико-комбинаторных задач. Показано, что отладка параллельных программ в силу их специфики, порождаемой параллельностью, требует методов и инструментов, значительно отличающихся от аналогичных инструментов для последовательных программ. Основная сложность при проектировании параллельных алгоритмов заключается в обеспечении нужной координации взаимодействия между вычислительными процессами и последовательности доступа к разделяемым ресурсам.

**Особенности разработки параллельных программ.** В настоящее время для выполнения высокопроизводительные вычислений широко используются кластеры, представляющие собой совокупности процессоров, объединенных компьютерной сетью и способных работать совместно над решением одной и той же задачи [3, 4]. Кластеры стали общедоступными и дешевыми аппаратными платформами для выполнения параллельных вычислений. Реализация алгоритмов решения задач на основе кластерных систем предполагает создание параллельных программ, ветви которой решают отдельные подзадачи и выполняются параллельно на разных процессорах. Основной моделью параллельного программирования на кластере является модель передачи сообщений, которая обеспечивается интерфейсом MPI (англ. Message Passing Interface) [5]. MPI

предоставляет средство взаимодействия задач в процессе их решения, выполняющегося параллельно в разных адресных пространствах.

Главная цель, которая ставится при разработке параллельных алгоритмов и реализации их параллельными программами заключается в достижении как можно более высокой скорости решения задачи. Параллельные программы обычно более сложны в разработке и отладке, чем их последовательные аналоги. Проанализировать код таких программ на предмет выявления ошибок довольно сложно. Одна из основных трудностей параллельного программирования заключается в организации хорошо структурированной синхронизации параллельных процессов [6].

При разработке параллельных программ для кластеров целесообразно стремиться к разработке «хороших» параллельных алгоритмов, эффективность которых целесообразно оценивать коэффициентом ускорения $K = t_s / t_p$, где $t_s$ – время, затрачиваемое на решение задачи последовательным алгоритмом на одном процессоре, $t_p$ – время решения параллельным алгоритмом с использованием $N$ процессоров. Чем больше значение $K$, тем выше эффективность применения мультипроцессорной вычислительной системы. Другой важной характеристикой алгоритма является эффективность распараллеливания, которая определяется как отношение ускорения к числу $N$ процессоров: $E = K / N$.

Тривиальные параллельные алгоритмы плохи потому, что в этих алгоритмах коэффициент загрузки процессоров стремится к нулю при росте размерности задачи. В «хороших» параллельных алгоритмах коэффициент загрузки процессоров должен не зависеть от параметров сложности задачи. Параллельное вычисление, так же как и последовательное, предполагает решение задачи путем применения плана (программы) действий, выбираемых из предопределенного набора, причем каждое из действий состоит в изменении среды вычислений детерминированным способом. При распараллеливании алгоритма для кластерного компьютера требуется декомпозировать проблему на ряд слабо зависящих друг от друга независимых последовательно выполняемых задач, которые могут выполняться одновременно. Если в последовательном алгоритме задано как должно интерпретироваться действие в контексте данной программы, то в параллельном алгоритме нужно указать не только это, но также и то, когда и где действие может (или не может) быть применено.

Для параллельного решения любой задачи на мультипроцессорной системе выделяется группа из затребованного количества $N$ процессоров. Хотя процессоры внутри группы принципиально не отличаются друг от друга, тем не менее, процессор номер 0 часто выделяется особо и считается корневым или главным, а остальные $N$–1 процессоров – подчиненными. Типичной является следующая ситуация. Корневой процессор получает из внешней среды исходные данные для решаемой задачи, декомпозирует ее на $N$–1 частных подзадач и распределяет их между подчиненными процессорами, которые решают каждый свою задачу параллельно и посылают результаты решений корневому процессору. Из полученных частичных решений корневой процессор формирует итоговое решение задачи.

Алгоритмы распараллеливания логико-комбинаторных задач по способу декомпозиции последовательно выполняемого процесса на подпроцессы, выполняемые подчиненными процессорами, и по правилам формирования из частичных решений итогового решения исходной задачи распадаются на два следующих класса [7]: алгоритмы совместного решения подзадач процессорами и алгоритмы, работающие в режиме соревнования. Первые алгоритмы основаны на разбиении задачи на частные подзадачи для каждого из процессоров, при этом итоговое решение задачи формируется на основе решений всех частных подзадач, и время вычислений определяется временем работы самого загруженного (и потому медленного) процессора. Вторые алгоритмы построены так, что решение частной подзадачи, получаемое любым из процессоров в ходе соревнования по времени с другими, является решением всей целой задачи. В этом случае время решения задачи определяется временем работы самого «самого быстрого» процессора.

**Особенности отладки параллельных программ.** Сложность параллельного программирования находит отражение и в сложности отладке таких программ. Для отладки MPI-программ можно использовать следующие программные средства [8]:

- имеющиеся отладчики классического типа, которые выполняют все процессы MPI-программы в режиме отладки. Наиболее известным из таких отладчиков параллельных программ является коммерческий отладчик *TotalView* и свободно доступный отладчик *glb*, который однако не имеет специальных средств для отладки MPI-программ;
- средства отладочной версии *mpich* библиотеки MPI, которые не позволяют перехватывать внутренние ошибки MPI, но обнаруживают неправильное использование MPI, такое как ошибки несоответствия типов данных при посылке и получении сообщений;
- программные средства типа *MPI-CHECK*, которые ориентированы на программы определенного языка (для *MPI-CHECK* – это Фортран) и выполняют проверку типов аргументов и обнаруживают некоторые другие некорректности, такие как тупики.

Недостатками этих средств является их дороговизна и ограниченность: в отношении используемой платформы, языка программирования и представления исходного текста программы. Кроме того, они не предотвращают неправильное использование MPI и не содержат средств анализа ситуации, приведшей к ошибке.

Опыт отладки параллельных программ (в частности, для решения логико-комбинаторных задач) показывает, что традиционные отладчики и методы отладки последовательных программ не эффективны для отладки параллельных. Причина заключается в том, что параллельные программы в целом отличаются от последовательных по методу разработки и модели программы, и, кроме того, программы для решения логико-комбинаторных задач имеют особенности в методе распараллеливания.

Программа для кластерного компьютера представляет собой ряд последовательных процессов, которые иногда взаимодействуют, но большей частью выполняются независимо. Процессы программы написаны с учетом условия, что все процессы выполняют асинхронно, то есть, если один процесс нуждается в получении информации от другого, он приостанавливается и ожидает поступления этой информации, и это взаимодействие не затрагивает другие процессы.

Наиболее частыми признаками неправильной работы параллельной программы являются тяжелые отказы (например, нарушение сегментации) или зависание программы. В обоих случаях традиционные отладчики не позволяют установить причину ошибки, так как эти синдромы вызваны, как правило, нарушениями в организации взаимодействия параллельных процессов, а не ошибками в организации последовательности вычислений, вызванными погрешностями реализации алгоритма решения задачи. Например, частой ошибкой является зависание программы, которое вызывается тупиком, порождаемым взаимной блокировкой параллельно выполняемых процессов.

Другой возможной фатальной ошибкой при работе параллельных программ является наличие гонок между подпроцессами, которое состоит в том, что результат вычислений в информационно связанных между собой процессах зависит от их относительных скоростей выполнения. Явления гонок и тупиков возникают из-за асинхронности параллельно выполняющихся процессов. Возможность гонок между процессами в параллельных программах представляет одну из главных проблем их разработки. Хотя программы с гонками процессов могут быть правильно организованными, имеются алгоритмы, в которых их наличие может приводить к неправильной работе программы в отдельных случаях. Ситуация возникновения гонок не может быть легко воспроизведена при отладке, потому что при каждом пуске программа может вести себя по-разному. Очевидно, что информацию о наличии гонок трудно получить с применением традиционных отладчиков, так как выполнение программы по шагам кардинально меняет всю динамику протекания процессов в отлаживаемой программе.

Единственным прямым методом получения данных о динамике выполнения подпроцессов параллельной программы является метод временной трассировки, который заключается во вставке фрагментов кодов в программу, которые фиксируют физическое время начала их выполнения. В этом случае используется старый способ отладки методом включения отладочных сообщений. В MPI имеются специальные средства для этой цели (например, MARMOT [9]). Однако индивидуально выбранные места для вставки, позволяют лучше представлять динамику взаимодействия процессов, так как временная трассировка операторов MPI с помощью таких средств как *MARMOT* дает слишком много ненужной информации. Метод ручной вставки трассирующих сообщений не требует применения специальных инструментов и при умелом применении вполне заменяет автоматизированные инструменты временной трассировки. Такой метод использовался при разработке программного комплекса решения комбинаторных задач логического проектирования для кластерного суперкомпьютера [10].

**Методика предотвращения ошибок в параллельных программах.** На момент создания параллельной программы, как правило, имеется уже отлаженная реализация ее последовательного прототипа. Однако это не исключает необходимости отладки созданной по нему параллельной реализации. Параллельный вариант отличается от последовательного хотя бы тем, что он разрабатывается для другой ЭВМ (как правило, более мощной) и включает механизм взаимодействия одновременно работающих ветвей программы. И основной причиной возникновения ошибок является отсутствие системы при проведении распараллеливания алгоритма.

Устранять влияние ошибочных действий программиста при распараллеливании последовательной программы призваны методики, которые представляют собой решение проблемы трансформации последовательной модели программы в параллельную. Большое количество работ и программных средств посвящено распараллеливанию специального типа программ – программ обработки данных. В этих программах основным источником массового параллелизма являются программные циклы. Для этих программ возможно статическое распределение процессоров.

Программы для решения логико-комбинаторных задач не могут быть распараллелены таким методом, потому что структура циклов в этих программах изменяется в ходе вычислений. Алгоритм обхода пространства решений обычно рекурсивен, так как первоначальная задача может быть решена методом декомпозиции проблемы на задачи того же типа, но меньшей размерности. Рекурсия из соображений эффективности частично заменяется циклическими вычислениями с использованием стека. Структура циклов в таком вычислении полностью зависит от используемых исходных данных. Тривиальный подход к распараллеливанию для почти всех известных комбинаторно-логических задач оказался неэффективным. Примером может служить задача выполнимости КНФ (конъюнктивной нормальной формы), алгоритмы решения которой интенсивно исследуются из-за важности практических применений: при проектировании СБИС, в задачах искусственного интеллекта, составлении расписаний и т.д. Из-за практической важности решения этой задачи интенсивно исследуются и параллельные алгоритмы ее решения. Однако несмотря на все усилия лучшими программами для решения задачи выполнимости КНФ пока остаются последовательные программы [11]. Следует отметить, что для некоторых задач комбинаторно-логического характера все же разработаны эффективные параллельные алгоритмы [7, 12, 13]. Для кластерного суперкомьютера разработан комплекс параллельных алгоритмов и реализующих их программ для решения четырех типов задач: минимизация систем булевых функций в различных постановках, перемножение ДНФ, анализ ДНФ на тавтологию, решение систем логических уравнений [8, 14].

Причиной многих трудностей параллельного программирования является организация структурированного взаимодействия параллельных ветвей программы,

которая требует некоторой формы их синхронизации, которая связана с их упорядоченностью. Синхронизация и упорядоченность прямо влияют на параметр масштабируемости, который характеризует увеличение производительности программной реализации в зависимости от числа занятых процессоров. При идеальной масштабируемости увеличение числа процессоров вдвое должно приводить к выполнению задачи по крайней мере в два раза быстрее, но так бывает редко, чаще имеется верхний предел числа процессоров, выше которого использование дополнительных процессоров уже не уменьшает время, требуемое для выполнения задачи. В настоящее время известны два подхода к решению проблемы синхронизации:

- синхронная параллельная модель, в которой параллелизм допустим только в пределах данного шага решения;
- асинхронная параллельная модель, в которой программа представляет собой ряд последовательных процессов, которые иногда взаимодействуют, но большую часть времени работают независимо.

В противоположность синхронным, асинхронные модели ориентированы на коммуникации между процессами как средства управления параллельным выполнением. Различают две модели коммуникации: обмен сообщениями (как в MPI) или общая память. Однако, очевидно, что независимо от модели, коммуникация в асинхронных языках связана с синхронизацией. Таким образом, ключевой аспект асинхронных языков, в противоположность синхронным языкам, состоит в задании средств синхронизации процессов. Синхронизация в параллельной программной модели представляет собой реальную проблему, так как без синхронизации отсутствует детерминизм выполнения. Хотя существует много механизмов синхронизации в большинстве асинхронных параллельных языков, но они не обеспечивают явной спецификации детерминизма. Обычно в асинхронной системе предполагается, что передача сообщения между процессами происходит через канал взаимодействия, который вызывает ненулевую задержку передачи.

**Заключение.** Проанализированы методы отладки параллельных программ, реализующих логико-комбинаторные алгоритмы, а также методы предотвращения ошибок при их разработке. Показано, что автоматические методы распараллеливания алгоритмов решения логико-комбинаторных задач не позволяют построить эффективные параллельные программы (лучшие, чем последовательные). Эффективные алгоритмы решения наиболее важных в практическом плане логических задач требуют разработки алгоритмов, в которых параллельная обработка специально спланирована.

## Список литературы

[1] Закревский А.Д., Поттосин Ю.В., Черемисинова Л.Д. Основы логического проектирования. Книга 1. Комбинаторные алгоритмы дискретной математики. Мн: ОИПИ НАН Беларуси; 2004. – 226 с.

[2] Banerjee P. Parallel Algorithms For VLSI Computer-Aided Design. – Prentice Hall, Englewoods Cliffs, NJ; 1994.

[3] Абрамов С.М., Парамонов Н.Н., Анищенко В.В., Абламейко С.В. Принципы построения суперкомпьютеров семейства «СКИФ» и их реализация. Информатика. 2004; 1: 89–106.

[4] Черемисинов Д.И. Программная среда для распределенных вычислений, использующих кластерный компьютер. Автоматизация проектирования дискретных систем (Computer-Aided Design of Discrete Devices – CAD DD'07): материалы Шестой Междунар. конф., 14–15 ноября 2007 г., Минск. В 2 т. Т. 1. Минск: ОИПИ НАН Беларуси; 2007: 38–45.

[5] Message Passing Interface Forum. MPI: A Message-Passing Interface standard, version 1.1. http://www.mpi-forum.org/docs/. 1995.

[6] Крюков В.А. Разработка параллельных программ для вычислительных кластеров и сетей. Информационные технологии и вычислительные системы. 2003; 1-2.

[7] Торопов Н.Р. Параллельные логико-комбинаторные вычисления в среде MPI. Информатика. 2005; 3: 82-90.

[8] Черемисинов Д.И. Программный комплекс решения комбинаторных задач логического проектирования и защиты информации. Информационные технологии программы Союзного государства

«Триада». Основные результаты и перспективы. Сб.науч. трудов. –Минск: ОИПИ НАН Беларуси; 2010: 96–110.

[9] Krammer B., Müller M. S., Resch M. M. MPI Application Development Using the Analysis Tool MARMOT. Proceedings of International Conference on Computational Science. 2004: 464–471.

[10] Cheremisinov D.I., Cheremisinova L.D. The Technology of Programming for Claster Computerby by the Remote Terminal with OS Windows. Information Theories & Applications (IJ ITA). 2007;14(4): 381–388.

[11] Singer D. Parallel Resolution of the Satisfiability Problem: A Survey. Parallel Combinatorial Optimization. Edited by El-Ghazali Talbi. – Wiley-Interscience; 2006.

[12] Тимошевская Н. Е. Разработка и исследование параллельных комбинаторных алгоритмов. Автореферат диссертации на соискание ученой степени кандидата технических наук. – Томск: ТГУ; 2007.

[13] Торопов Н.Р. Параллельная проверка ДНФ на тавтологию. Информатика. 2005; 2: 35-42.

[14] Cheremisinov D.I., Cheremisinova L.D. Using Parallel Computing in VLSI Computer-Aided Design. Problems of Advanced Micro- and Nanoelectronic Systems Development. Moscow, IPPM RAS; 2017, Part 1: 2–8.

## Авторский вклад

**Черемисинов Дмитрий Иванович** – разработка алгоритмов и программ распараллеливания логикого-комбинаторных задач в среде MPI.

**Черемисинова Людмила Дмитриевна** – постановка задачи исследования и подготовка статьи.

# DEBUGGING PARALLEL PROGRAMS FOR SOLVING LOGICAL-COMBINATORIAL PROBLEMS

***D.I. Cheremisinov***
*UIIP of NAS of Belarus, Leading Researcher,*
*PhD of Technical Sciences, Associate Professor,*

***L.D. Cheremisinova***
*UIIP of NAS of Belarus, Principal Researcher,*
*Doctor of Technical Sciences, Professor,*

**Abstract.** The problems arising in the development and debugging of parallel programs for solving logical-combinatorial problems are considered. The peculiarities of methods for developing and debugging such parallel programs, as well as methods for preventing errors during their development, are discussed. It is shown that algorithms for solving logical-combinatorial problems are difficult to effectively parallelize completely automatically using traditional methods, and the use of known methods for debugging sequential programs does not allow for high-quality verification of such parallel programs.

**Keywords**: parallel program, cluster, MPI, program debugging, verification.

УДК 004.522:2

# BIG DATA FROM MULTI-OMICS TECHNOLOGIES IS ADVANSING MEDICINE AND HEALTHCARE

***O.M. Abbasova***
*Lecturer of the Department of Software of information technologies,*
*The Institute of Telecommunications and Informatics of Turkmenistan*
ogultaganabbasowa@g.mail.com

***O.M. Abbasova***
*She graduated from the Institute of Telecommunications and Informatics of Turkmenistan. The area of scientific of interests is associated with the use and development of artificial intelligence, the organization of educational and research processes at the institute.*

**Annotation.** The global surge in access to digital technologies has led to the collection and documentation of information in various public and private sectors. The healthcare industry has been substantially influenced by the advent of big data. Clinical records, patient lifestyle data recorded from wearable devices, IoTs, electronic health records, imaging analyses, molecular and diagnostic profiling, population studies and clinical trial information together have contributed to major advances and disruption in the healthcare sector.

**Keywords:** Big data, big data in biomedicine, data analytics in biomedical research, multi-omics big data, biomedical data management, big data in personalized medicine

**Introduction.** Big data is defined by massive amounts of data generated in the private and public sectors. In the healthcare industry, sources of big data include clinical records, health records of patients, results of medical examinations, and data collected by using diagnostic and health management devices as part of the internet of things (IoTs). Additionally, the field of biomedical research or biomedicine generates a substantial amount of data, thus contributing to data overload, which is relevant to public healthcare. Big data aims to encourage the use of new technologies to analyze large numbers of available datasets and extract new information on a variety of topics. Big data can take the form of any high volume, complex data representing a condition of interest. Collections of data are often characterized by the five “big Vs” of big data in designing computational systems for big data analytic: volume refers to the size of an overall dataset collection, velocity is the speed at which computational pipelines can load and process the datasets, veracity describes the reliability of datasets in reflecting the true nature of a condition of interest, the variety of data refers to how diverse the datasets sets are with respect to one another in the collection, and variability is a data quality representing continual changes in data and the consequent effects on the performance on a computational system. Together, these qualities contribute to the collective value of big data when applied in any discipline and must be thoroughly addressed in designing big data operational workflows.

Historically, the evolution of big data has markedly changed all aspects of personal and professional life. The collection of information followed by its documentation and analysis dates to 18,000 BCE, when people in tribes used markings, sticks and engravings to track days and food supplies. Moreover, civilizations such as the Egyptians and Romans used libraries to store information that was essential for governing and the military. In the modern era, the birth of big

data has been attributed to some of the first reports of data analysis by John Graunt during the London plague in 1663, after which the innovation and invention of several hardware processing and storage devices paved the way for managing the data generated. Subsequently, the emergence of computational language along with database management systems aided in the global evolution of information technologies and the big-technology sector. Today, technologies that use big data revolve around the gathering, processing, computing, visualizing, and storing of datasets, ideally in the form of an efficient high-speed system. The sheer volume of data available is increasing exponentially each year. Currently, the field of biomedical research is flooded with large experimental datasets obtained from patient tissue, blood, body fluids and laboratory investigations. These data include molecular expression changes obtained from high-throughput "omics"-based studies and clinical parameters from wearable devices. Hence, several large-scale cloud-based platforms have been developed to store, process, curate and manage massive amounts of biomedical data. Some of these platforms include ELIXIR, which is supported and funded by the European Molecular Biology Laboratory, the Global Alzheimer's Association Interactive Network (GAAIN) and Genomics Data Commons funded by the US National Cancer Institute, among others. As such, a growing need exists to develop new technologies that can handle such large-scale datasets while minimizing the costs and demands on computation.

Impact of big data on enterprises and the health care industry. The global surge in access to digital technologies has led to the collection and documentation of information in various public and private sectors. The healthcare industry has been substantially influenced by the advent of big data. Clinical records, patient lifestyle data recorded from wearable devices, IoTs, electronic health records, imaging analyses, molecular and diagnostic profiling, population studies and clinical trial information together have contributed to major advances and disruption in the healthcare sector.

Healthcare and medical research are shifting focus to improving disease outcomes through finding hidden associations or patterns within data derived from a wealth of available data resources. Such data can be used in prediction studies to determine highly probable outcomes of disease. These investigations are leading to advances in delivering personalized medicine tailored to individual patients' needs and the development of automated analysis to aid in the diagnosis and treatment of diseases. Big data has also contributed to better understanding of epidemiology within populations for a variety of pathogens . Immunological surveillance centers routinely process big data to identify pathogens at high risk of becoming endemic within populations. Similarly, big data from genomic libraries, such as whole-genome sequencing and whole-exome sequencing libraries, are playing key roles in accelerating data-driven biomedical discovery. Whole genome sequencing, transcriptome sequencing (RNA-seq and ribosome profiling), proteome profiling (mass spectrometry) and interactome profiling (chromosome conformation capture, ChIP-seq and hybrid assays) are enabling faster, more economical access to the biomedical information of each patient or a specific cohort. Recent advances in genomics, including single-cell genome and transcriptome sequencing of circulating tumor DNA in liquid biopsies, and metagenomics, are already markedly affecting medicine and are steadily being integrated into standard medical practice for early-stage diagnosis and various stages of disease treatment.

The application of big data in genomics is rapidly changing therapeutic development. In the past, drugs were developed according to knowledge of the biological pathways involved in disease, often derived from experiments performed in cells or. Very often, these experimental findings fail to be replicated in humans. Furthermore, drug development is limited to biological processes that are known or believed to be known. The ability to access and analyze the whole genome has allowed scientists and clinicians to identify novel pathways involved in disease, and to develop therapies relevant to humans. This approach supersedes the classical method of drug development, which is conducted according to knowledge of biological pathways involved in disease, based on results obtained from *in vitro* or *in vivo* experiments. However, the major

challenge in the classical method has been the poor rate of success, given the species differences and biological and technical variables. Because genetic variants are often rare, or have small effect sizes, large datasets are required to make valid inferences about the roles of these variants in disease. The collective large amount of genetic information (genomics big data) is in turn guiding the development of therapeutic modalities that are, or could be, individualized to patients according to their genetic profiles.

The sequence of the human genome is the foundation for the understanding of how genetic (DNA and RNA) instructions lead to biological function in humans. In-depth understanding of the genetic landscape of Homo sapiens has been gained from the Human Genome Project. Continued genomic research using this genome map has increased understanding of the functions of genes and the molecular factors that regulate them; variations in DNA and RNA sequences among individuals, and their roles in diseases; and nucleic acid-protein interactions and their regulation in an environment specific manner. Specifically, several international initiatives have been undertaken to evaluate the genetic profiles of patient populations and identify patterns associated with disease risk and therapeutic approaches and responses. For example, The Cancer Genome Atlas (TCGA) is a project initiated by the National Institutes of Health and the National Human Genome Research Institute to sequence specimens from multiple disease sites and gain in-depth understanding of the genetic alterations associated with cancer. Similarly, the genome-wide association study (GWAS) approach allows the whole genomes of individuals to be compared to identify genetic variations called single-nucleotide polymorphisms that are associated with disease. The International HapMap Project was established to determine the patterns in human genome variation and determine their effects on health and disease, including responses to drugs or environmental factors . The Encyclopedia of DNA Elements project was designed to identify the functional and regulatory elements in the human genome sequence, including proteins and noncoding RNA molecules that affect genetic function and outcomes in health and disease. Beyond genomic DNA regions, substantial evidence indicates that RNAs, including messenger RNAs and noncoding RNAs, such as microRNA (miRNA), may also enable effective disease diagnostics in early and late stages of various diseases, including cancer, neurological disorders, and infectious diseases.

With ongoing improvements in the field of bioinformatics and the combination of basic and clinical science with machine learning and high-power computing, major advances have been made in disease diagnosis and the monitoring of molecular changes in organ regions, tissues and circulating body fluids, such as the blood, plasma/serum, urine and cerebrospinal fluid. Collectively, the use of big data in biomedicine involves extracting the essential information from large amounts of research and clinical data to assist in decisions advancing the field toward effective patient-centric healthcare and disease management.

**Data integration across different “omics” branches advances biomedical research.**

Omics technology generates big data that require sophisticated bioinformatic analysis and has enabled the detailed study of many cellular components in their roles and behaviors across many cells and tissues. The availability of high-quality omics data has substantially contributed to understanding across the many tiers of biological conditions in health and disease. Integration of big data across different branches through multi-omics has accelerated outcomes in biomedical research; this integration is relevant for pathogen detection, disease diagnosis and timely therapeutic intervention.

**Analytical pipelines for typical** data integration across different omics branches advanc**e** biomedical research. The analytical procedures, including commonly used analysis packages for various omics data, are summarized in this schematic. Overall, all omics data analysis requires raw data acquisition (gray boxes); quality control (green boxes); identification, quantification, and statistical inference pipelines (yellow boxes); and functional enrichment analysis (red boxes). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Data sources and applications of multi-omics.** Multi-omics data generated from various experimental pipelines (genomics, epigenomics, transcriptomics, proteomics and metabolomics) in pre-clinical and clinical investigations provide useful insights into the flow of biological information at multiple levels (genotype to phenotype), thus aiding in clarifying the mechanisms and factors that contribute to disease pathology.

Integration of data in multi-omics approaches provides a multi-faceted readout by combining results from multiple data layers (e.g., genome, transcriptome, epigenome, and metabolome) of a biological condition. Indeed, several research groups have successfully used multi-omics data to profile biological phenotypes and gain a better understanding of the mechanisms underlying various diseases. Biological data integration can be performed in several ways, depending on the biological question to be answered. Several R software packages have been developed to facilitate easier integration of several types of omics data. One example includes the R package HiCeekR, which enables the integration of Hi-C, RNA-seq and ChIP-seq datasets to explore chromatin configuration changes and their effects on the transcriptome of a host. These packages are usually characterized by each of the different data types (e.g., Hi-C, RNA-seq or ChIP-seq) analyzed in parallel with their results being subjected to a post-integration process. A key downstream analysis is functional annotation of the genes of interest identified by various omics data. The mix Omics R package uses a variety of multivariate methods to account for the increased complexity of integrated datasets, owing to the accumulation of more experimental variables from each of the datasets. Multivariate approaches have been shown to be perform well in the integration of large complex datasets. Currently, increasing omics profiling endeavors have focused on the single-cell level, to uncover the heterogeneity of multi-cellular tissues. Simultaneous profiling of transcriptomes and epigenomes can be performed at the single-cell, or a given spatial location. In this regard, Seurat, a powerful R package with versions (V1 to V3) developed by an interdisciplinary team of investigators and data scientists, can integrate single-cell RNA-seq and single cell ATAC-seq data. More recently, a multimodal reference mapping approach, which can integrate the datasets of single-cell RNA and protein profiles, has been included in Seurat V4.

Other methods use a variety of statistical measurements to infer connections between different datatypes such as Bayesian inference, correlation coefficients and similarity indices. These data integration tools have successfully identified new disease subtypes, such as those in cancer, but are also applicable in the study of other diseases. In viral research, the availability of omics data has contributed to the formation of several sequence databases, such as GISAID, to facilitate the tracking of viral evolution and the emergence of new influenza and SARS-CoV-2 strains. Omics big data has also influenced the understanding of how viral infections develop over. Multi-omics approaches have been successful in characterizing changes in cell-type populations during SARS-CoV-2 infection. Because most diseases are characterized beyond the findings from a single omics data type, combining omics datatypes helps to provide more comprehensive details regarding the development and progression of diseases and related pathological features.

The key downstream analysis involves annotation of the functions of the genes of interest identified by various omics data. The popularity and performance of several gene set analyses have been summarized in a recent review. Of note, one of the most used and powerful algorithms is GSEA. DAVID, a pathway enrichment tool, remains very popular, although it appears to be out of date, given that the most updated version (DAVID 6.8) was released 5 years ago. In contrast, the R package clusterProfiler (version 4.0), which has recently been substantially updated, is an outstanding enrichment tool in this regard. In addition to gene set enrichment analysis and the comparison of enrichment results from multiple gene lists, clusterProfiler provides interfaces for data operation and visualization.

**Conclusion.** In conclusion, big data has revolutionized the healthcare industry, significantly impacting various sectors from clinical practice to biomedical research. With its vast sources and diverse types such as a genomic data, clinical records, and we are able device outputs, big data

allows for a deeper understanding all of disease mechanisms and the development of personalized medicine. The integration of multiple “omics” data types has facilitated comprehensive insides into disease pathology, aiding in more accurate diagnostics, treatment options, and epidemiological surveillance. As computational technologies continue to evolve, the efficient management and analysis of large-scale data are poised to further enhance healthcare outcomes and accelerate the development of tailored therapeutic approaches. Ultimately, the ongoing integration of big data into biomedical research promises to transform patient care, enabling more precise, individualized, and effective healthcare solutions.

## List of references

[1] PC Magazine Encyclopedia. http://www.pcmag.com/encyclopedia/term/62849/big-data.

[2] O’Driscoll A, Daugelaite J, Sleator RD. ‘Big data’, Hadoop and cloud computing in genomics. J Biomed Inform 2013

# БОЛЬШИЕ ДАННЫЕ В MULTI – OMICS ТЕХНОЛОГИЯ – ЭТО ПРЕИМУЩЕСТВО В МЕДИЦИНЕ И ЗДРАВООХРАНЕНИЯ

***О. Аббасова***

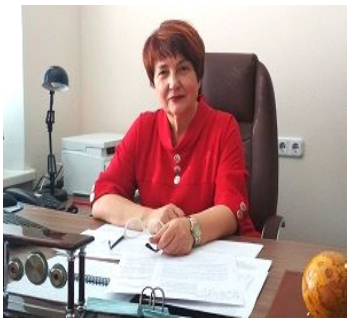
*Преподаватель кафедры «Программное обеспечение информационных технологий»*
*Институт Телекоммуникаций и информатики Туркменистана*

**Аннотация.**. Глобальный всплеск доступа к цифровым технологиям привел к сбору и документированию информации в различных государственных и частных секторах. На индустрию здравоохранения существенно повлияло появление больших данных. Клинические записи, данные об образе жизни пациентов, записанные с носимых устройств, IoTs, электронных медицинских карт, анализов изображений, молекулярного и диагностического профилирования. Популяционные исследования и информация о клинических испытаниях вместе способствовали крупным достижениям и прорывам в секторе здравоохранения.

**Ключевые слова:** Большие данные, большие данные в биомедицине, аналитика данных в биомедицинских исследованиях, мулти – омикс большие данные, управление биомедицинскими данными, большие данные в персонализированной медицине.

УДК 004.8-027.231(476)

# ВОЗМОЖНЫЕ ПЕРСПЕКТИВЫ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РЕСПУБЛИКЕ БЕЛАРУСЬ



***Т.В. Казак***
*Заведующий кафедрой инженерной психологии и эргономики, доктор психологических наук Республики Беларусь, доктор психологических наук Российской Федерации, профессор, БГУИР*
kazak@bsuir.by

***А.Н. Василькова***
*Старший преподаватель кафедры инженерной психологии и эргономики, магистр,*
*БГУИР*
a.vasilkova@bsuir.by

***Т.В. Казак***

*Заведующий кафедрой инженерной психологии и эргономики, доктор психологических наук Республики Беларусь, доктор психологических наук Российской Федерации, член-корреспондент Международной академии психологических наук, профессор.*

*Образование: Высшее; Аспирантура, специальность: 19.00.03 - психология труда, инженерная психология, эргономика; Докторантура, специальность: 19.00.03 - психология труда, инженерная психология, эргономика.*

*Область профессиональных интересов / исследований: Психология труда. Инженерная психология. Эргономика. Психология управления. Юридическая психология. Социальная и организационная психология. Клиническая психология.*

***А.Н. Василькова***

*Старший преподаватель кафедры инженерной психологии и эргономики.*

*Образование: 2007 - МГВРК по специальности «Программное обеспечение информационных технологий»,*
*2022 - магистратура БГУИР по специальности «Охрана труда и эргономика».*

*Область профессиональных интересов / исследований: языки программирования, искусственный интеллект, технологии виртуальной реальности, большие данные.*

**Аннотация.** Стремясь оценить, как динамика GenAI может развиваться к 2030 году, необходимо учитывать, как факторы, способствующие развитию, которые играют каталитическую роль, так и факторы, препятствующие развитию GenAI. Основные ускоряющие факторы включают достижения в новых технологиях, которые расширяют технические возможности GenAI, цифровую грамотность, крупные инвестиции (как государственные, так и прямые частные инвестиции, а также государственно-частные партнерства) со стороны технологического сектора, и волю государственной политики, экономики как макромасштаба, предприятий (коммерческих или социальных) и общества принять и использовать GenAI. Основные ингибиторы включают отсутствие стандартов взаимодействия и конкурентоспособности, отсутствие цифровой грамотности, стимулов и долгосрочных планов, оборонительное или технофобное мышление, а также усталость от использования технологий. Хотя все информанты предпочитают открытую, инклюзивную и демократическую разновидность GenAI, значительная доля участников считает и предвидит, что технологическая сфера будет закрытой и централизованной, в которой будут доминировать несколько корпоративных субъектов, а также множество цифровых неравенств и алгоритмических несправедливостей.

Вопросы цифровой этики и регулирования вызывают большую озабоченность у экспертов. С одной стороны, чрезмерное политическое или корпоративное вмешательство и контроль могут ограничить уровень функциональности и инноваций GenAI. С другой стороны, низкий уровень регулирования может заставить пользователей чувствовать себя неуверенно в надежности контента (например, фейковых новостей и дипфейков), их конфиденциальности, что приведет к менее позитивной и здоровой культуре использования GenAI. К 2030 году целенаправленные действия и проактивная политика вокруг GenAI будут важны и необходимы как для частных компаний, так и для государственного сектора.

Глобальный геополитический и геоэкономический порядок становится все более неопределенным, сложным и нестабильным. GenAI, по-видимому, усиливает эти системные черты, подчеркивая необходимость «экспоненциального» и ориентированного на будущее образа мышления. Мы не можем быть уверены в том, что принесет нам будущее, и не можем избежать постоянных сбоев в эпоху пермакризиса. Но нам предстоит использовать эту обобщенную неопределенность/сложность, ставить долгосрочные цели, адаптировать институты и образ мышления, быть должным образом подготовленными и работать над наиболее благоприятным сценарием - т. е. сценарием «техносоциального ускорения» - чтобы повысить способность страны к устойчивому росту и устойчивому процветанию на основе динамической диагностики тенденций, неопределенностей и возможностей, возникающих вокруг нас.

**Ключевые слова:** искусственный интеллект, GenAI, перспективы генеративного искусственного интеллекта, блокчейн.

**Введение.** Последние разработки в области искусственного интеллекта (ИИ), а точнее в области генеративного искусственного интеллекта (Генеративного ИИ), открывают волну технологических инноваций, способных изменить структуру общества, экономики и работы. Поскольку эта преобразующая технология продолжает развиваться быстрыми темпами, Республика Беларусь, похоже, находится на заре новой эры, где Генеративный ИИ готов сыграть центральную роль в ускорении автоматизации, повышении производительности и революционизации различных отраслей.

Системы генеративного ИИ эволюционировали — и продолжают эволюционировать — экспоненциально в новую форму техно-социального субъекта, демонстрируя беспрецедентную способность составлять и производить контент, который напоминает человеческие результаты, выполнять команды (в вероятностных терминах) и творить без размышлений, чувств, понимания, объяснений или вынесения моральных суждений, давать «почему» тому, что происходит. Генеративный ИИ демонстрирует постоянно расширяющийся диапазон возможностей, от создания в первую очередь реальных текстов до создания изображений, видео, музыки и даже компьютерного кода.

Влияние непрерывного развития генеративного ИИ огромно и затрагивает множество различных секторов деятельности, таких как здравоохранение, финансы, развлечения, СМИ, образование, государственное управление, защита окружающей среды, судоходство, туризм, культура и т. д.

Эта технология представляет собой крупный скачок вперед в сотрудничестве и конвергенции человека и машины, сравнимый по значимости с появлением персональных компьютеров, интернета и мобильных технологий. Однако, как и любая революционная технология, генеративный ИИ перестраивает весь наш горизонт восприятия, представляя свой собственный набор проблем, включая потенциальную возможность для его систем генерировать неточную или ложную информацию, поднимая вопросы о технической и эпистатической надежности, а также этике.

Поскольку все больше организаций в Республике Беларусь и по всему миру интегрируют генеративный ИИ в свою повседневную деятельность, мы можем представлять и предвидеть глубокие технологические, социальные, политические и экологические изменения. Высокая степень неопределенности и сложности, сопровождающая эти разработки, требует внедрения стратегии проактивного подхода, а именно процессов участия, которые позволяют систематически понимать (но не прогнозировать) окружающую среду и силы, ее формирующие, альтернативные перспективы будущего и сценарии для формирования политики, и стратегии.

«Будущее» генеративного ИИ в Республике Беларусь можно представить, как наполненное обещаниями, но также и множеством рисков и потенциальных ловушек. Оно

открывает просторы для дальновидного мышления и исследовательской мысли и предоставляет философию и инструменты, необходимые для эффективной навигации и использования генеративного ИИ нашей страной. Дискуссия о технологиях и о том, как они интегрируют и изменяют нашу повседневную жизнь, усилилась в последние годы. Драматический прогресс, наблюдаемый в области искусственного интеллекта (ИИ) за последнее десятилетие, и его первые практические применения в исследованиях, промышленности и различных аспектах нашей жизни стали катализатором. Эти разработки приносят такие значительные изменения в то, как мы производим, потребляем, работаем, учимся, остаемся в курсе, взаимодействуем и удовлетворяем свои потребности, что сейчас широко распространено мнение, что человечество находится в середине четвертой промышленной революции, движущими силами которой являются ИИ и большие данные.[1]

**Основная часть.** ИИ меняет «ткань» человеческих знаний и находится в постоянной эволюции, предлагая новые возможности и открывая новые горизонты во многих секторах. Тенденции отражают спрос на более высокое качество и разнообразие в создании контента, а также желание расширить возможности ИИ в различных областях.

Согласно недавнему исследованию McKinsey22, как подчеркнул Атанасиос Чимис, исследователь Центра планирования и экономических исследований (KEPE): «влияние ИИ на производительность во всем мире может добавить от 2,6 до 4,4 триллиона долларов в год.

Более того, ИИ увеличит общее влияние искусственного интеллекта на 15% до 40%. В том же исследовании подсчитано, что примерно 75% ценности, которую может предложить ИИ, приходится на четыре сектора: операции с клиентами, маркетинг и продажи, проектирование программного обеспечения, а также исследования и разработки. Это означает, согласно тому же исследованию, что отраслями с наибольшим влиянием ИИ будут банковское дело, высокие технологии и биологические науки. Причина в том, что ИИ по определению способен поддерживать взаимодействие с клиентами, создавать креативный контент для маркетинга и продаж и составлять компьютерный код на основе естественного языка». [2]

Алгоритмы развиваются и включают новые методы, предоставляя новые возможности для приложений, которые требуют высокой креативности и естественности в создаваемом контенте, таком как текст, музыка, видеопроизводство и другие формы мультимедийного контента.

В этой статье описываются технологические, экологические, социальные и экономические тенденции ИИ. Эти тенденции представляют собой направления изменений в ИИ с течением времени, а также основы для определения возможностей в белорусском контексте, а также вызовы, с которыми мы призваны столкнуться. Последнее, вместе с позитивным сценарием, которому Республика Беларусь может следовать и которого должна придерживаться, формируют оси политики, которые необходимо учитывать.

***Технологические аспекты.*** *Сотрудничество человека и искусственного интеллекта.* По мере развития возможностей ИИ возникает важный вопрос: каковы будут последствия для человеческой жизни, работы и взаимодействия с ИИ? ИИ все чаще вовлекается в деятельность человека, и можно отметить, что более половины (53%) организаций уже начали внедрять роботизированную автоматизацию процессов. Ожидается, что эта тенденция будет дальше расширяться, при этом ИИ потенциально автоматизирует до 65% задач и значительно сократит расходы. Развивающиеся отношения между людьми и ИИ подчеркивают необходимость тщательно изучить меняющийся ландшафт работы и повседневной жизни в этом все более усовершенствованном ИИ мире.

*Мультимодальный ИИ.* Согласно данным об использовании фундаментальных моделей ИИ, или «базовых моделей», лидерство занимают большие языковые модели, 71%

компаний ИИ используют и разрабатывают их, за ними следуют мультимодальные модели с 38%.[3]

Мультимодальная модель — это тип ИИ, способный принимать и генерировать информацию из нескольких источников, таких как текст, изображения, звук, видео и и т. д. Тенденция к переходу к мультимодальным моделям возникает из убежденности, что эти модели достигнут более высокой точности по сравнению с традиционными моделями, тем самым расширяя их применение на разнообразный набор приложений. Кроме того, мультимодальные модели не имеют исключительной зависимости от текстовых данных и ограничений, которые с этим связаны, таких как исключительное использование для генерации текста, отсутствие извлечения знаний из контекстных мультимедиа и ограниченная способность к обобщению.

Помощь в написании кода GPT предоставляет разработчикам ценный инструментарий для повышения их производительности. Сюда входят такие возможности, как преобразование кода из одного языка программирования в другой, автоматизация задач написания и объяснение кода.

В области производства кода появились заметные исследования и инструменты для дальнейшей помощи разработчикам. В частности, такие инструменты, как Codegen29 и CodeGeeX30, представляют собой значительный прогресс в этой области. Кроме того, широко признанные инструменты, такие как Co-Pilot31, ChatGPT32 и Llama33, занимают видное место в сообществе разработчиков за свою способность помогать в различных действиях, связанных с написанием и исправлением кода.

*Документированное резюме решений.* Процесс обобщения для принятия решений на основе фактических данных включает в себя сжатие большого объема информации в обобщенный формат. Эти резюме предоставляют лицам, принимающим решения, основные знания, выводы и факты, необходимые для принятия достаточно обоснованных решений на основе имеющихся данных. Этот подход находит особый резонанс в таких областях, как здравоохранение, бизнес и академическое сообщество, где анализ данных играет решающую роль в формировании выбора и стратегий.

Примером, который служит вышеуказанной цели, является PolicyGPT34, который предоставляет краткие сведения об информации для политик конфиденциальности. [3]

Обнаружение контента, сгенерированного ИИ. Обнаружение контента, сгенерированного ИИ, относится к способности систем ИИ точно идентифицировать и различать контент, который был сгенерирован или обработан другими алгоритмами. Эта тенденция особенно важна в усилиях по борьбе с распространением дезинформации, «дипфейков» и других форм контента, созданного ИИ, которые могут обманывать, дезинформировать или манипулировать людьми.

Кроме того, она важна для поддержания и восстановления доверия и аутентичности в цифровых медиа.

Расцвет служб Web3, использующих продуктивный искусственный интеллект. Расцвет служб Web3 (Spatial Web)36, использующих ИИ, представляет собой значительную эволюцию в цифровом мире. Используя технологию блокчейн, Web3 стремится создать Интернет, в котором пользователи контролируют свои данные и транзакции. Кроме того, ИИ может улучшить пользовательский опыт на платформах Web3, предоставляя персонализированный контент и взаимодействия. Наконец, ИИ фундаментально трансформирует сферу маркетинга Web3, увеличивая возможности маркетинговых стратегий. Недавнее исследование пытается пролить свет на эту тенденцию, предоставляя полезные сведения о том, как инструменты ИИ используются в торговле для повышения производительности и ускорения маркетингового цикла.

*Дальнейшее развитие обработки естественного языка.* Прогресс новых методологий в обработке естественного языка (NLP) продолжается в исследовательском пространстве, изучая различные аспекты LLM, такие как способность к обобщению, способность к

рассуждению, понимание решений ИИ, сокращение затрат и т. д. В то же время одним из наиболее полезных и широко признанных вариантов использования LLM является перевод текста. Обширные исследования показали, что перевод с использованием LLM часто превосходит популярное программное обеспечение для перевода, такое как Google Translate. Это подчеркивает замечательный потенциал LLM в преодолении языковых барьеров и содействии эффективному межъязыковому общению.

*Автономные системы ИИ.* Это знаменует собой заметный скачок в области продуктивных приложений ИИ. Эти системы работают независимо, непрерывно генерируя контент в ответ на запросы без необходимости постоянного взаимодействия с пользователем. В отличие от традиционных приложений чат-ботов, которые полагаются на инициированные пользователем вопросы и ответы, автономные приложения ИИ выполняют более сложные и продвинутые функции, что делает их очень гибкими.

Выдающимся примером является Auto-GPT39, автономная система ИИ, способная генерировать контент и выполнять широкий спектр задач, не требуя постоянного руководства со стороны пользователей.

Автономная природа этих моделей открывает путь для нового поколения приложений ИИ, способных обрабатывать все более сложные и динамичные сценарии с минимальным человеческим вмешательством.

***Экономические аспекты.*** Согласно исследованию, проведенному в 2019 году Accenture Greece совместно с Microsoft в странах на юго-востоке Европы, предполагается, что ИИ может добавить 195 миллиардов долларов к ВВП страны за 15-летний период, в то время как прогнозируется, что к 2030 году рынок продуктов ИИ составит 500 миллиардов долларов.

*Инвестиции в исследования и разработки для ИИ.* Как показывают финансовые данные крупных корпораций, значительное внимание уделяется исследованиям и разработкам для приложений ИИ. В частности, и Alphabet, и Meta значительно увеличили свои расходы на исследования и разработки в последние годы. [2]

Эти цифры указывают на приверженность инновациям и технологическому прогрессу в быстро развивающемся секторе, и никто не хочет оставаться позади.

*Демократизация ИИ.* В настоящее время существует огромный разрыв между технологическими гигантами и теми, у кого нет финансовых средств и вычислительных ресурсов, необходимых для обучения модели ИИ. Эту проблему также подтверждает представитель многонациональной технологической компании, подчеркивающий необходимость «демократизации цифровых технологий и доступа к ним для гораздо большей части населения».

Преобладающая тенденция направлена на содействие демократизации ИИ посредством распределенного обучения, сообществ с открытым исходным кодом и создания центров обработки данных, доступных для всех граждан.

*Финансовые аспекты.* ИИ может улучшить обнаружение и предотвращение мошенничества, анализируя огромные объемы данных транзакций для выявления подозрительных схем и потенциальных угроз в реальном времени. Кроме того, он может оптимизировать инвестиционные стратегии, обрабатывая рыночные данные, новости и исторические тенденции для предоставления информации и прогностических моделей на основе данных, помогая финансовым учреждениям и инвесторам принимать обоснованные решения. В сфере обслуживания клиентов чат-боты и виртуальные помощники могут предоставлять персонализированные финансовые консультации, отвечать на запросы и помогать в управлении счетами, улучшая общий клиентский опыт.

Ниже приведены некоторые примеры применения ИИ в финансовом секторе:

1 Capital One и JPMorgan Chase44 - использовали ИИ для улучшения своих систем обнаружения мошенничества и мониторинга подозрительной активности с помощью ИИ. Как сообщается, эти усилия привели к значительному сокращению ложных срабатываний,

улучшению коэффициентов обнаружения, снижению затрат и повышению удовлетворенности клиентов.

2 Morgan Stanley Wealth Management будет использовать технологию OpenAI для использования своих обширных источников данных для поддержки финансовых консультантов с информацией о компаниях, секторах, классах активов, рынках капитала и регионах по всему миру.

3 Wells Fargo разрабатывает возможности автоматизации обработки документов, экономические аспекты, включая предоставление сводных отчетов и масштабирование виртуальных помощников чат-ботов.

4 Goldman Sachs и Citadel - изучают приложения ИИ для внутренней разработки программного обеспечения и анализа информации.

***Экологические аспекты.*** Он фокусируется на устойчивости центров обработки данных в рамках Пакта о климатически нейтральных центрах обработки данных, направленного на установление стандартов для их устойчивой работы. Среди ключевых целей пакта – повышение энергоэффективности, переход на зеленые источники энергии и содействие повторному использованию и переработке оборудования.

Мониторинг окружающей среды. Он вносит значительный вклад в такие виды деятельности, как мониторинг выбросов метана, мониторинг качества воздуха, измерение экологического следа и сокращение выбросов. В контексте экологической устойчивости ИИ служит ценным инструментом для сбора и обработки данных из различных источников, что позволяет предоставлять в режиме реального времени информацию и прогнозы относительно экологических условий.

Появление «зеленого» ИИ По мере того, как модели ИИ продолжают развиваться, растущая сложность и требование вычислительной мощности вызывают экологические проблемы, касающиеся значительно возросшего углеродного следа.

В ответ на это научно-исследовательское сообщество уделяет значительное внимание разработке энергоэффективных алгоритмов и архитектур.

*Оптимизация ресурсов.* В свете инициатив по устойчивому развитию существует острая необходимость в сокращении отходов, снижении затрат и повышении эффективности, при этом ИИ находится на переднем крае, как ключевой союзник в достижении этих целей. Это достигается путем анализа обширных наборов данных и создания практической информации для оптимизации управления ресурсами.

***Социальные аспекты.*** *Продуктивный ИИ в искусстве и культуре.* Эта новая тенденция проистекает из эволюции систем ИИ для создания чрезвычайно сложных и оригинальных форм искусства, музыки и литературы. Благодаря обширному обучению на обширных наборах данных, содержащих различные произведения искусства и культурные объекты, ИИ продвигает стандарты и стили, охватывающие разные жанры. Следовательно, он может создавать совершенно новый или новый контент, который часто очень похож на творения, созданное людьми. Эта эволюция привела к появлению нового ландшафта цифрового искусства, которое не только раздвигает границы творчества, но и нарушает традиционные художественные процессы.

Похоже, что ИИ меняет широкую область культурных и творческих отраслей, переопределяя вековую заботу человечества о природе и направлении его творчества, а также его помощь посредством технологий. В этом контексте недавно были сформулированы четыре сценария будущего ИИ с акцентом на ИИ в этих типах отраслей (искусство и культура):

1 ИИ может стать помощником и усилить человеческую креативность.

2 Мы можем быть вынуждены медленно внедрять ИИ из-за недостатка соответствующих навыков у рабочей силы, а также этических и юридических сложностей, связанных с правами интеллектуальной собственности создателей.

3 Сверхразумный ИИ почти полностью заменяет человеческий талант в творческих отраслях, поскольку он может быть более рентабельным в среднесрочной перспективе, а человеческие творческие усилия часто оказываются непродуктивными и не дают измеримых результатов. [2]

4 Ограниченное влияние ИИ на культурное производство, поскольку ИИ может предлагать низкокреативные - просто имитационные - результаты (узкий GenAI).

В контексте этого обсуждения возникающего воздействия ИИ на сферы культурных/творческих индустрий, даже правовая база прав интеллектуальной собственности сталкивается с новой моделью: «материал, созданный искусственным интеллектом, не может использоваться для замены автора-человека.

В частности, материал, созданный искусственным интеллектом, не может быть классифицирован как исходный материал для адаптации никаким образом, и если студия хочет использовать сценарий, созданный ИИ, не может быть указан автор, и, следовательно, нет права интеллектуальной собственности.

*Этические проблемы.* С непрерывной эволюцией и расширением ИИ, кажется, что обсуждения и проблемы, касающиеся этики, значительно возрастают. Важно отметить, что многие из этих обсуждений превосходят старые «дуалистические» или «манихейские» тенденции, признавая двойственность ИИ: одна и та же технология влечет за собой как хорошее, так и плохое использование. Однако, здесь мы должны ввести аналитическое различие, согласно которому структура руководящих принципов делится на:

а) процедурные принципы, которые управляют функциональностью систем, такие как подотчетность, прозрачность, объяснимость, совместимость, управление предвзятостью и научная обоснованность как ценность;

б) этические принципы, которые перспективно сосуществуют и гармонизируются, такие как автономия и уважение человеческого достоинства, защита частной жизни и прав от несправедливого стороннего управления, принцип непричинения вреда и пользы, избежание не только цифрового разрыва, но и более широких социальных последствий алгоритмической предвзятости и создание автоматизированных неравенств в распределении выгод, солидарность и чувство совместной принадлежности, справедливости, ответственности в направлении создания защитных замков для человеческого контроля автоматизации в алгоритмическом принятии решений, а также интеграция этических стандартов уже из процесса проектирования создаваемых технологических систем.

Существенная текущая проблема вращается вокруг потенциального злоупотребления контентом, созданным ИИ, особенно при создании поддельных видео, которые могут распространять ложную информацию или служить злонамеренным целям. Создание фильтрующих механизмов имеет жизненно важное значение, позволяя системам ИИ автономно адаптироваться для обеспечения соблюдения этических правил. Кроме того, концепция надежного использования ИИ выходит за рамки этики, охватывая философские, юридические и технологические аспекты.

*Предрассудки и дискриминация.* Некоторые из основных проблем в системах ИИ являются предрассудками, стереотипными обобщениями и ошибочными суждениями. Эти проблемы возникают из-за предвзятости, присутствующей в данных и связанной с человеческим поведением. Однако, усилия по устранению всех источников предвзятости из данных становятся все более интенсивными с целью создания беспристрастных моделей. Кроме того, после обучения базовых моделей обычной практикой является тонкая настройка моделей экспертами, чтобы избежать нежелательного поведения.

*Конфиденциальность.* Обеспечение защиты конфиденциальности имеет первостепенное значение. Согласно недавним исследованиям, существует риск нарушения конфиденциальности и прав собственности третьих лиц, а именно конечных пользователей услуги. Существует проблема нарушения данных пользователя, использующего цифровую

услугу без его согласия (например, компания, создающая цифровую игру, нацеленную на конечных пользователей с основной косвенной целью сбора их поведенческих данных и продажи их другой компании без согласия пользователей игры).

В целом, кажется, что ИИ усиливает эти опасения, смягчение которых связано с использованием множественных стратегий фильтрации и анонимизации.

**Заключение.** Глобальные тенденции в области ИИ определяют среду, в которой появляются новые возможности на национальном уровне для инноваций, развития, а также для модернизации и модернизации секторов экономики. Эти возможности, подчеркнутые прогрессивным подходом, могут внести значительный вклад в совместную эволюцию технико-социо-экономической структуры страны и сформировать новые столпы долгосрочного и устойчивого развития. В этом понимании секторы и специализации возможностей упоминаются через европейскую точку зрения, принимая во внимание местные особенности, потребности и возможности, путем анализа ответов специалистов, представляющих различные заинтересованные стороны.

ИИ может создавать новый, оригинальный контент, такой как текст, изображения или музыка, и это сильная будущая тенденция. Сочетание этих двух технологий открывает захватывающие возможности, такие как создание уникальных цифровых активов, которые могут принадлежать и продаваться на блокчейне. Как заявил представитель многонациональной технологической компании в контексте нашего исследования, «ИИ имеет потенциал для поддержки более высокого качества, большего объема контента и более высокой производительности труда работников». Кроме того, согласно недавнему исследованию, проведенному Deloitte Digital67 в США среди 640 маркетинговых компаний, ИИ уже меняет форму производства контента для маркетинга: 26% респондентов уже используют ИИ, а еще 45% планируют использовать его к концу 2024 года. Это же исследование показывает 54%-ное увеличение объема контента, произведенного компаниями в предыдущем году, при этом 65% из них обеспокоены интеллектуальной собственностью или юридическими рисками, связанными с использованием ИИ. Однако пользователи ИИ сообщают об экономии времени примерно в 11,4 часа в неделю, что позволяет им посвятить себя более ценным или более стратегически значимым задачам. Тем не менее, по словам нашего вышеупомянутого собеседника: «Инструменты ИИ требуют человеческого вмешательства для непосредственного производства и оценки результатов и, следовательно, являются инструментами, которые помогают людям в их повседневной жизни. В то время как контент, созданный с использованием ИИ в будущем, вероятно, продолжит управляться людьми».

Например, текст и изображения, опубликованные на различных веб-сайтах (новости, записи в блогах и т. д.), вероятно, будут в значительной степени определяться GenAI гораздо раньше 2030 года. Другие категории цифрового контента, такие как полнометражные фильмы, потребуют больше времени, хотя даже в этих случаях контент будет включать в себя обширные элементы GenAI.

## Список литературы

[1] Алейникова, Д. В. К проблеме общения человека с искусственным интеллектом / Д. В. Алейникова // Вестник Московского государственного лингвистического университета. Образование и педагогические науки. – 2023. – Вып. 3 (848). – С. 9-15.

[2] Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. – Москва : Юрайт, 2024. – 243 с.

[3] Беликова, Е. К. Проблема отнесения искусственного интеллекта к числу современных феноменов культуры / Е. К. Беликова // Общество: философия, история, культура. – 2024. – № 2(118). – С. 60-65.

## Авторский вклад

**Казак Тамара Владимировна** – руководство и постановка задачи исследования аналитического подхода на основе больших данных для прогнозирования факторов успеха студентов.

**Василькова Анастасия Николаевна** – постановка задачи исследования, описание принципа работы Big Data в улучшении эффективности прогнозирования факторов успеха студентов, анализ полученных результатов, формирование структуры статьи.

# POSSIBLE PROSPECTS OF GENERATIVE ARTIFICIAL INTELLIGENCE IN THE REPUBLIC OF BELARUS

***T.V. Kazak***
*Head of the Department of Engineering Psychology and Ergonomics, doctor of psychological sciences of the Republic of Belarus, doctor of psychological sciences of the Russian Federation, Professor, Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus.*

***A.N. Vasilkova***
*Senior Lecturer, Department of Engineering Psychology and Ergonomics, BSUIR*

**Abstract.** In seeking to assess how the GenAI dynamics might evolve by 2030, it is necessary to consider both the enablers that play a catalytic role and the inhibitors of GenAI. The key accelerators include advances in new technologies that enhance GenAI's technical capabilities, digital literacy, large investments (both public and private equity, as well as public-private partnerships) from the tech sector, and the will of public policy, the macro-scale economy, businesses (commercial or social), and society to embrace and use GenAI. The key inhibitors include the lack of interoperability and competitiveness standards, lack of digital literacy, incentives, and long-term plans, defensive or technophobic thinking, and technology fatigue. Although all informants prefer an open, inclusive, and democratic version of GenAI, a significant proportion of participants believe and foresee a closed and centralized tech landscape dominated by a few corporate actors and with multiple digital inequalities and algorithmic injustices. Digital ethics and regulation are of great concern to experts. On the one hand, excessive political or corporate interference and control may limit the level of functionality and innovation of GenAI. On the other hand, low levels of regulation may make users feel insecure about the reliability of content (e.g. fake news and deepfakes), their privacy, leading to a less positive and healthy culture of GenAI use. By 2030, targeted actions and proactive policies around GenAI will be important and necessary for both private companies and the public sector. The global geopolitical and geoeconomic order is becoming increasingly uncertain, complex and unstable. GenAI seems to amplify these systemic features, highlighting the need for an "exponential" and future-oriented mindset. We cannot be sure of what the future will bring us, and we cannot avoid constant disruptions in the era of permacrisis. But we have to use this generalized uncertainty/complexity, set long-term goals, adapt institutions and mindsets, be properly prepared, and work towards a best-case scenario – i.e. a "technosocial acceleration" scenario – to enhance the country's capacity for sustainable growth and sustainable prosperity based on a dynamic diagnosis of the trends, uncertainties, and opportunities emerging around us.

**Keywords**: artificial intelligence, GenAI, prospects of generative artificial intelligence, blockchain.

УДК 004.522:2

# РАСКРЫТИЕ ПОТЕНЦИАЛА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ АНАЛИТИКИ БОЛЬШИХ ДАННЫХ

***М.Б. Таджиева***
*Преподаватель кафедры «Программное обеспечение информационных технологий»*
*Институт Телекоммуникаций и информатики Туркменистана*
gurbangylyjovam@gmail.com

***М.Б. Таджиева***
*Окончила университет инженерных технологий Туркменистана имена Огуз хана. Область научных интересов связана с использованием и разработкой искусственного интеллекта, организацией учебного и научно-исследовательского процессов в институте.*

**Аннотация.** В данной статье всесторонне рассматривается использование искусственного интеллекта (ИИ) в аналитике больших данных. Основное внимание уделяется методам машинного обучения и глубокого обучения, которые используются для разработки инновационных алгоритмов и решений в таких областях, как финансы, здравоохранение, окружающая среда и образование. В статье рассматриваются преимущества применения ИИ для анализа больших данных, такие как повышение эффективности и точности прогнозов, а также оптимизация решений. Однако в нем также подчеркиваются недостатки и проблемы, такие как обработка и безопасность информации, вопросы конфиденциальности и этические соображения. Возможности и технологические проблемы, связанные с обработкой огромных объемов данных, весьма многочисленны. В целом статья представляет собой сбалансированный взгляд на современный ИИ в условиях революции больших данных. Он информирует технических и нетехнических читателей о том, как добиться успеха на стыке больших данных и ИИ, реалистично оценивая проблемы, следуя принципам ответственного ИИ и уделяя особое внимание проектированию, ориентированному на человека.

**Ключевые слова:** Искусственного интеллекта (ИИ), большие данные, обучение, машинное обучение (МО).

**Введение.** Область анализа больших данных с помощью искусственного интеллекта охватывает различные методы, в основном относящиеся к областям машинного обучения и глубокого обучения. Эти методы играют ключевую роль в современных приложениях ИИ для анализа огромных наборов данных, предлагая больше, чем просто инструменты — они представляют собой основу анализа данных в эпоху ИИ. Машинное обучение — важная область искусственного интеллекта — использует статистические методы, позволяющие компьютерам выполнять определенные задачи на основе предоставленных данных. Этот подход автоматизирует построение аналитических моделей, основанных на принципе, согласно которому системы могут обучаться на основе данных, выявлять закономерности и принимать решения с минимальным вмешательством человека. Область машинного обучения можно условно разделить на три типа: контролируемое обучение, неконтролируемое обучение и обучение с подкреплением.

При контролируемом обучении алгоритмы обучаются на маркированных данных, которые включают в себя входные данные и соответствующие им правильные выходные данные. Алгоритм учится сопоставлять эти входные данные с выходными данными, что широко используется в приложениях, где исторические данные предсказывают вероятные будущие события, например, при обнаружении мошенничества в банковских транзакциях. С другой стороны, неконтролируемое обучение имеет дело с немаркированными данными.

Система пытается изучить структуру и закономерности таких данных, что обычно используется в исследовательском анализе данных для поиска скрытых закономерностей или группировок в данных, например, при сегментации клиентов. Обучение с подкреплением подразумевает, что агент учится вести себя в окружающей среде, выполняя действия и наблюдая за результатами. Этот метод в основном применяется в сценариях, где принятие решений является последовательным, а цель — долгосрочной, например, в стратегиях робототехники.

Глубокое обучение — специализированный подвид машинного обучения — основано на искусственных нейронных сетях с несколькими слоями, что позволяет модели обучаться и принимать разумные решения самостоятельно. Он произвел революцию в области искусственного интеллекта благодаря своей способности обрабатывать и анализировать сложные и неструктурированные данные, такие как изображения, звук и текст.

Сверточные нейронные сети в основном используются для обработки изображений, видео, аудио и других двумерных данных, прекрасно справляясь с такими задачами, как распознавание изображений и видео, классификация изображений и анализ медицинских изображений. Рекуррентные нейронные сети, используемые для последовательных данных, таких как анализ временных рядов, языковое моделирование и перевод текста, отличаются от традиционных нейронных сетей тем, что имеют память для хранения предыдущих входных данных во внутренней памяти, что делает их идеальными для последовательных данных. Трансформеры, изначально разработанные для задач обработки естественного языка, таких как перевод и реферирование текстов, произвели революцию в машинном обучении. Они известны своей эффективностью в распараллеливании обработки данных и обработке долгосрочных зависимостей в данных.

Интеграция машинного обучения и глубокого обучения в анализ больших данных открыла новые горизонты в исследованиях и практическом применении искусственного интеллекта. Эти методы меняют наш подход к большим наборам данных и их понимание — от предиктивной аналитики в финансах до расширенного распознавания изображений в здравоохранении. По мере развития технологий увеличивается потенциал более сложного и эффективного анализа больших данных на основе искусственного интеллекта, что обещает дальнейшие инновации и прорывы в различных областях.

***Применение искусственного интеллекта для анализа больших данных в различных отраслях.***

Использование искусственного интеллекта для анализа больших данных получило широкое распространение во многих отраслях. У каждого сектора есть свои уникальные проблемы и требования, и ИИ предлагает индивидуальные решения для эффективного удовлетворения этих потребностей. В финансовом секторе ИИ играет решающую роль в анализе больших объемов данных для различных целей. Он играет важную роль в прогнозировании цен на акции, оценке кредитных рисков и выявлении мошеннических действий. Методы машинного обучения, включая классификацию и регрессию, применяются для анализа финансовых данных, помогая принимать стратегические решения. Эти анализы на основе искусственного интеллекта предоставляют информацию, имеющую решающее значение для управления рисками, инвестиционных стратегий и обслуживания клиентов в финансовой отрасли.

Здравоохранение — еще одна область, где влияние ИИ на анализ больших данных имеет преобразующее значение. Алгоритмы ИИ способны выявлять закономерности в медицинских данных, которые могут помочь в ранней диагностике заболеваний, разработке новых лекарств и определении наиболее эффективных методов лечения. Глубокое обучение, в частности, применяется для анализа медицинских изображений и генетических данных, предлагая расширенные возможности диагностики и персонализированной медицины. Это не только улучшает результаты лечения пациентов, но и способствует развитию медицинских исследований. Роль ИИ в анализе экологических данных

становится все более важной в борьбе с изменением климата и сохранением окружающей среды. Методы машинного обучения используются для прогнозирования погодных условий, мониторинга состояния экосистем и разработки стратегий по сокращению выбросов загрязняющих веществ. Анализируя данные, связанные с климатом, загрязнением и биоразнообразием, ИИ помогает разрабатывать более эффективную экологическую политику и стратегии сохранения природы.

Сфера образования переживает стремительные перемены благодаря интеграции искусственного интеллекта. Алгоритмы ИИ анализируют индивидуальные характеристики учащихся, включая их знания, навыки и стили обучения, чтобы предоставлять персонализированные учебные материалы и задания. Такой адаптивный подход к обучению помогает учащимся учиться в своем собственном темпе и сосредотачиваться на тех областях, где им необходимо улучшить знания. Автоматизированные системы оценивания, работающие на основе искусственного интеллекта, эффективно оценивают ответы учащихся, особенно при решении задач с множественным выбором или математических задач, экономя преподавателям драгоценное время. Алгоритмы ИИ также предлагают персонализированные рекомендации по курсам, видео лекциям, учебникам и другим образовательным ресурсам на основе предпочтений и интересов учащихся. Для обеспечения академической честности ИИ используется для обнаружения плагиата в эссе и научных работах, а также для мониторинга студентов во время онлайн-экзаменов с целью предотвращения списывания. Виртуальные помощники и чат-боты, работающие на основе искусственного интеллекта, отвечают на вопросы студентов, помогают с домашними заданиями и предоставляют информацию о курсе. Это особенно полезно в массовых открытых онлайн-курсах, где личное взаимодействие с преподавателями может быть ограничено. Возможности предиктивной аналитики ИИ распространяются на прогнозирование успеваемости учащихся, оптимальной нагрузки и рисков отсева путем анализа таких данных, как посещаемость, оценки и активность в системах управления обучением.

Технологии искусственного интеллекта также используются для развития социальных навыков и эмоционального интеллекта в виртуальных средах, что особенно полезно для учащихся с особенностями развития или социальными тревожностями. Более того, само по себе образование в области ИИ является развивающейся областью,

совместно с образовательными учреждениями разрабатывают курсы и программы, направленные на подготовку студентов к карьере в области искусственного интеллекта, машинного обучения и робототехники. Наконец, технологии искусственного интеллекта анализируют данные учащихся, такие как успеваемость, интересы и стили обучения, чтобы предоставить ценную обратную связь педагогам. Это помогает учителям совершенствовать методы обучения и адаптировать материалы для максимальной эффективности. ИИ также используется при создании контента, такого как видеоролики, анимация, тексты и визуализации, что позволяет педагогам сосредоточиться на стратегиях обучения, а не на рутинных задачах.

Поскольку искусственный интеллект в образовании продолжает развиваться, он привносит инновации в традиционные методы обучения, предлагая возможности для персонализированного обучения, повышения эффективности преподавания и улучшения взаимодействия между преподавателями и учениками. Однако при разработке и внедрении решений на основе ИИ в образовании важно учитывать потенциальные проблемы, такие как конфиденциальность данных, этические вопросы и неравный доступ к технологиям.

***Аппаратные и программные технологии для применения искусственного интеллекта в анализе больших данных.***

В сфере искусственного интеллекта и анализа больших данных синергия между передовыми аппаратными и программными технологиями играет ключевую роль. Эти технологии не только обеспечивают необходимую вычислительную мощность, но и

предоставляют необходимые инструменты и структуры для эффективной обработки и анализа данных с помощью ИИ. Высокопроизводительные вычислительные системы находятся на переднем крае аппаратных технологий, позволяющих применять ИИ в работе с большими данными. Эти системы, включая суперкомпьютеры и распределенные вычислительные среды, обладают способностью обрабатывать и анализировать большие наборы данных гораздо быстрее, чем стандартные вычислительные системы. Графические процессоры, изначально разработанные для рендеринга графики, стали краеугольным камнем в обработке ИИ, особенно для алгоритмов глубокого обучения. Их возможности параллельной обработки особенно подходят для выполнения сложных вычислений, необходимых в ИИ. Тензорные процессоры, специально разработанные для задач машинного обучения, оптимизируют операции ИИ, особенно при вычислениях нейронных сетей. Кроме того, развивающаяся область квантовых вычислений открывает будущее, в котором сложные вычисления ИИ можно будет выполнять за доли текущего времени, что обещает значительный скачок в возможностях обработки данных. С точки зрения программного обеспечения незаменимыми стали такие фреймворки машинного обучения, как TensorFlow, PyTorch и Keras. Они предлагают готовые функции и структуры, которые упрощают создание и обучение моделей машинного обучения. Инструменты обработки больших данных, такие как Apache Hadoop и Apache Spark, обеспечивают необходимую структуру для распределенного хранения и обработки, что имеет решающее значение для обработки и анализа огромных объемов данных. Инструменты визуализации данных, такие как Tableau и Power BI, играют важную роль в переводе результатов анализа с помощью ИИ в понятные и применимые на практике форматы. Они позволяют преобразовывать сложные шаблоны данных в более понятные визуальные представления.

Платформы облачных вычислений, такие как AWS, Google Cloud и Microsoft Azure, играют решающую роль, предлагая масштабируемые и гибкие ресурсы. Эти платформы предоставляют целый ряд услуг — от хранения данных до расширенных возможностей машинного обучения, позволяя пользователям получать доступ к вычислительной мощности высокого уровня без необходимости в обширной физической инфраструктуре. Эффективное применение ИИ в анализе больших данных зависит от бесшовной интеграции этих аппаратных и программных технологий. Такая интеграция не только обеспечивает эффективную обработку больших наборов данных, но и позволяет применять сложные алгоритмы и получать полезные выводы. Это требует стратегического подхода для обеспечения совместимости и оптимальной производительности различных технологий.

Ландшафт аппаратных и программных технологий для ИИ в анализе больших данных разнообразен и динамичен. Поскольку эти технологии продолжают развиваться, они будут способствовать дальнейшему расширению возможностей и сфер применения ИИ в различных областях, делая анализ больших данных более эффективным, точным и информативным. Понимание и эффективное использование этих технологий является ключом к раскрытию полного потенциала ИИ в сфере больших данных. На схеме, представленной на рисунке, показано взаимодействие между ключевыми компонентами, участвующими в применении ИИ и машинного обучения для анализа больших данных. Все начинается с того, что пользователь генерирует новые данные, которые собираются различными источниками данных. Источники данных создают новые объекты для сохранения необработанных данных в компоненте хранения данных.

Параллельно компонент модели обработки данных и машинного обучения извлекает данные из хранилища для предварительной обработки и очистки. Затем он инициализирует модель машинного обучения, обучает ее на предварительно обработанных данных, настраивает гиперпараметры и завершает модель. После обучения модели ее развертывают путем создания обслуживающего компонента модели для составления прогнозов на основе новых данных.

Параллельно с этим выполняется анализ новых данных путем получения новых данных из хранилища, составления прогнозов с использованием развернутой модели, визуализации результатов и предоставления аналитических сведений пользователю.

Компонент развертывания модели также получает данные в реальном времени из источников, делает прогнозы и отправляет их на визуализацию.

В целом, он дает общее представление о применении ИИ и машинного обучения к большим данным, разбивая этапы и демонстрируя взаимодействие между компонентами.

**Заключение.** Анализ больших данных с использованием искусственного интеллекта и машинного обучения открыл новые горизонты для извлечения полезной информации из больших и разнообразных наборов данных во многих отраслях и приложениях. Однако, каким бы многообещающим ни казалось это сочетание, существуют определенные ограничения и риски, которые необходимо признать и смягчить для его эффективного и этичного использования.

Одной из ключевых задач является разработка моделей ИИ/МО (машинное обучение), которые могут предоставлять объяснимые прогнозы и решения вместо непостижимых «черных ящиков». Такие методы, как диагностические объяснения с использованием локально интерпретируемых моделей, могут пролить свет на то, почему модели делают определенные прогнозы, повышая доверие и подотчетность. Кроме того, необходимо использовать такие методы, как состязательное обучение, проверки надежности и анализ чувствительности, чтобы уменьшить предвзятость и построить более обобщенные модели, менее подверженные неожиданным ошибкам.

### Список литературы

[1] Z.M.Gizatullin, M.S.Shkinderov, R.R.Mubarakov, Proceedings of the 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering, 1350-1353, 2022.

[2] R. F. Gibadullin, N. S. Marushkai, 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 404-409, 2021.

[3] R. F. Gibadullin, I. S. Vershinin, R. Sh. Minyazev, 2017 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 1-6, 2017.

## UNLOCKING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE FOR BIG DATA ANALYTICS

***M.Tajieva***

*Lecturer of the Department of Software of information technologies,*
*The Institute of Telecommunications and Informatics of Turkmenistan*

**Annotation.** This article comprehensively examines the use of artificial intelligence (AI) in big data analytics. It focuses on machine learning and deep learning methods that are leveraged to develop innovative algorithms and solutions across domains like finance, healthcare, environment, and education. The article discusses the benefits of applying AI to big data analysis such as improved efficiency and accuracy of predictions, as well as optimization of decisions. However, it also highlights downsides and challenges such as information processing and security, privacy concerns, and ethical considerations.

**Keywords**: Artificial intelligence (AI), big data, learning, machine learning (ML).

УДК 004.041:655.26: 004.51

# ВЛИЯНИЕ БОЛЬШИХ ДАННЫХ НА СОЗДАНИЕ ИНТЕРФЕЙСОВ

***Ю.Р. Пигарева***
*Магистр университета СПБГУТ*
pigareva.julia.spb.p@mail.ru

***Б.А. Аль-Нами***
*Доцент кафедры информатики и компьютерного дизайна СПбГУТ, кандидат технических наук, доцент*
alnomibasheer@gmail.com

***Ю. Р. Пигарева***

*Окончила Санкт-Петербургский государственный университет телекоммуникаций имени профессора М. А. Бонч-Бруевича по направлению «дизайн графических и пользовательских интерфейсов информационных систем».*

***Б.А. Аль-Нами***

*Окончил Санкт-Петербургский электротехнический университет «ЛЭТИ». Научные интересы связаны с проектированием виртуальных средств информационных технологий, 3d полигональным моделированием и эргономикой, стандартизацией и проектированием (юзабилити) программных продуктов, Web-проектированием и дизайном, информатикой и компьютерным дизайном. Визуальное проектирование интерфейсов цифровых продуктов, информационные технологии, Информационные технологии визуализации данных и исследовательских процессов в Санкт-Петербургском государственном университете телекоммуникаций им. профессора М. А. Бонч-Бруевича.*

**Аннотация.** В данной работе рассматривается влияние больших данных на создание интерфейсов. Анализируются ключевые характеристики Big Data и их применение в UX/UI-дизайне. Рассмотрены примеры использования big data в персонализации контента, оптимизации пользовательского опыта.

**Ключевые слова:** дизайн, интерфейс, большие данные, big data, пользователь, UI/UX, пользовательский опыт.

**Введение.** Цифровые платформы активно используют большие данные для повышения удобства и эффективности пользовательских интерфейсов. Big Data позволяет анализировать поведение пользователей, выявлять их предпочтения и адаптировать контент под индивидуальные потребности. Взаимосвязь между анализом больших данных и дизайном интерфейсов играет важную роль в развитии торговых площадок[1,2].

Целью данной работы является изучение влияния больших данных на проектирование интерфейсов и выявление ключевых способов их применения для улучшения пользовательского опыта. Для достижения этой цели решаются следующие задачи:

– рассмотрение основных характеристик и источников больших данных;
– анализ методов использования Big Data в UX/UI-дизайне;
– исследование примеров успешного применения данных в цифровых продуктах;
– оценка влияния персонализированных интерфейсов на вовлеченность пользователей.

**Определение и характеристики big data.** Теперь необходимо понять, что такое big data. Big data (большие данные) – это огромный объем информации, который сложно

обработать стандартными методами. Эти данные могут быть структурированными (таблицы, базы данных) и неструктурированными (тексты, фото, видео, аудио) [3]. Они поступают из различных источников, таких как социальные сети, интернет вещей (IoT), поисковые системы, медицинские и финансовые системы [4].

Существует несколько основных характеристик больших данных:

– volume (Объем) - большие данные представляют собой огромные массивы информации, накапливаемые из различных источников;

– variety (Разнообразие) - данные поступают в разных форматах: текст, изображения, аудио, видео, базы данных;

– velocity (Скорость) - данные обрабатываются в режиме реального времени, что требует высокой вычислительной мощности;

– veracity (Достоверность) - важно учитывать точность и надежность данных, так как ошибки могут привести к неправильным выводам [5];

– value (Ценность) - анализ больших данных позволяет извлекать полезную информацию для оптимизации процессов и принятия решений.

Использование Big Data открывает новые возможности для автоматизации и оптимизации процессов, в том числе в сфере дизайна. Например, анализ данных позволяет выявлять предпочтения пользователей, персонализировать контент, прогнозировать тренды и разрабатывать более удобные и эффективные пользовательские интерфейсы.

Применение больших данных охватывает практически все сферы жизни: здравоохранение, транспорт, логистику, маркетинг, социальные сети, образование, развлечение, науку. В бизнесе они помогают анализировать поведение клиентов, в промышленности – оптимизировать производственные процессы, а в медицине – ставить точные диагнозы и прогнозировать развитие [6,7].

**Дизайн и его связь с big data.** Чтобы понять взаимосвязь между дизайном и большими данными, стоит понять, что такое дизайн. Дизайн – это процесс создания удобных, понятных и эстетически привлекательных решений для разных задач. Проще говоря, дизайн помогает сделать вещи более удобными и приятными в использовании. В контексте интерфейсов это означает, что дизайн делает взаимодействие пользователя с программами, сайтами и приложениями максимально интуитивным и эффективным. Дизайн сочетает в себе творческий и аналитический подход, включая работу с цветами, формами, композицией, а также анализ поведения пользователей и их потребностей [8].

Дизайн не может существовать без пользователей, так как дизайн направлен на решение задач и удовлетворение потребностей пользователей, а также привлечение внимания. Чтобы персонализировать дизайн и адаптировать информацию под каждого нужны данные о каждом человеке.

Откуда берутся те самые большие данные? Практически везде, где используются цифровые технологии, присутствует big data, а именно:

– социальные сети (посты, комментарии, сообщения);
– датчики (фитнес-браслеты, GPS);;
– транзакционные данные (банковские операции, покупки в интернет-магазинах);
– логи веб-сайтов (время на сайте, переходы между страницами);
– медицинские данные (диагнозы, анализы, лечение);
– мультимедийные данные (просмотры, жанры, предпочтения);
– данные из поисковых систем (поисковые запросы);
– данные из игровой индустрии (время в игре, достижения);
– данные из транспортных систем (маршруты, время в пути);
– образовательные платформы (успеваемость студентов, предпочтение по курсам);
– данные из рекламных систем (клики).

**Влияние больших данных на понимание пользователей.** Большие данные помогают проанализировать пользовательский опыт. Понять потребности, ожидания, поведение пользователей. С помощью этой информации дизайнеры и разработчики смогут создавать улучшенные интерфейсы, которые будут интуитивно понятными и удобными [9]. Основные факторы, влияющие на создание и улучшение пользовательского опыта:

1 **Анализ пользовательского поведения** – благодаря этому пункту можно понять, что чаще ищут, какими функциями пользуются больше всего, а какими меньше, с какой целью посещают сайт и с какими трудностями сталкивается человек в поисках информации.

2 **Разделение аудитории** – на различных сайтах важно понимать всю свою целевую аудиторию. Где-то это конкретная группа лиц, а где-то их несколько (как в интернет-магазинах). Даже если целевая аудитория одна, то интересы все равно будут местами отличаться от других людей. Другие главные факторы, влияющие на персонализацию: местоположение человека, возраст, устройство. Все эти факторы позволяют адаптировать интерфейсы под каждого члена целевой аудитории [10].

3 **Оптимизация пользовательского пути** – здесь проводится анализ пути пользователя, чтобы выявить сложности на каком-либо этапе, которые необходимо упростить. Например, если в процессе регистрации пользователи часто останавливаются или прерываются, это станет сигналом для команды разработчиков о необходимости смены или упрощения этого процесса.

Примеры использования больших данных в интерфейсах. Чтобы понимать роль big data в интерфейсах, важно проанализировать существующие интерфейсы, насколько большие данные помогают компаниям понимать поведение пользователей, предсказывать их предпочтения и оптимизировать взаимодействие с продуктом.

Большинство современных маркетплейсов применяют анализ больших данных для рекомендаций товаров. Система отслеживает какие товары смотрел пользователь, что добавил в избранное, что купил, какие поисковые запросы вводил человек. Так, на основе этой информации, платформа рекомендует товары, которые наверняка должны понравиться конкретному будущему покупателю. И у каждого человека будут свои рекомендации [2,5].

На рисунке 1, в качестве примера, представлена популярная торговая площадка Ozon [12]. Практически каждая страница данного приложения персонализирована. Платформа предлагает продукты, которые должны значительно повысить конверсию и стимулировать дополнительные покупки. Ozon активно использует большие данные для улучшения пользовательского опыта. Анализируя историю покупок, предпочтения, отзывы и даже географическое местоположение, маркетплейс формирует индивидуальные подборки товаров. Например, если пользователь часто заказывает спортивные товары, система предложит аксессуары или аналогичные продукты. Также Ozon использует алгоритмы прогнозирования, чтобы предлагать пользователям товары, которые могут быть востребованы в ближайшем будущем, например, сезонные товары, подарки на определенные праздники, продукты с ограниченным сроком годности. Все эти факторы влияют на рост продаж [9].

Ozon эффективно и успешно использует big data на своей площадке, об этом свидетельствует рейтинг и количество скачиваний приложения (рисунок 2) [13]. Большинство пользователей довольно как самой площадкой, так и ее интерфейсом.

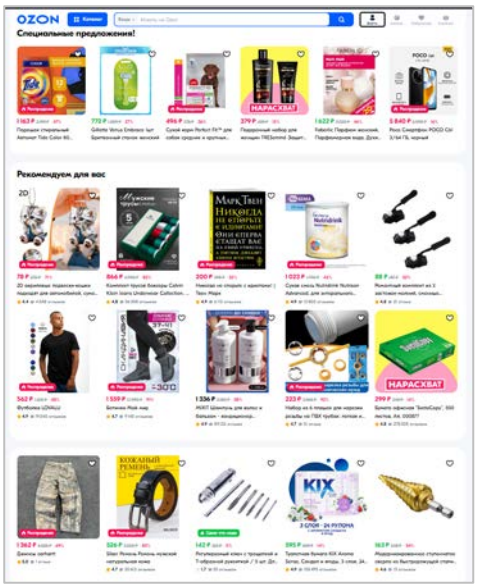



*Рисунок 1.* Маркетплейс Ozon

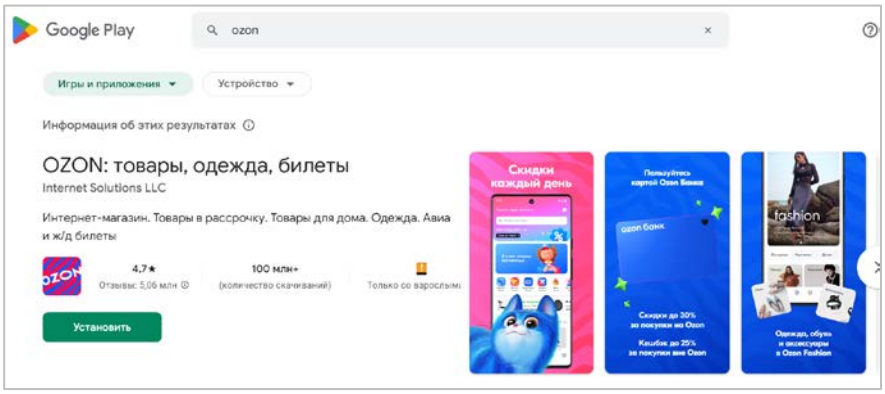



*Рисунок 2.* Рейтинг маркетплейса Ozon в Google Play

Также большие данные используют онлайн-кинотеатры. Они анализируют просмотры, паузы, перемотки, поисковые запросы, на основе которых платформа рекомендует фильмы или сериалы, которые могут понравится посетителю, что приведет к увеличению времени, проведенного на платформе, привлечению новых пользователей. Такими платформами являются Wink (рисунок 3), Okko и так далее. По такому же принципу работают музыкальные площадки Яндекс.Музыка, Spotify, но в другом направлении.

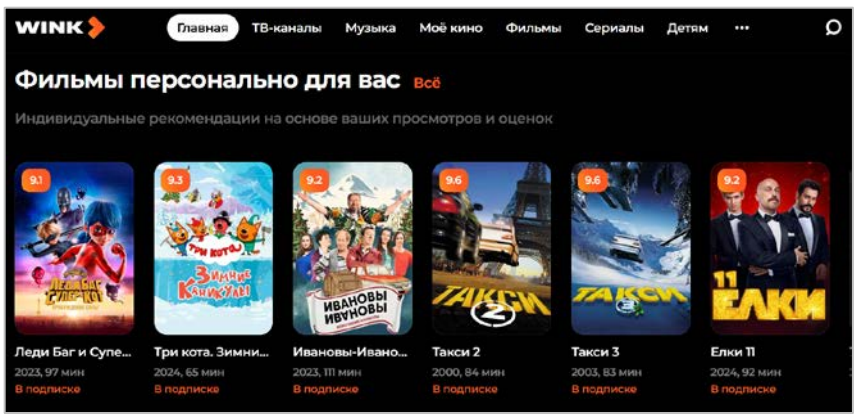



*Рисунок 3.* Онлайн-кинотеатр Wink

**Заключение.** Использование больших данных в дизайне интерфейсов открывает широкие возможности для персонализации, анализа поведения пользователей и оптимизации пользовательского пути. Маркетплейсы, онлайн-кинотеатры и музыкальные сервисы применяют алгоритмы обработки данных для создания более удобных и адаптированных интерфейсов. Анализ успешных кейсов показывает, что эффективное использование Big Data способствует росту вовлеченности пользователей, повышению конверсии и конкурентоспособности цифровых платформ.

## Список литературы

[1] Пигарева Ю.Р., Аль-Нами Б.А., Калинина С.С. Влияние пользовательского дизайна на веб-сайт логистической компании.// Информатика, управляющие системы, математическое и компьютерное моделирование (ИУСМКМ-2024). XV Международная научно-техническая конференция в рамках X Международного Научного форума Донецкой Народной Республики. Донецк, 2024. С. 225-228.

[2] Al-Nami B.A. AI - The future entity // Информатика, управляющие системы, математическое и компьютерное моделирование (ИУСМКМ-2024). XV Международная научно-техническая конференция в рамках X Международного Научного форума Донецкой Народной Республики. Донецк, 2024. С. 230-234.

[3] Аль-Нами Б.А., Пигарева Ю.Р. Разработка пользовательского интерфейса веб-сайта логистической компании // Актуальные проблемы инфотелекоммуникаций в науке и образовании. Материалы XIII Международной научно-технической и научно-методической конференции. Санкт-Петербург, 2024. С. 41-44.

[4] Аль-Нами Б.А., Попова Е.В. Инновационные тенденции в веб-дизайне: будущее пользовательского опыта в цифровой среде // Актуальные проблемы инфотелекоммуникаций в науке и образовании. Сборник научных статей XIII Международной научно-технической и научно-методической конференции. Санкт-Петербург, 2024. С. 506-510.

[5] Al-Nami B.A. Can big data and advanced analytics address problems of ethical, privacy, and equity concerns in university education? // BIG DATA и анализ высокого уровня. сборник научных статей X Международной научно-практической конференции в двух частях. Минск, 2024. С. 55-66.

[6] Al-Nami B.A. The problems and obstacles facing the designers of the virtual world // Информатика, управляющие системы, математическое и компьютерное моделирование (ИУСМКМ-2023). Материалы XIV Международной научно-технической конференции в рамках IX Международного Научного форума Донецкой Народной Республики. Донецк, 2023. С. 367-371.

[7] Al-nami B.A. Converting images from one extension to another without losing its quality, problems and solutions // В сборнике: BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ. Сборник научных статей IX международной научно-практической конференции: в 2х частях . Минск, 2023. С. 51-61.

[8] Аль-Нами Б.А., Григорьева С.Д. Влияние дизайна сайта на продажи онлайн-бизнеса // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2023). Сборник научных статей XII Международной научно-технической и научно-методической конференции. В 4-х томах. Под редакцией С.И. Макаренк, сост. В.С. Елагин, Е.А. Аникевич. Санкт-Петербург, 2023. С. 113-116.

[9] Al-Nami B.A., Musaeva T. Modern methods to solve the problems of adapting images in web pages and the possibility of using artificial intelligence to do so in the future // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2023). Сборник научных статей XII Международной научно-технической и научно-методической конференции. В 4-х томах. Под редакцией С.И. Макаренк, сост. В.С. Елагин, Е.А. Аникевич. Санкт-Петербург, 2023. С. 180-185.

[10] Аль-Нами Б.А., Рстакян А.А. Взаимосвязь между дизайном и маркетингом: как дизайн влияет на продажи // Актуальные проблемы инфотелекоммуникаций в науке и образовании (АПИНО 2023). Сборник научных статей XII Международной научно-технической и научно-методической конференции. В 4-х томах. Под редакцией С.И. Макаренк, сост. В.С. Елагин, Е.А. Аникевич. Санкт-Петербург, 2023. С. 206-209.

# THE IMPACT OF BIG DATA ON INTERFACE DESIGN

***J.R. Pigareva***
*Master's Degree,*
*St. Petersburg State University of Telecommunications*

***B.A. Al-Nami***
*Associate Professor,*
*Department of Informatics and Computer Design, SPBSUT, Ph.D. (in technical sciences),*

**Abstract.** This paper examines the impact of big data on interface design. The key characteristics of Big Data and their application in UX/UI design are analyzed. Examples of using big data for content personalization and user experience optimization are considered.

**Keywords**: design, interface, big data, user, UI/UX, user experience.

УДК 004.522

# РОЛЬ БОЛЬШИХ ДАННЫХ В АНАЛИЗЕ ГЕНОМНЫХ ДАННЫХ

***О.М. Аббасова***
*Преподаватель кафедры «Программное обеспечение информационных технологий»*
*Институт Телекоммуникаций и информатики Туркменистана*
ogultaganabbasowa@gmail.com

***Г.Г. Аннасапаров***
*Преподаватель кафедры «Программное обеспечение информационных технологий»*
*Институт Телекоммуникаций и информатики Туркменистана*
guwancannasaparow187@gmail.com

***О.М. Аббасова***

*Окончил институт Телекоммуникаций и информатики Туркменистана. Область научных интересов связана с програмированием и использованием искусственного интеллекта.*

***Г.Г. Аннасапаров***

*Окончил Туркменский государственный университет имени Махтумкули. Область научных интересов связана с математикой и программированием.*

**Аннотация.** Системы здравоохранения, генетика и геномика, население и общественное здравоохранение; все области биомедицины могут выиграть от Больших данных и связанных с ними технологий. При анализе геномных данных часто требуется интеграция различных типов данных (например, геномные, клинические, экологические). Поэтому важно владеть языками и инструментами, которые позволяют эффективно обрабатывать и интегрировать разнородные данные. Машинное обучение играет все более важную роль в анализе геномных данных. Python с его библиотеками машинного обучения является отличным выбором для этой задачи.

**Ключевые слова:** Медицинская информатика, интеллектуальный анализ данных, интеллектуальный анализ текстов, информационные системы, хранение и поиск информации, bioPython, Python.

**Введение.** Big Data — это горячая тема в здравоохранении и биомедицинских исследованиях. Более частое использование термина «Big Data» в биомедицинской литературе свидетельствует о растущей важности крупномасштабных наборов данных в здравоохранении и биомедицине, а также растет понимание роли, которую большие данные могут играть в научных и клинических исследованиях. Также было отмечено, что термин «Big Data» может означать разные вещи для разных групп людей. Однако было общее признание того, что здравоохранение, биомедицинские исследования и здоровье населения генерируют огромные, сложные, распределенные и часто динамические наборы данных, и что размер и сложность этих данных будут создавать как проблемы, так и возможности для организаций здравоохранения.

Считается, что термин «большие данные» возник в поисковых компаниях, которым приходилось запрашивать очень большие распределенные агрегации слабо структурированных данных.

С тех пор этот термин используется для обозначения огромных объемов данных, собранных с течением времени, которые трудно анализировать и обрабатывать с помощью обычных инструментов управления базами данных. Хотя может показаться, что этот

термин относится к объему данных, это не всегда так. В исследовательском отчете 2001 года аналитик Gartner (ранее META Group) Дуг Лэйни определил проблемы и возможности роста данных как трехмерные, т. е. увеличивающийся объем (количество данных), скорость (скорость входящих и исходящих данных) и разнообразие (диапазон типов данных и источников). Большая часть отрасли продолжает использовать эту модель «3 Vs» для описания больших данных. В 2012 году Gartner обновил свое определение следующим образом: «Большие данные — это информационные активы большого объема, высокой скорости и/или большого разнообразия, которые требуют новых форм обработки для обеспечения улучшенного принятия решений, обнаружения информации и оптимизации процессов». Кроме того, некоторые организации добавили новую букву V для «Veracity» для описания. Термин «Большие данные» может также относиться к технологии (например, хранилищам, инструментам и процессам), которые требуются организации для обработки больших объемов данных. Более прагматичное определение определяет Большие данные с точки зрения требования к аналитическим приложениям для обработки новых типов данных. С момента публикации первой последовательности генома человека в 2003 году область геномики стала основной движущей силой генерации больших данных в биомедицине. Прогресс в лабораторных аналитических методах (например, анализ последовательности ДНК) и мобильных технологиях (например, данные с мониторов физической активности и приложений) в настоящее время в значительной степени ответственны за постоянно растущее производство данных в реальном времени в больших объемах.

Однако использование больших данных теперь достигло всех областей здравоохранения, биомедицинских исследований и здоровья населения. Исследователи служб здравоохранения могут объединять административные и клинические базы данных для разработки прогностических моделей с целью улучшения политики здравоохранения. Фармацевтическая промышленность управляет огромными хранилищами клинических и молекулярных данных для рационального дизайна лекарств и подходов фармакогеномики. В области здоровья населения регистры заболеваний и данные из клинических записей используются для измерения воздействия вмешательств в здравоохранение. Биомедицинские исследователи имеют доступ к новым источникам геномных данных (например, микробиом, эпигеномика) и могут изучать новые гипотезы для понимания молекулярных причин заболеваний. Наконец, экологические данные начинают интегрироваться с генетическими и клиническими данными, которые (организация) ранее не отслеживала.

Хотя формального определения нет, обычно термин «структурированные данные» относится к данным с определенной схемой или моделью данных (т. е. явной семантикой). Данные, хранящиеся в базе данных, обычно структурированы. Измерения и сигналы являются примерами структурированных данных. Напротив, неструктурированные данные относятся к данным, содержащим информацию, которая нелегко доступна для систем управления вычислительными данными — содержащаяся в них информация не представлена в форме с четкой схемой данных, которая позволяет проводить прямую вычислительную интерпретацию и анализ. Этот тип данных обычно требует специализированных аналитических методов для извлечения содержащейся в них информации и преобразования ее в вычислимую форму. Текст на естественном языке, изображения и аудиопотоки являются примерами неструктурированных данных.

Существует огромное количество информации, относящейся к пониманию здоровья человека, зафиксированной в неструктурированных ресурсах. Тексты, изображения, аудио- и видеопотоки обычно производятся в клиническом контексте. Эти ресурсы требуют разработки стратегий для извлечения и обобщения содержащейся в них информации, чтобы наложить структуру и значение, используя внутреннюю структуру или закономерности, присущие источнику. Технологические решения, позволяющие автоматически

интерпретировать такие ресурсы, в дополнение к помощи людям, которым поручено интерпретировать эти источники данных, позволяют масштабировать анализ и рассматривать гораздо более широкий набор данных — по больнице, населению или научно-исследовательскому сообществу. Это привело к некоторым весьма инновационным исследованиям, демонстрирующим силу крупномасштабного анализа данных в медицине.

Анализ геномных данных — это процесс изучения и интерпретации генетической информации, закодированной в ДНК организма. Он включает в себя использование различных методов и инструментов для выявления закономерностей, вариаций и мутаций в геноме, а также для определения их влияния на биологические процессы, здоровье и болезни.

Геномные данные относятся к полной генетической информации организма, хранящейся в его ДНК. Эта информация включает в себя последовательность оснований ДНК, гены, регуляторные элементы и другие функциональные элементы.

**Проблемы анализа геномных данных.** Анализ геномных данных представляет собой уникальные проблемы из-за огромного объема данных, сложности генома и необходимости интеграции данных из различных источников.

Большие данные — это термин, используемый для описания огромных объемов данных, которые настолько сложны, что традиционные методы обработки данных не могут их эффективно обрабатывать. Большие данные предлагают несколько преимуществ для анализа геномных данных:

Масштабируемость: системы больших данных могут обрабатывать и анализировать огромные объемы геномных данных, что позволяет исследователям изучать большие популяции и выявлять редкие генетические вариации.

Скорость: технологии больших данных могут анализировать геномные данные гораздо быстрее, чем традиционные методы, что позволяет исследователям быстрее получать результаты.

Интеграция данных: платформы больших данных могут интегрировать геномные данные с другими типами данных, такими как клинические данные и данные о стиле жизни, для получения более полного представления о здоровье человека.

Машинное обучение: алгоритмы машинного обучения могут анализировать геномные данные для выявления закономерностей и взаимосвязей, которые трудно обнаружить с помощью традиционных статистических методов.

Для анализа геномных данных с использованием технологий больших данных применяется широкий спектр языков программирования и инструментов. Выбор конкретного языка часто зависит от специфики задачи, предпочтений аналитика и доступных ресурсов. Вот наиболее популярные и востребованные языки:

**Основные языки программирования:**

**Python:**

– широко используется в биоинформатике и анализе данных благодаря богатому набору библиотек (например, NumPy, Pandas, SciPy, scikit-learn);

– идеален для статистического анализа, машинного обучения и визуализации данных;

– библиотеки, такие как Biopython, специально разработаны для работы с биологическими данными.

**R:**

– специализированный язык для статистического анализа и визуализации данных;

– имеет обширный набор пакетов для биоинформатики и геномного анализа (например, Bioconductor);

– часто используется для анализа экспрессии генов, генетической вариации и других геномных данных.

**Java:**

– используется для разработки высокопроизводительных приложений и систем обработки больших данных;

– применяется в разработке биоинформатических инструментов и платформ;

– используется для таких технологий как Apache Hadoop.

**C/C++:**

– обеспечивают высокую производительность и используются для разработки низкоуровневых биоинформатических инструментов и алгоритмов;

– используются, когда критична скорость выполнения.

Python — мощный инструмент для анализа геномных данных, благодаря своему широкому набору библиотек и простоте использования.

**Заключение.** Анализ геномных данных на Python открывает широкие возможности для исследований в области биоинформатики, медицины и других наук. Благодаря мощным библиотекам, таким как Biopython, Pandas и Scikit-learn, Python позволяет эффективно обрабатывать, анализировать и визуализировать сложные геномные данные. Мы рассмотрели основные шаги анализа, начиная от загрузки данных из распространенных форматов, таких как FASTA и VCF, до анализа последовательностей и генетических вариаций. Визуализация данных с помощью Matplotlib и Seaborn помогает лучше понять результаты анализа, а применение алгоритмов машинного обучения позволяет выявлять скрытые закономерности и делать прогнозы. Однако важно помнить, что анализ геномных данных требует не только технических навыков, но и глубокого понимания биологических процессов. Интерпретация результатов должна проводиться с учетом контекста и с использованием соответствующих биологических знаний.

## Список литературы

[1] PC Magazine Encyclopedia. http://www.pcmag.com/encyclopedia/term/62849/big-data.

[2] O'Driscoll A, Daugelaite J, Sleator RD. 'Big data', Hadoop and cloud computing in genomics. J Biomed Inform 2013

# THE ROLE OF BIG DATA IN GENOMIC DATA ANALYSIS

***O. Abbasova***
*Lecturer of the Department of Software of information technologies,*
*The Institute of Telecommunications and Informatics of Turkmenistan*

***G. Annasaparov***
*Lecturer of the Department of Software of information technologies,*
*The Institute of Telecommunications and Informatics of Turkmenistan*

**Annotation.** Health systems, genetics and genomics, population and public health; all areas of biomedicine can benefit from Big Data and related technologies. Genomic data analysis often requires the integration of different types of data (e.g. genomic, clinical, environmental). It is therefore important to have languages and tools that allow efficient processing and integration of heterogeneous data. Machine learning is playing an increasingly important role in genomic data analysis. Python with its machine learning libraries is an excellent choice for this task..

**Keywords**: Medical informatics, data mining, text mining, information systems, information storage and retrieval, bioPython, Python.

УДК 004.855.5

# ВЫБОР ФУНКЦИИ АКТИВАЦИИ И ФУНКЦИИ ПОТЕРЬ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ АКТИВНОСТЕЙ ЧЕЛОВЕКА

***В.В. Венгеренко***
*Мл. науч. сотр. ОИПИ НАН Беларуси, аспирант ОИПИ НАН Беларуси*
vengerenkovadim1@gmail.com

***М.М. Лукашевич***
*Доцент кафедры информационных систем управления БГУ, ст. науч. сотр. лаборатории №222 ОИПИ НАН Беларуси, кандидат технических наук, доцент*
LukashevichMM@bsu.by

***В.В. Венгеренко***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов машинного обучения для решения задач анализа данных и компьютерного зрения.*

***М.М. Лукашевич***
*Является доцентом кафедры информационных систем управления БГУ, ст. науч. сотр. лаборатории №222 ОИПИ. Область научных интересов связана с разработкой методов и алгоритмов машинного обучения для решения задач анализа данных и компьютерного зрения.*

**Аннотация.** Цель работы – исследовать влияние выбора функции активации и функции потерь на показатели нейросетевых моделей при решении задачи классификации активностей человека.

Представлены результаты экспериментов по оценке показателей нейросетевых моделей, которые классифицируют активности человека и при этом используют различные комбинации функции активации и функции потерь. В экспериментальных исследованиях использовалась выборка из набора данных HMDB51 для 10 классов. Для рассматриваемой нейросетевой архитектуры была выявлена комбинация гиперпараметров, позволяющая достичь максимальной точности классификации.

**Ключевые слова:** нейронная сеть, функция активации, функция потерь, точность предсказаний, классификация.

**Введение.** Актуальность темы исследования заключается в том, что классификация активностей людей повышает уровень безопасности посредством приложений в видеонаблюдении реального времени, позволяя обнаруживать потенциально опасные ситуации, а также способствует достижениям в таких областях, как здравоохранение и робототехника.

Задача исследования состоит в оценке показателей нейросетевых моделей, которые классифицируют активности человека и при этом используют различные комбинации функции активации и функции потерь.

Распознавание активностей человека является актуальной задачей в компьютерном зрении и особенно полезно в системах видеонаблюдения. Можно выделить два подхода для распознавания активностей человека: на основе акселерометров и гироскопов и на основе визуальных данных (изображения и видео). Во втором подходе используется компьютерное

зрение, позволяющее определять сложные контекстно-зависимые активности. В настоящее время подход на основе нейронных сетей дает наиболее высокие результаты. Среди архитектур следует выделить архитектуру трансформер и сверточные нейронные сети [1–5], на которых строятся современные решения.

Для рассматриваемой задачи используются следующие наборы данных: HMDB51 [6], UCF101 [7], Kinetics [8], ActivityNet [9], Charades [10], THUMOS [11], HACS [12], MOD20 [13], Sports-1M [14], HAA500 [15].

Перечисленные наборы содержат видео различной длительности из таких источников, как YouTube, архивы Прелингера (the Prelinger archives), Google видео и так далее.

Среди перечисленных наборов MOD20 содержит наименьшее количество классов (20) и экземпляров (2324 видео). В наборе Kinetics-700 700 классов, что является наибольшим значением данного показателя для перечисленных наборов. Из крупных наборов данных стоит отметить HACS, в котором содержится 1,55 млн клипов длительностью 2 с, полученных из 504 тыс. видео, и 140 тыс. сегментов, полученных из 50 тыс. видео.

Примерами классов могут послужить такие классы UCF101, как “Typing” (набор на клавиатуре), “Playing Guitar” (игра на гитаре), “Rope Climbing” (лазание по канату), “Walking with a Dog” (прогулка с собакой), “Biking” (езда на велосипеде) [7].

**Экспериментальный набор данных.** Для проведения экспериментов была выбрана часть набора данных HMDB51, который в исходном виде включает в себя 6766 видео и 51 класс [6]. Экспериментальный набор содержит 500 видео из первых 10 классов (“brush_hair”, “cartwheel”, “catch”, “chew”, “clap”, “climb”, “climb_stairs”, “dive”, “draw_sword”, “dribble”) набора HMDB51. Данный экспериментальный набор делится на обучающую, валидационную и тестовую выборки в соотношении 60/20/20 соответственно. При этом применяется стратификация по классам. Каждая выборка заключается в 1 пакет данных (batch). Как видно из рисунка 1, данные в экспериментальном наборе и выборках сбалансированы по классам.

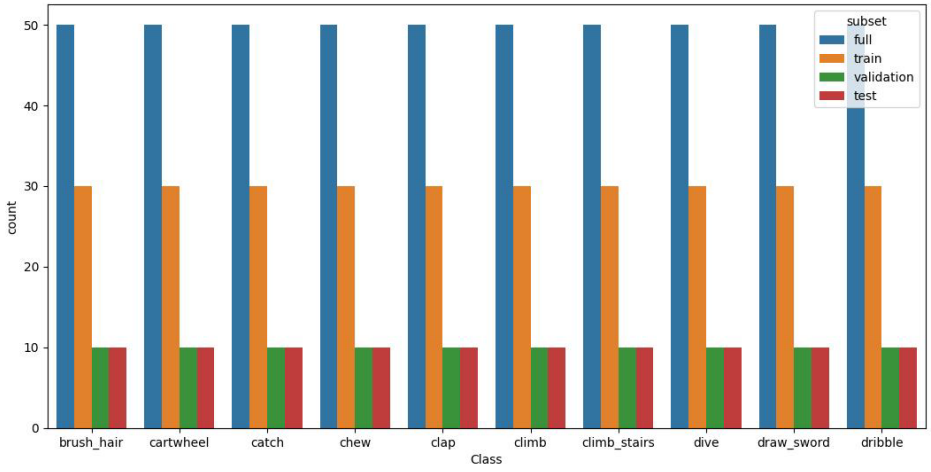



*Рисунок 1.* Распределение экспериментальных данных по классам и выборкам

Каждое видео из экспериментального набора содержит 40 кадров размером 224×224 пикселей. В случае, если длительность видео из исходного набора данных превышает установленное ограничение в 40 кадров, то последовательность кадров этой длины выбирается из середины видео. Иначе вся последовательность кадров из исходного видео зацикливается до нужной длины.

**Результаты исследований.** Эксперименты проводились с использованием языка программирования Python и его пакетов, из которых стоит отметить tensorflow, numpy, cv2, scikit-learn, pandas, seaborn, matplotlib. Для выполнения вычислений с помощью библиотеки

TensorFlow использовался графический процессор (Graphical Processing Unit, GPU) NVIDIA GeForce MX150.

Общая нейросетевая архитектура для проведения экспериментов по примеру в [16] представлена на рисунке 2.

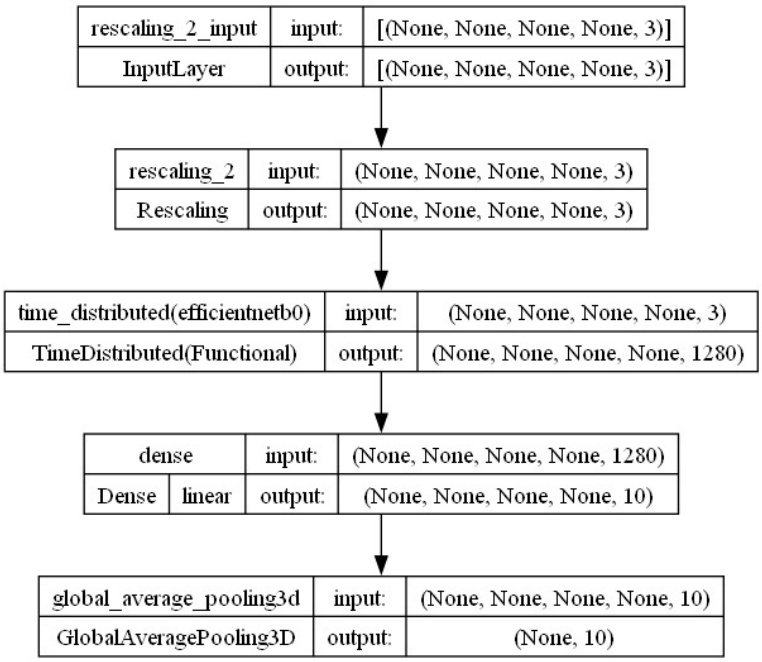



*Рисунок 2.* Общая архитектура нейронной сети для проведения экспериментов

Для предварительной обработки входных данных используется слой *Rescaling* с параметром *scale* со значением 255, что позволяет отмасштабировать входные значения из исходного диапазона [0, 1] в новый диапазон [0, 255].

Слой *TimeDistributed* является оберткой над слоем, передаваемым в качестве параметра, то есть в данном случае над предварительно обученной на наборе данных ImageNet (1000 классов) моделью EfficientNetB0, возвращаемой соответствующей функцией библиотеки TensorFlow. Это позволяет независимо применить оборачиваемый слой к каждому временному срезу входных данных, используя одинаковый набор весовых коэффициентов. Таким образом, в рассматриваемой задаче модель EfficientNetB0 независимо задействуется для каждого кадра отдельного видео из поступающего на нее набора данных.

Использование предварительно обученной модели EfficientNetB0 в составе экспериментальных нейросетевых моделей позволяет ускорить достижение высоких показателей при решении поставленной задачи. Параметры *include_top* и *trainable* данной модели принимают значение *False* для исключения раннего полносвязного слоя из этой модели и обеспечения неизменности значений весовых коэффициентов ее оставшихся слоев на этапе обучения экспериментальных моделей.

Предпоследний полносвязный слой *Dense* предназначен для соотнесения признаков, извлеченных на предыдущих слоях для каждого кадра отдельного видео, с 10 классами, обозначенными в решаемой задаче.

Последний слой *GlobalAveragePooling3D* выполняет операцию усреднения данных с предыдущего слоя по всем кадрам отдельного видео таким образом, что это позволяет соотнести анализируемое видео с 10 классами, обозначенными в решаемой задаче.

Все экспериментальные модели обучаются в течение 15 эпох с применением оптимизатора *Adam* с параметрами по умолчанию.

Точность предсказаний рассчитывается следующим образом [17]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (1)$$

где *TP* – True Positives (истинно положительные), количество положительных наблюдений, которые модель правильно предсказала как положительные;

*TN* – True Negatives (истинно отрицательные), количество отрицательных наблюдений, которые модель правильно предсказала как отрицательные;

*FP* – False Positives (ложноположительные или ошибка 1-го рода), количество отрицательных наблюдений, которые модель неверно предсказала как положительные;

*FN* – False Negatives (ложноотрицательные или ошибка 2-го рода), количество положительных наблюдений, которые модель неверно предсказала как отрицательные.

В TensorFlow для оценки точности предсказаний присутствуют такие метрики, как *SparseCategoricalAccuracy* (*SCA*) [18] и *CategoricalAccuracy* (*CA*) [19]. В первом случае истинная принадлежность экземпляра определенному классу должна определяется как число, соответствующее этому классу, например, его порядковый номер. Во втором – как one-hot вектор, длина которого равна количеству классов в задаче, при этом в случае истинной принадлежности рассматриваемого экземпляра классу с порядковым номером *i* метка класса в *i*-й позиции данного вектора принимает значение 1, иначе – 0.

В экспериментах использовались функции активации [20] и функции потерь [21; 22], представленные в таблицах 1 и 2 соответственно.

Таблица 1. Функции активации, используемые в экспериментах

| Название функции активации | Формула |
|---|---|
| *Softmax* | $\frac{e^x}{\sum_{i=0}^{K-1} e^x}$ |
| *Swish* (или *SiLU* – Sigmoid-weighted Linear Unit) | $x \cdot \frac{1}{1+e^{-x}}$ |
| *Hard Swish* (или *Hard SiLU*) | $\begin{cases} 0, & x < -3, \\ \frac{x \cdot (x+3)}{6}, & -3 \le x \le 3, \\ x, & x > 3 \end{cases}$ |
| *Mish* | $x \cdot tanh(\log(e^x + 1))$ |
| *GELU* (Gaussian Error Linear Unit) | $0{,}5 \cdot x \cdot (1 + \frac{2}{\sqrt{\pi}} \int_0^{\frac{x}{\sqrt{2}}} e^{-t^2} dt)$ |
| *CELU* (Continuously-differentiable Exponential Linear Unit) | $\begin{cases} \alpha \cdot (e^{\frac{x}{\alpha}} - 1), & x < 0, \\ x, & x \ge 0 \end{cases}$ |

Таблица 2. Функции потерь, используемые в экспериментах

| Название функции потерь | Формула |
|---|---|
| *CategoricalCrossentropy* (категориальная перекрестная энтропия, *CCE*) | $-\sum_{i=0}^{C-1} y_{true\,i} \cdot \log(y_{pred\,i})$ |
| *SparseCategoricalCrossentropy* (разреженная *CCE*, *SCCE*) | $-\log(y_{pred\,y_{true}})$ |
| *CategoricalFocalCrossentropy* (категориальная фокальная перекрестная энтропия, *CFCE*) | $-\alpha \cdot \sum_{i=0}^{C-1} (1 - y_{pred\,i})^{\gamma} \cdot y_{true\,i} \cdot \log(y_{pred\,i})$ |

| | |
|---|---|
| *CategoricalHinge* (категориальная верхняя граница, *CH*) | $\max\left(1 + \max\left((1 - y_{true}) \odot y_{pred}\right) - \sum_{i=0}^{C-1}\left(y_{true} \odot y_{pred}\right)_i, 0\right)$ |
| *KLDivergence* (Kullback-Leibler Divergence, *KLD*; расстояние Кульбака-Лейблера) | $KLD(y_{true}, y_{pred}) = \sum_{i=0}^{C-1} y_{true\,i} \cdot \log\left(\frac{y_{true\,i}}{y_{pred\,i}}\right)$ |

В первой части экспериментов, посвященной исследованию влияния выбора применяемой между сверточными слоями EfficientNetB0 функции активации на показатели экспериментальных моделей, для оценки точности предсказаний использовалась метрика *SCA* и функция потерь *SCCE* (разреженная категориальная перекрестная энтропия) с параметром *from_logits* со значением *True* (в таком случае к предсказаниям модели перед непосредственным расчетом значения *SCCE* будет применяться функция *Softmax*). На предпоследнем полносвязном слое используется линейная (*Linear*) функция активации, пропускающая входные данные без изменений [20].

Функция активации задается параметром *activation* функции tensorflow.keras.applications.EfficientNetB0.

Результат применения функции активации *Swish* между сверточными слоями EfficientNetB0 представлен на рисунке 3.

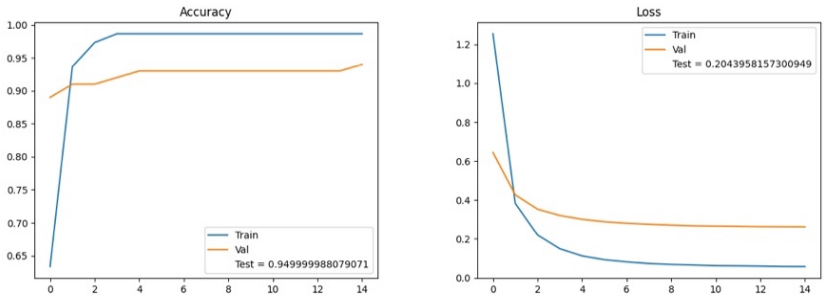



*Рисунок 3.* Графики точности (слева) и функции потерь (справа) при применении между сверточными слоями EfficientNetB0 функции активации *Swish*

Результат применения функции активации *Hard Swish* между сверточными слоями EfficientNetB0 представлен на рисунке 4.

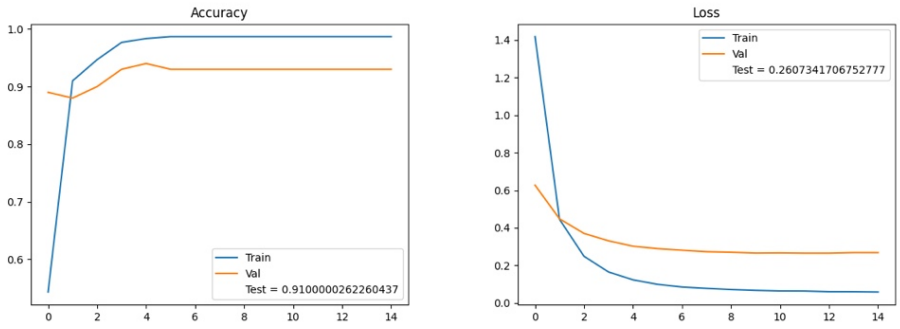



*Рисунок 4.* Графики точности (слева) и функции потерь (справа) при применении между сверточными слоями EfficientNetB0 функции активации *Hard Swish*

Результат применения функции активации *Mish* между сверточными слоями EfficientNetB0 представлен на рисунке 5.

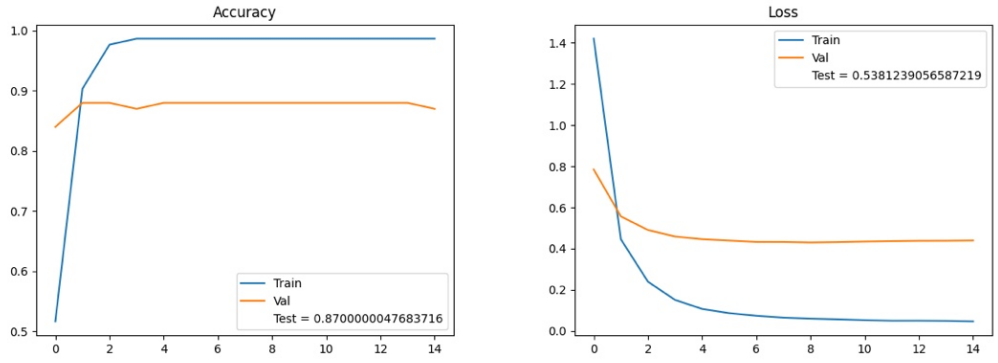



*Рисунок 5.* Графики точности (слева) и функции потерь (справа) при применении между сверточными слоями EfficientNetB0 функции активации *Mish*

Результат применения функции активации *GELU* между сверточными слоями EfficientNetB0 представлен на рисунке 6.

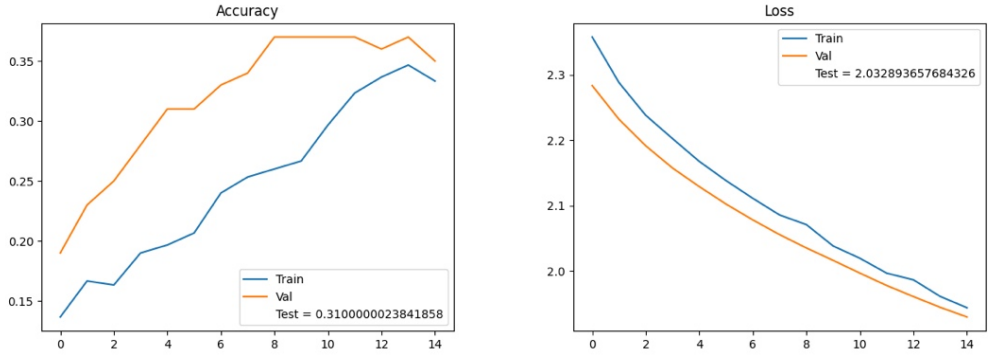



*Рисунок 6.* Графики точности (слева) и функции потерь (справа) при применении между сверточными слоями EfficientNetB0 функции активации *GELU*

Результат применения функции активации *CELU* (параметр $\alpha$ равен 0,5) между сверточными слоями EfficientNetB0 представлен на рисунке 7.

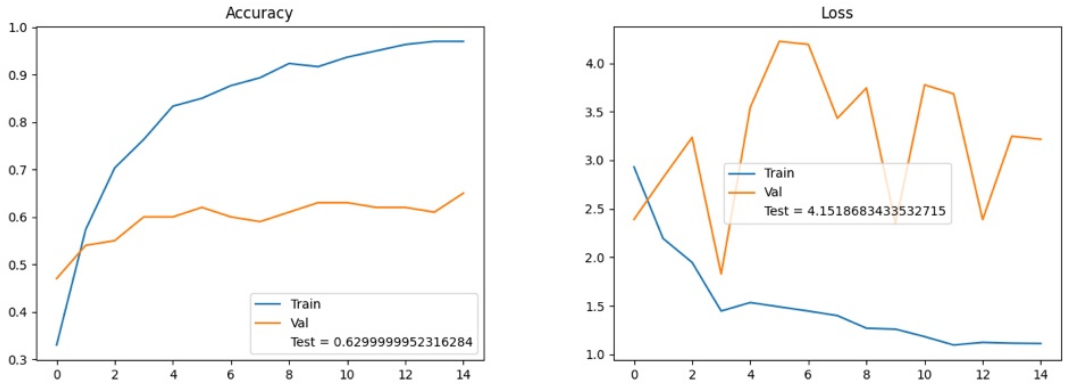



*Рисунок 7.* Графики точности (слева) и функции потерь (справа) при применении между сверточными слоями EfficientNetB0 функции активации *CELU* (параметр $\alpha$ равен 0,5)

Во второй части экспериментов, посвященной исследованию влияния выбора функции потерь на показатели экспериментальных моделей, на предпоследнем слое EfficientNetB0 используется функция активации *Softmax*, а параметр *from_logits* при его наличии у функции потерь принимает значение *False*. Между сверточными слоями EfficientNetB0 применяется функция активации *Swish*, поскольку в первой части экспериментов она позволила добиться лучших показателей.

Результат выбора *SCCE* в качестве функции потерь представлен на рисунке 8. Метрикой оценки точности в данном случае выступает *SCA*.

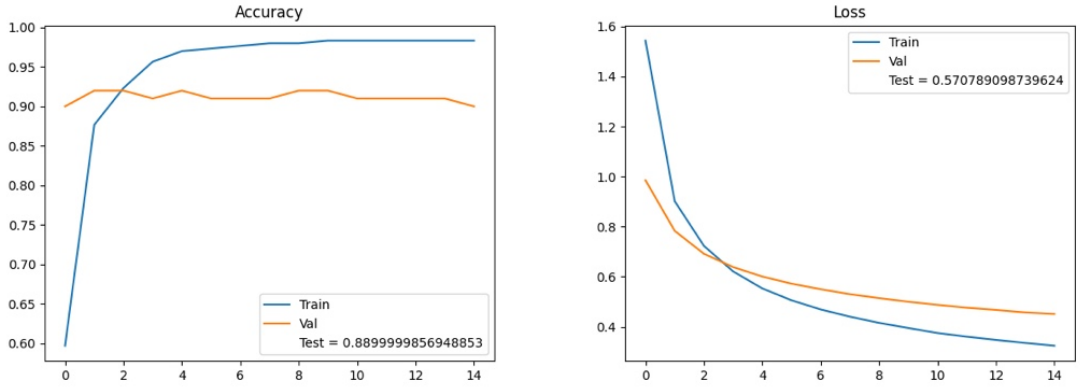



*Рисунок 8.* Графики точности (слева) и функции потерь *SCCE* (справа)

В оставшихся экспериментах этой части метрикой оценки точности выступает *CA*, а $y_{true}$ задается как one-hot вектор.

Результат выбора в качестве функции потерь *CCE* представлен на рисунке 9.

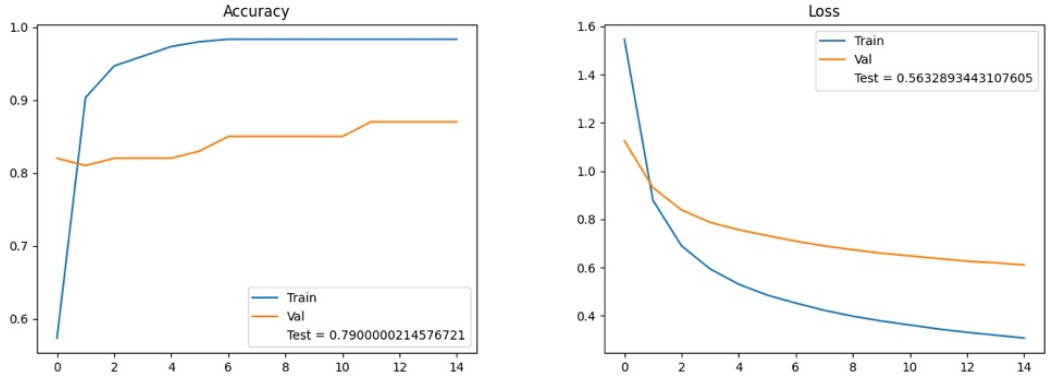



*Рисунок 9.* Графики точности (слева) и функции потерь *CCE* (справа)

Результат выбора в качестве функции потерь *CFCE* с параметрами $\alpha$ со значением 0,25 и $\gamma$ со значением 2 представлен на рисунке 10.

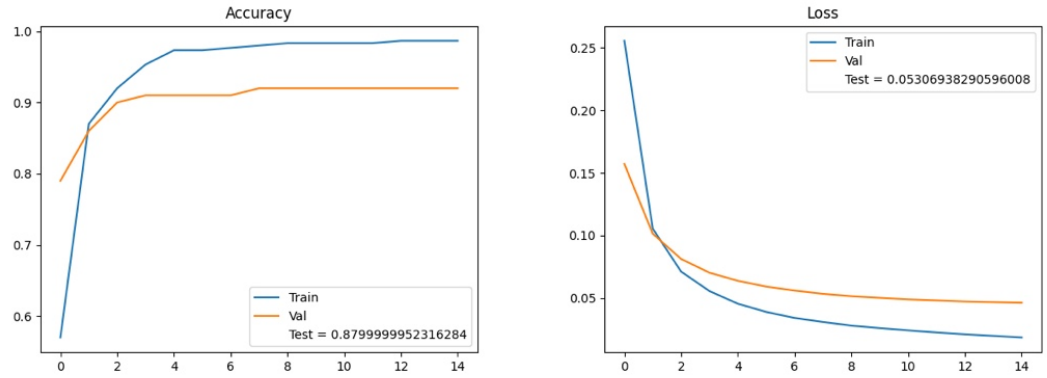



*Рисунок 10.* Графики точности (слева) и функции потерь *CFCE* (справа), $\alpha$ со значением 0,25 и $\gamma$ со значением 2

Результат выбора в качестве функции потерь *CH* представлен на рисунке 11.

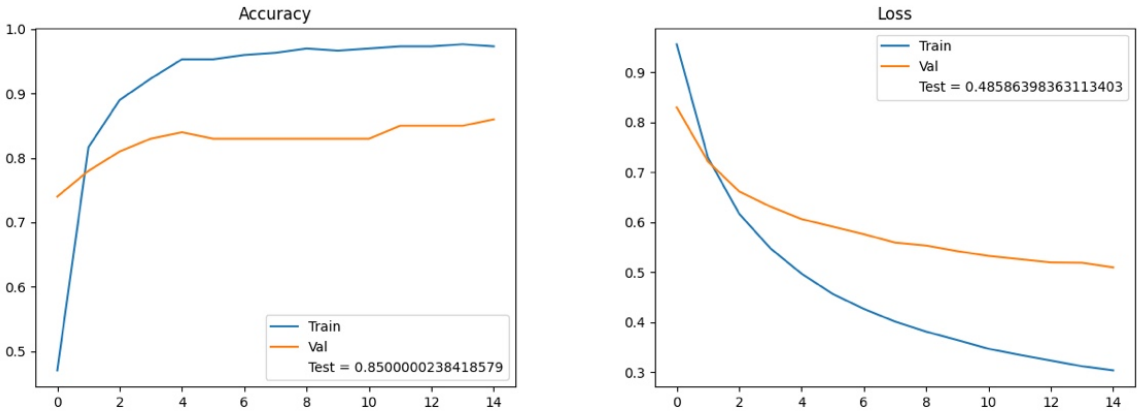



*Рисунок 11.* Графики точности (слева) и функции потерь *CH* (справа)

Результат выбора в качестве функции потерь *KLD* представлен на рисунке 12.

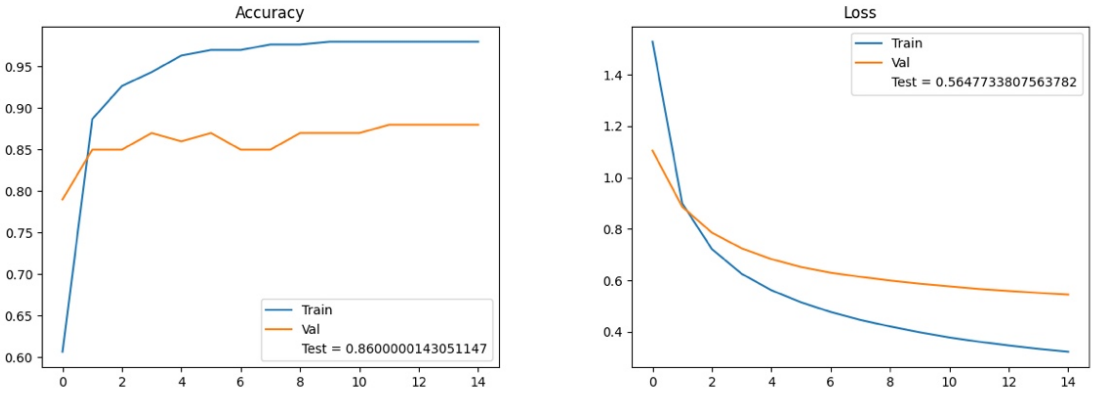



*Рисунок 12.* Графики точности (слева) и функции потерь *KLD* (справа)

В TensorFlow для *KLD* значения, содержащиеся в $y_{true}$ и $y_{pred}$, ограничиваются до диапазона [0, 1] при их несоответствии этому диапазону.

Результаты экспериментов сведены в таблицу 3.

Таблица 3. Результаты экспериментов

| Функция активации между сверточными слоями EfficientNetB0 | Функция потерь | Метрика оценки точности предсказаний, функция активации на предпоследнем полносвязном слое экспериментальной модели | Точность предсказаний на тестовой выборке | Значение функции потерь на тестовой выборке |
|---|---|---|---|---|
| *Swish* | *SCCE* (*from_logits=True*) | *SCA, Linear* | 0,95 | 0,20 |
| *Hard Swish* | | | 0,91 | 0,26 |
| *Mish* | | | 0,87 | 0,54 |
| *GELU* | | | 0,31 | 2,03 |
| *CELU* ($\alpha$ = 0,5) | | | 0,63 | 4,15 |
| *Swish* | *SCCE* (*from_logits=False*) | *SCA, Softmax* | 0,89 | 0,57 |
| | *CCE* (*from_logits=False*) | *CA, Softmax* | 0,79 | 0,56 |
| | *CFCE* ($\alpha$ = 0,25, $\gamma$ = 2, *from_logits=False*) | | 0,88 | 0,05 |
| | *CH* | | 0,85 | 0,49 |
| | *KLD* | | 0,86 | 0,56 |

Таким образом, по результатам всех экспериментов совместное использование функции активации *Swish* между сверточными слоями EfficientNetB0, функции потерь *SCCE* с параметром *from_logits* со значением *True*, метрики оценки точности предсказаний *SCA* и линейной функции активации на предпоследнем полносвязном слое позволяет достичь лучших показателей для данной задачи и нейросетевой архитектуры.

**Заключение.** Выбор функции активации *Swish* и функции потерь *SCCE* позволил достичь наилучших показателей как в первой, так и во второй части экспериментов. При этом для данной архитектуры применять функцию *Softmax* предпочтительнее не на предпоследнем полносвязном слое, а после прохождения данными всех слоев модели, что обеспечивается установкой значения *True* для параметра *from_logits* функции потерь *SCCE*. Выбор функций активации *GELU* и *CELU* с параметром $\alpha$ со значением 0,5 в первой части экспериментов значительно ухудшил показатели соответствующих экспериментальных моделей. Стоит отметить несущественное влияние выбора функции потерь на точность предсказаний во второй части экспериментов.

**Список литературы**

[1] Weon I.-S., Lee S.-G., Ryu J.-K. Object Recognition Based Interpolation With 3D LIDAR and Vision for Autonomous Driving of an Intelligent Vehicle. IEEE Access. 2020;8:65599-65608. DOI:10.1109/ACCESS.2020.2982681.

[2] Eltzner B., Huckemann S. F. A SMEARY CENTRAL LIMIT THEOREM FOR MANIFOLDS WITH APPLICATION TO HIGH-DIMENSIONAL SPHERES. The Annals of Statistics. 2019;47(6):3360-3381. DOI:10.1214/18-AOS1781.

[3] Simonyan K., Zisserman A. Two-Stream Convolutional Networks for Action Recognition in Videos. Proceedings of the 28th International Conference on Neural Information Processing Systems. 2014;1:568-576.

[4] Yedder H. B., Cardoen B., Hamarneh G. Deep learning for biomedical image reconstruction: a survey. Artificial Intelligence Review. 2021;54:215-251. DOI:10.1007/s10462-020-09861-2.

[5] Wang L., Xiong Y., Wang Z., Qiao Y., Lin D., Tang X., Van Gool L. Temporal segment networks: Towards good practices for deep action recognition. Computer Vision – ECCV 2016. 2016;9912:20-36. DOI:10.1007/978-3-319-46484-8_2.

[6] Serre Lab » HMDB: a large human motion database [Electronic resource]. – Mode of access: https://serre-lab.clps.brown.edu/resource/hmdb-a-large-human-motion-database/. – Date of access: 30.01.2025.

[7] CRCV | Center for Research in Computer Vision at the University of Central Florida [Electronic resource]. – Mode of access: https://www.crcv.ucf.edu/data/UCF101.php. – Date of access: 30.01.2025.

[8] GitHub - cvdfoundation/kinetics-dataset [Electronic resource]. – Mode of access: https://github.com/cvdfoundation/kinetics-dataset. – Date of access: 30.01.2025.

[9] Activity Net [Electronic resource]. – Mode of access: http://activity-net.org/. – Date of access: 03.02.2025.

[10] Charades Challenge [Electronic resource]. – Mode of access: http://vuchallenge.org/charades.html. – Date of access: 03.02.2025.

[11] THUMOS Challenge 2015 [Electronic resource]. – Mode of access: http://www.thumos.info/home.html. – Date of access: 05.02.2025.

[12] HACS Dataset [Electronic resource]. – Mode of access: http://hacs.csail.mit.edu/. – Date of access: 07.02.2025.

[13] MOD20 Dataset [Electronic resource]. – Mode of access: https://asankagp.github.io/mod20/. – Date of access: 08.02.2025.

[14] Large-scale Video Classification with Convolutional Neural Networks (CVPR 2014) [Electronic resource]. – Mode of access: https://cs.stanford.edu/people/karpathy/deepvideo/. – Date of access: 08.02.2025.

[15] HAA500 Project Website [Electronic resource]. – Mode of access: https://www.cse.ust.hk/haa/. – Date of access: 10.02.2025.

[16] Load video data | TensorFlow Core [Electronic resource]. – Mode of access: https://www.tensorflow.org/tutorials/load_data/video. – Date of access: 11.02.2025.

[17] Classification Metrics using Sklearn - GeeksforGeeks [Electronic resource]. – Mode of access: https://www.geeksforgeeks.org/sklearn-classification-metrics/. – Date of access: 11.02.2025.

[18] tf.keras.metrics.SparseCategoricalAccuracy | TensorFlow v2.16.1 [Electronic resource]. – Mode of access: https://www.tensorflow.org/api_docs/python/tf/keras/metrics/SparseCategoricalAccuracy. – Date of access: 14.02.2025.

[19] tf.keras.metrics.CategoricalAccuracy | TensorFlow v2.16.1 [Electronic resource]. – Mode of access: https://www.tensorflow.org/api_docs/python/tf/keras/metrics/CategoricalAccuracy. – Date of access: 15.02.2025.

[20] Layer activation functions [Electronic resource]. – Mode of access: https://keras.io/api/layers/activations/. – Date of access: 15.02.2025.

[21] Probabilistic losses [Electronic resource]. – Mode of access: https://keras.io/api/losses/probabilistic_losses/. – Date of access: 16.02.2025.

[22] Hinge losses for "maximum-margin" classification [Electronic resource]. – Mode of access: https://keras.io/api/losses/hinge_losses/. – Date of access: 17.02.2025.

**Авторский вклад**

**Венгеренко Вадим Владимирович** – проведение экспериментов, анализ полученных результатов.

**Лукашевич Марина Михайловна** – руководство исследованием, постановка задачи исследования, формирование структуры статьи.

# SELECTION OF ACTIVATION FUNCTION AND LOSS FUNCTION FOR THE PROBLEM OF HUMAN ACTIVITIES CLASSIFICATION

***V.V. Vengerenko***
*Junior Researcher at the UIIP NASB, Postgraduate Student at the UIIP NASB*

***M.M. Lukashevich***
*Associate Professor of the Department of Information Management Systems of BSU, Senior Researcher of the Laboratory No. 222 of the UIIP NASB PhD of Technical sciences, Associate Professor*

**Abstract.** The purpose of the research is to investigate the influence of the selection of activation function and loss function on the performance of neural network models when solving the problem of classifying human activities.

The results of experiments on assessing the performance of neural network models that classify human activities and use various combinations of activation function and loss function are presented. The experimental studies used a sample from the HMDB51 dataset for 10 classes. A combination of hyperparameters was identified for the considered neural network architecture that allows achieving maximum classification accuracy.

**Keywords**: neural network, activation function, loss function, predictions accuracy, classification.

УДК 004.522

# ОПТИМИЗАЦИЯ IT-ИНФРАСТРУКТУРЫ С ПОМОЩЬЮ BIG DATA: ОТ МОНИТОРИНГА ДО ПРОГНОЗИРОВАНИЯ СБОЕВ

***Аралиев Эсенали***
*Преподаватель кафедры*
*«Программное обеспечение информационных технологий»*
*Институт телекоммуникаций и информатики*
*Туркменистана*
araliyew@icloud.com

***Э.Аралиев***

*Окончил Туркменский государственный университет имени Махтумкули. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.*

**Аннотация.** Статья посвящена применению технологий Big Data для оптимизации IT-инфраструктуры. Рассматриваются методы мониторинга, анализа производительности и прогнозирования сбоев в IT-системах с использованием алгоритмов машинного обучения, кластеризации и анализа временных рядов. Описание методов обработки потоковых данных и адаптивного масштабирования ресурсов позволяет создавать гибкие и эффективные IT-инфраструктуры, что критически важно для устойчивости и бесперебойной работы системы.

**Ключевые слова:** Big Data, мониторинг IT-инфраструктуры, машинное обучение, прогнозирование сбоев, алгоритмы кластеризации, временные ряды, адаптивное масштабирование.

**Введение.** Современные IT-инфраструктуры играют ключевую роль в обеспечении устойчивой работы бизнеса и предоставлении высококачественных услуг пользователям. Однако с ростом объема данных и сложности систем необходимость в эффективном управлении IT-системами становится все более актуальной. Одним из наиболее перспективных направлений оптимизации является использование технологий обработки больших данных (Big Data). Big Data предоставляет возможности для мониторинга, анализа и прогнозирования состояния IT-систем в реальном времени, что позволяет значительно повысить эффективность эксплуатации инфраструктуры и уменьшить вероятность сбоев.

Целью данной статьи является анализ методов и алгоритмов, применяемых для оптимизации IT-инфраструктуры с помощью Big Data, начиная от мониторинга работы системы и заканчивая прогнозированием сбоев. В работе рассматриваются современные подходы, включающие машинное обучение, кластеризацию и методы анализа временных рядов, а также примеры их применения для улучшения производительности и надежности IT-инфраструктуры.

**Методы и алгоритмы мониторинга IT-инфраструктуры с использованием Big Data.** Мониторинг IT-инфраструктуры предполагает непрерывный сбор и анализ данных о ее работе. В традиционных системах мониторинга для этих целей часто использовались простые метрики, такие как загрузка процессора, использование памяти и пропускная способность сети. Однако с развитием Big Data появились новые возможности для более детализированного и точного анализа.

Одним из эффективных методов мониторинга является использование потоковых данных. Потоковая аналитика позволяет обрабатывать данные в реальном времени, что критически важно для оперативного реагирования на изменения в инфраструктуре. Для этого широко применяются алгоритмы машинного обучения и статистического анализа. Например, алгоритм k-Nearest Neighbors (k-NN) используется для классификации состояний системы на основе измерений с различных датчиков и источников данных. Он помогает определить текущий статус инфраструктуры и классифицировать ее состояние, например, как «нормальное», «предсказуемое ухудшение» или «сбой».

Другим важным инструментом для мониторинга является алгоритм кластеризации, такой как k-means. Этот метод позволяет группировать объекты (например, серверы или виртуальные машины) в кластеры на основе схожих характеристик их работы. Кластеры с отклонениями от нормы могут быть быстро обнаружены, что позволяет своевременно выявлять потенциальные проблемы.

Для реализации более сложных методов мониторинга активно используются платформы обработки потоковых данных, такие как Apache Kafka и Apache Flink, которые позволяют интегрировать данные из множества источников и применять различные аналитические модели для их анализа в реальном времени.

**Использование Big Data для анализа производительности системы.** После того как данные собраны и классифицированы, следующим шагом является анализ производительности IT-инфраструктуры. Одним из ключевых аспектов анализа является выявление узких мест, которые могут снижать эффективность работы системы.

Для этого часто применяются методы статистического анализа и кластеризации. Например, с помощью алгоритма k-means можно выделить группы серверов или виртуальных машин, которые имеют схожие проблемы с производительностью. Это может помочь не только в мониторинге состояния системы, но и в прогнозировании потенциальных проблем.

Данные, полученные из различных сенсоров, могут быть обработаны с использованием алгоритмов для оценки использования ресурсов. Например, алгоритм регрессии может быть использован для анализа зависимости между нагрузкой на процессор, память и пропускную способность сети, что позволяет выявлять паттерны использования ресурсов и прогнозировать возможные перегрузки.

Кроме того, для визуализации и представления данных можно использовать методы аналитики, такие как диаграммы контроля, гистограммы и тепловые карты, что позволяет сделать процесс анализа данных более интуитивно понятным и доступным для технических специалистов.

**Прогнозирование сбоев IT-инфраструктуры с использованием методов машинного обучения.** Один из самых мощных инструментов для оптимизации IT-инфраструктуры с помощью Big Data – это прогнозирование сбоев. Сбор и хранение больших объемов данных позволяет создавать предсказательные модели, которые могут предсказывать вероятные сбои на основе анализа исторических данных о работе системы.

Для этой цели широко используются методы машинного обучения, такие как регрессия, нейронные сети и методы временных рядов, например, ARIMA. Эти методы позволяют строить модели, которые предсказывают поведение системы на основе предыдущих наблюдений и текущих данных.

Одним из наиболее эффективных подходов является использование нейронных сетей, которые способны выявлять сложные зависимости в данных и учитывать множество факторов, влияющих на работу системы. Нейронные сети могут быть использованы для прогнозирования сбоев на уровне отдельных серверов или более масштабных проблем, таких как сбои в работе всей инфраструктуры.

**Алгоритмы для адаптивного масштабирования ресурсов на основе Big Data.** С помощью алгоритмов машинного обучения и обработки Big Data можно не только предсказывать сбои, но и оптимизировать использование ресурсов. Адаптивное

масштабирование позволяет динамически изменять ресурсы инфраструктуры в ответ на изменения нагрузки.

Для этого широко применяются алгоритмы кластеризации и обучения с подкреплением. К примеру, алгоритм кластеризации может быть использован для группировки ресурсов по уровням нагрузки, что позволит автоматизировать процессы перераспределения нагрузки между серверами.

Методы обучения с подкреплением могут быть использованы для разработки систем, которые автоматически адаптируют использование ресурсов на основе изменений в рабочих нагрузках, эффективно управляя ресурсами в реальном времени.

**Заключение.** Использование Big Data для мониторинга и прогнозирования сбоев в IT-инфраструктуре позволяет значительно повысить эффективность работы системы, уменьшить вероятность сбоев и оптимизировать ресурсы. Применение алгоритмов машинного обучения, анализа потоковых данных и прогнозирования временных рядов открывает новые возможности для создания гибких и устойчивых IT-инфраструктур. В будущем развитие технологий Big Data и машинного обучения будет способствовать дальнейшему совершенствованию методов управления и оптимизации IT-систем.

## Список литературы

[1] Марц Н, Уоррен Д. Большие данные. Принципы и практика построения масштабируемых систем обработки данных в реальном времени. М.: Вильямс, 2016. – 368 с.

[2] Бенджамин К., Ли М. Прогнозирование сбоев в IT-инфраструктуре с использованием машинного обучения. М.: ТехноЛогика, 2018. – 250 с.

[3] Фрэнк С., Бёрнс Р. Методы анализа и прогнозирования временных рядов. М.: Научный мир, 2017. – 300 с.

## Авторский вклад

**Аралиев Эсенали** – основная концепция и разработка методов мониторинга IT-инфраструктуры, а также предложенные подходы к применению машинного обучения для анализа данных в реальном времени. Разработка и тестирование алгоритмов прогнозирования сбоев с использованием моделей временных рядов и нейронных сетей. Анализ применения Big Data для адаптивного масштабирования IT-ресурсов. Исследование методов кластеризации и алгоритмов анализа производительности системы, а также практическое внедрение предложенных решений для мониторинга и анализа работы IT-инфраструктуры.

# OPTIMIZING IT INFRASTRUCTURE WITH BIG DATA: FROM MONITORING TO FAILURE PREDICTION

***Araliev Esenali***
*Lecturer, Department of*
*“Software for Information Technologies”,*
*Institute of Telecommunications and Informatics of*
*Turkmenistan*

**Abstract.** This article is dedicated to the application of Big Data technologies for optimizing IT infrastructure. The methods for monitoring, performance analysis, and failure prediction in IT systems using machine learning algorithms, clustering, and time series analysis are considered. The description of stream data processing methods and adaptive resource scaling enables the creation of flexible and efficient IT infrastructures, which is critically important for the stability and uninterrupted operation of the system.

**Keywords**: Big Data, IT infrastructure monitoring, machine learning, failure prediction, clustering algorithms, time series, adaptive scaling.

УДК 004.932.2

# ИССЛЕДОВАНИЕ ВЛИЯНИЯ ВХОДНЫХ ПАРАМЕТРОВ НА КЛАССИФИКАЦИЮ ИЗОБРАЖЕНИЙ ПРОДУКТОВ

***А.В. Деркач***
*Магистрант кафедры электронных вычислительных машин БГУИР*
i.angelika897@gmail.com

***И.И. Фролов***
*Доцент кафедры электронных вычислительных машин БГУИР, кандидат технических наук, доцент*
frolov@bsuir.by

***А.В. Деркач***

*Окончила Белорусский государственный университет информатики и радиоэлектроники по специальности «Вычислительные машины, системы и сети». Обучается в магистратуре БГУИР по специальности «Компьютерная инженерия», выполняет исследования в рамках магистерской диссертации на тему «Автоматизированная система анализа продуктов по изображениям для рецептурного подбора».*

***И.И. Фролов***

*Окончил Белорусский государственный университет информатики и радиоэлектроники. Доцент кафедры электронных вычислительных машин БГУИР. Проводит научные исследования в области машинного обучения и компьютерного зрения, участвует в проектировании и разработке систем технического зрения.*

**Аннотация.** Проведен комплексный анализ влияния параметров входных изображений на качество классификации изображений продуктов питания с использованием разных моделей нейронных сетей. Описаны основные этапы подготовки изображений, включая изменение разрешения, добавление гауссовского шума и изменение средней яркости. Проведены эксперименты с различными параметрами входных изображений для оценки их влияния на точность классификации.

**Ключевые слова:** обработка изображений, машинное обучение, нейронные сети, классификация продуктов, качество классификации, параметры входных данных.

**Введение.** Автоматизированная классификация продуктов питания на основе изображений играет важную роль в ряде приложений, включая системы самообслуживания, контроль качества продукции, управление запасами и персонализированные рекомендации. Однако качество классификации во многом зависит от характеристик входных данных [1]. В реальных условиях съёмки изображения могут содержать шумы, варьироваться по разрешению и освещённости, что потенциально снижает эффективность работы нейросетевых моделей, а влияние таких искажений на классификацию часто остаётся недостаточно изученным.

В данном исследовании анализируется, как изменение разрешения, добавление гауссовского шума и коррекция средней яркости изображений влияют на точность работы модели. Эти параметры выбраны, поскольку они моделируют распространённые искажения, возникающие в процессе сбора данных. Полученные результаты позволяют количественно оценить степень деградации классификации и определить, насколько устойчивы современные модели к данным изменениям.

Практическая значимость исследования заключается в возможности учитывать влияние подобных факторов при подготовке обучающих выборок, разработке методов предобработки изображений и создании моделей, менее чувствительных к изменению входных данных. Это особенно важно для автоматизированных систем, где невозможно контролировать условия съёмки, но требуется высокая точность классификации.

**Цель исследования.** Целью проведения исследования является оценка влияния различных характеристик входных изображений на эффективность и точность классификации продуктов в автоматизированной системе анализа изображений продуктов. Эксперимент направлен на выявление ключевых факторов, оказывающих наибольшее влияние на производительность моделей классификации, с последующей настройкой системы для оптимальных условий работы.

Проведено 8 отдельных экспериментов, каждый из которых сосредоточен на одном типе продукта. Это позволяет изолировать влияние исключительно параметров изображений, исключая дополнительные переменные, связанные с различиями между продуктами. Такой подход обеспечивает точные и репрезентативные результаты, поскольку в каждом эксперименте будут варьироваться только параметры входных изображений при неизменном объекте классификации.

Для анализа будут использованы три различных архитектуры нейронных сетей для обнаружения объектов (object detection), что позволит оценить, как изменения параметров изображений влияют на результаты классификации на разных моделях. Каждая модель получает одинаковые изображения с одинаковыми параметрами, что позволит более полно исследовать воздействие характеристик изображений на точность классификации, а также поможет определить оптимальные условия для работы системы в реальных сценариях. Однако сравнение самих алгоритмов не является целью исследования – основное внимание сосредоточено на анализе параметров изображений, оказывающих влияние на результаты классификации.

Кроме того, исследование позволит оценить устойчивость моделей к различным условиям съёмки изображений, таким как шумы, изменения яркости и качество. Результаты помогут в дальнейшей оптимизации автоматизированных систем анализа изображений, а также могут быть полезными для практического применения в области распознавания и классификации изображений продуктов питания в реальных условиях.

**Описание исследования.** В рамках эксперимента будут исследоваться три ключевых фактора, влияющих на качество классификации изображений:

1 *Разрешение изображений*. Позволяет оценить, насколько детализация изображения влияет на точность классификации. В рамках эксперимента будут рассмотрены три уровня разрешения: 128×128, 256×256, 512×512 пикселей.

2 *Гауссовский шум (стандартное отклонение шума)*. Исследуется влияние добавления случайных вариаций пикселей на способность модели классифицировать объекты. Гауссовский шум был выбран для моделирования реальных искажений изображений, так как он имитирует случайные вариации пикселей, характерные для условий съемки, например,
из-за недостаточной освещенности или качества сенсоров камеры. Для этого используются три уровня шума, заданные стандартным отклонением (σ): 0 (без шума), 30 (умеренный шум), 60 (высокий шум).

3 *Освещенность изображения (средний уровень яркости)*. Влияет на способность модели распознавать объекты при разной интенсивности освещения. В качестве исследуемых значений примем три уровня яркости, измеряемые в среднем значении интенсивности пикселей (от 0 до 255 в 8-битном представлении): 30 (тёмное изображение), 130 (естественное освещение), 210 (очень яркое изображение).

Для исследования влияния параметров изображений на качество классификации будут использованы три нейросетевых модели. Использование нескольких алгоритмов

позволит нивелировать зависимость результатов от специфики одной конкретной модели и выявить универсальные закономерности, влияющие на качество классификации. Модели, применяемые в эксперименте:

1 YOLOv8 (You Only Look Once) – модель для детекции объектов в реальном времени, использующая подход одноэтапной обработки [2]. YOLO делит изображение на сетку и прогнозирует координаты объектов и классы объектов для каждого сегмента. Особенностью модели является высокая скорость работы при относительно хорошем качестве детекции, что делает ее популярной для задач, требующих быстрой обработки.

2 DETR (DEtection TRansformer) – трансформер-архитектура, разработанная Facebook [3]. DETR применяет трансформеры для обработки изображений и выявления объектов без необходимости в традиционных этапах постобработки, таких как non-maximum suppression (NMS). Это позволяет модели быть более гибкой и эффективной в сложных сценах с множеством объектов и разнообразными контекстами, но может иметь большую вычислительную нагрузку по сравнению с более традиционными методами.

3 RetinaNet – модель, основанная на архитектуре ResNet-50 с использованием Feature Pyramid Networks (FPN) для улучшенной обработки признаков на разных уровнях масштаба [4]. RetinaNet использует механизм потерь focal loss, который помогает бороться с дисбалансом между большим количеством фонов и меньшим количеством объектов на изображении. Эта модель подходит для детекции объектов в условиях, когда важно правильно классифицировать малые или трудные для распознавания объекты.

Эксперимент будет проводиться с использованием 8 различных изображений (рис. 1) для каждой комбинации факторов.

*Рисунок 1.* Изображения для исследования

Это позволит учесть естественную вариативность данных и рассчитать усреднённые показатели качества классификации, минимизируя влияние случайных факторов. Для большего разнообразия в эксперимент были выбраны 5 изображений обычных продуктов и 3 изображения готовых блюд. Такой выбор обеспечит более определённую картину, отражающую как распознавание простых объектов, так и сложных составных объектов.

Изображения были выбраны высокого качества, с равными размерами по высоте и ширине (1000×1000 пикселей), чтобы исключить влияние исходного разрешения на результат. Эти изображения были выбраны, поскольку на них все модели в первоначальной обработке продемонстрировали отличные результаты, с вероятностью предсказания объектов более 75%.

На изображениях, где присутствуют несколько одинаковых объектов, будет учитываться только наиболее уверенное предсказание модели, класс которого совпадает с ожидаемым. В случае нескольких объектов одного типа выбирается тот, у которого вероятность предсказания наиболее высока.

Ниже приводится диаграмма предсказания моделей без предварительной обработки исходных изображений:

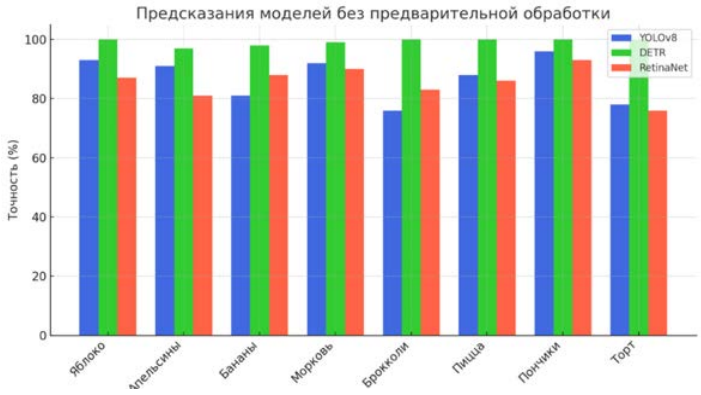



*Рисунок 2.* Результаты экспериментов для исходных изображений разрешением 1000×1000

Для каждой из моделей (YOLOv8, DETR и RetinaNet) использовались предобученные веса, что позволило сразу приступить к оценке их работы без необходимости в проведении нескольких итераций на одном изображении. Это значительно ускоряет процесс тестирования и исключает дополнительную тренировку моделей для конкретного эксперимента.

**Изменение разрешения изображения.** Для изменения разрешения изображений использовалась библиотека PIL (Python Imaging Library). Каждое изображение было масштабировано до трёх заранее заданных разрешений: 128×128, 256×256 и 512×512 пикселей. Масштабирование выполнялось с помощью метода *resize()* из PIL, который изменяет размер изображения до целевого, сохраняя пропорции. Все изображения имели одинаковые размеры в пикселях, что обеспечивало их единообразную обработку в ходе эксперимента.

**Изменение средней яркости изображения.** Для изменения яркости изображений в эксперименте использовался метод коррекции с помощью библиотеки PIL (Python Imaging Library). Яркость изображения оценивалась как среднее значение интенсивности пикселей, вычисляемое по всем цветовым каналам (RGB). Для этого изображение сначала преобразовывалось в массив данных с помощью библиотеки NumPy, и затем рассчитывалась средняя яркость всех пикселей. После этого применялась коррекция яркости с целью достижения заданного значения.

Для каждого изображения был выбран целевой уровень яркости (30, 130 или 210), который представлял собой соответствующие уровни: тёмное изображение, естественное освещение и яркое изображение. Чтобы привести изображение к заданной яркости, рассчитывался коэффициент коррекции, который представлял собой отношение целевой яркости к текущей средней яркости изображения. Этот коэффициент использовался для масштабирования исходного изображения с помощью метода *ImageEnhance.Brightness()*, что позволяло изменить яркость изображения на нужный уровень. Таким образом, изображения были приведены к заданной яркости, при этом сохранялась естественная цветовая насыщенность.

**Наложение гауссовского шума.** Для наложения гауссовского шума на изображения использовалась библиотека NumPy для работы с массивами данных и PIL для работы с изображениями. Сначала изображение было загружено с помощью библиотеки PIL, и оно преобразовывалось в массив с помощью *np.array()* с типом данных *np.float32* для точности операций. Затем для каждого пикселя изображения генерировался гауссовский шум с помощью функции *np.random.normal()*, где в качестве параметров задавались: среднее значение шума (0) и стандартное отклонение (σ), которое определяет интенсивность шума. Этот шум добавлялся к исходному массиву изображения для всех трех цветовых каналов (красного, зеленого и синего), что приводило к изменению цветовых значений пикселей.

После добавления шума, чтобы избежать выхода значений за пределы допустимого диапазона [0, 255] для изображений в формате RGB, использовалась функция *np.clip()*,

которая ограничивает значения в пределах этого диапазона. Далее массив преобразовывался обратно в изображение с помощью функции *Image.fromarray()*, и результат сохранялся как новое изображение.

Пример измененных изображений представлен на рисунке 2. На картинке слева – изображение размером 512×512 с средней яркостью пикселей 30 и отклонением гауссовского шума равным 30, по центру – изображение размером 512×512 с средней яркостью пикселей 130 и отклонением гауссовского шума равным 60, справа – изображение размером 512×512 с средней яркостью пикселей 210 и отклонением гауссовского шума равным 0.

*Рисунок 3.* Пример обработанных входных изображений

**Применение моделей для детекции объектов.** Для детекции объектов использовались три модели с предобученными весами:

1 YOLOv8. Модель была загружена через библиотеку Ultralytics YOLO с использованием предобученных весов yolov8n.pt. Для обработки изображений использовалась функция *show()*, которая отображала детектированные объекты с рамками и текстовыми метками.

2 DETR. Модель загружена через библиотеку Transformers с использованием предобученных весов *facebook/detr-resnet-50*. Для предсказания использовались тензоры PyTorch [5], а выходные данные модели использовались для дальнейшей обработки и получения ограничивающих рамок объектов на изображениях.

3 RetinaNet. Модель была загружена с предобученными весами *RetinaNet_ResNet50_FPN_Weights.DEFAULT* через библиотеку TorchVision. Эти веса были обучены на наборе данных COCO и содержат информацию о категориях объектов, что позволяет модели классифицировать и локализовать объекты на изображениях.

**Результаты проведения исследования.** В таблице приведены результаты тестирования трех моделей детекции объектов на изображениях с различными параметрами. В столбце 6 указаны значения предсказаний моделей для 8 изображений, расположенных в том же порядке, как на рисунке 1. Процентные значения в этом столбце отражают степень успешности детекции объектов для каждого изображения. В случае, если модель не распознала объект, указано значение 0%.

| Разрешение | σ | Средняя яркость [0-255] | Модель | Яблоки | Апельсин | Банан | Морковь | Брокколи | Пицца | Пончик | Торт |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 128×128 | 0 | 30 | DETR | 1 | 0.99 | 0.96 | 0 | 0 | 0.94 | 0 | 0 |
| 128×128 | 0 | 130 | DETR | 0.99 | 0.99 | 0.96 | 0.98 | 0.99 | 0.99 | 1 | 0 |
| 128×128 | 0 | 210 | DETR | 0.94 | 0.91 | 0.97 | 0.58 | 1 | 0.98 | 1 | 0.84 |
| 128×128 | 30 | 30 | DETR | 0 | 0 | 0.9 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 30 | 130 | DETR | 0 | 0.72 | 0.98 | 0.91 | 0 | 0.98 | 0 | 0 |
| 128×128 | 30 | 210 | DETR | 0 | 0.9 | 0.97 | 0 | 0 | 0.53 | 0 | 0 |
| 128×128 | 60 | 30 | DETR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 130 | DETR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 210 | DETR | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 0 | 30 | RetinaNet | 0.7 | 0.47 | 0.75 | 0.06 | 0.13 | 0.82 | 0.31 | 0.05 |
| 128×128 | 0 | 130 | RetinaNet | 0.87 | 0.7 | 0.88 | 0.76 | 0.56 | 0.85 | 0.71 | 0.13 |
| 128×128 | 0 | 210 | RetinaNet | 0.86 | 0.72 | 0.79 | 0.34 | 0.79 | 0.76 | 0.73 | 0.26 |
| 128×128 | 30 | 30 | RetinaNet | 0.09 | 0.21 | 0.13 | 0 | 0.07 | 0.17 | 0.12 | 0 |
| 128×128 | 30 | 130 | RetinaNet | 0.12 | 0.61 | 0.44 | 0.25 | 0.6 | 0.74 | 0.37 | 0 |
| 128×128 | 30 | 210 | RetinaNet | 0.2 | 0.59 | 0.68 | 0.07 | 0.66 | 0.66 | 0.4 | 0 |
| 128×128 | 60 | 30 | RetinaNet | 0.09 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 130 | RetinaNet | 0.06 | 0.28 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 210 | RetinaNet | 0 | 0.4 | 0.15 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 0 | 30 | YOLOv8 | 0 | 0 | 0.67 | 0 | 0 | 0.61 | 0 | 0 |
| 128×128 | 0 | 130 | YOLOv8 | 0.63 | 0.81 | 0.66 | 0.36 | 0 | 0.7 | 0.39 | 0 |
| 128×128 | 0 | 210 | YOLOv8 | 0 | 0.75 | 0.64 | 0 | 0 | 0.42 | 0.76 | 0 |
| 128×128 | 30 | 30 | YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 30 | 130 | YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 30 | 210 | YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 30 | YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 130 | YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 128×128 | 60 | 210 | YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

*Рисунок 4.* Результаты экспериментов для изображений с разрешением 128×128

| Разрешение | σ | Средняя яркость [0-255] | Модель | Яблоко | Апельсин | Банан | Морковь | Брокколи | Пицца | Пончик | Торт |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 256×256 | 0 | | 30 YOLOv8 | 0.9 | 0.53 | 0.72 | 0.61 | 0 | 0.9 | 0 | 0 |
| 256×256 | 0 | | 30 DETR | 1 | 1 | 0.98 | 0.94 | 0 | 0.98 | 0 | 0 |
| 256×256 | 0 | | 30 RetinaNet | 0.84 | 0.68 | 0.82 | 0.18 | 0.2 | 0.85 | 0.33 | 0.25 |
| 256×256 | 0 | | 130 YOLOv8 | 0.94 | 0.83 | 0.67 | 0.89 | 0.72 | 0.85 | 0.54 | 0.61 |
| 256×256 | 0 | | 130 DETR | 0.98 | 0.97 | 0.97 | 0.99 | 1 | 1 | 0.99 | 0.86 |
| 256×256 | 0 | | 130 RetinaNet | 0.88 | 0.82 | 0.89 | 0.89 | 0.87 | 0.85 | 0.9 | 0.71 |
| 256×256 | 0 | | 210 YOLOv8 | 0.93 | 0.86 | 0.71 | 0.41 | 0.74 | 0.8 | 0.85 | 0.32 |
| 256×256 | 0 | | 210 DETR | 1 | 0.99 | 0.98 | 0.98 | 1 | 1 | 0.98 | 0.68 |
| 256×256 | 0 | | 210 RetinaNet | 0.89 | 0.78 | 0.83 | 0.74 | 0.87 | 0.75 | 0.92 | 0.33 |
| 256×256 | 30 | | 30 YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 30 | | 30 DETR | 1 | 0.99 | 1 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 30 | | 30 RetinaNet | 0.56 | 0.59 | 0.6 | 0.08 | 0.07 | 0.49 | 0.33 | 0.1 |
| 256×256 | 30 | | 130 YOLOv8 | 0 | 0 | 0.6 | 0 | 0 | 0.5 | 0.31 | 0 |
| 256×256 | 30 | | 130 DETR | 1 | 0.99 | 0.99 | 0.87 | 0.99 | 1 | 1 | 0 |
| 256×256 | 30 | | 130 RetinaNet | 0.58 | 0.69 | 0.83 | 0.46 | 0.73 | 0.8 | 0.75 | 0.15 |
| 256×256 | 30 | | 210 YOLOv8 | 0 | 0 | 0.83 | 0 | 0 | 0.39 | 0 | 0 |
| 256×256 | 30 | | 210 DETR | 0.98 | 0.94 | 0.96 | 0 | 0.6 | 0 | 0 | 0.83 |
| 256×256 | 30 | | 210 RetinaNet | 0.29 | 0.71 | 0.82 | 0.07 | 0.61 | 0.48 | 0.67 | 0.17 |
| 256×256 | 60 | | 30 YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 60 | | 30 DETR | 0 | 0 | 0.98 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 60 | | 30 RetinaNet | 0.16 | 0.47 | 0.23 | 0 | 0 | 0.34 | 0.25 | 0 |
| 256×256 | 60 | | 130 YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 60 | | 130 DETR | 0 | 0.98 | 0.96 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 60 | | 130 RetinaNet | 0.12 | 0.42 | 0.26 | 0.09 | 0.13 | 0.14 | 0.16 | 0.1 |
| 256×256 | 60 | | 210 YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 60 | | 210 DETR | 0 | 0.97 | 0.91 | 0 | 0 | 0 | 0 | 0 |
| 256×256 | 60 | | 210 RetinaNet | 0.14 | 0.52 | 0.42 | 0 | 0.06 | 0.25 | 0.06 | 0.06 |

*Рисунок 5.* Результаты экспериментов для изображений с разрешением 256×256

| Разрешение | σ | Средняя яркость [0-255] | Модель | Яблоко | Апельсин | Банан | Морковь | Брокколи | Пицца | Пончик | Торт |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 512×512 | 0 | | 30 YOLOv8 | 0.92 | 0.69 | 0.65 | 0.89 | 0.64 | 0.85 | 0 | 0.64 |
| 512×512 | 0 | | 30 DETR | 1 | 0.99 | 0.95 | 0.98 | 1 | 1 | 0 | 0 |
| 512×512 | 0 | | 30 RetinaNet | 0.82 | 0.69 | 0.89 | 0.55 | 0.68 | 0.89 | 0.5 | 0.38 |
| 512×512 | 0 | | 130 YOLOv8 | 0.93 | 0.89 | 0.72 | 0.88 | 0.66 | 0.87 | 0.87 | 0.7 |
| 512×512 | 0 | | 130 DETR | 1 | 0.93 | 0.96 | 0.99 | 1 | 1 | 1 | 1 |
| 512×512 | 0 | | 130 RetinaNet | 0.84 | 0.82 | 0.88 | 0.88 | 0.84 | 0.87 | 0.92 | 0.68 |
| 512×512 | 0 | | 210 YOLOv8 | 0.93 | 0.87 | 0.75 | 0.78 | 0.74 | 0.8 | 0.94 | 0.52 |
| 512×512 | 0 | | 210 DETR | 1 | 0.98 | 0.98 | 1 | 1 | 1 | 1 | 0.99 |
| 512×512 | 0 | | 210 RetinaNet | 0.86 | 0.77 | 0.87 | 0.68 | 0.88 | 0.84 | 0.92 | 0.44 |
| 512×512 | 30 | | 30 YOLOv8 | 0.73 | 0.49 | 0.61 | 0.26 | 0.4 | 0.85 | 0 | 0 |
| 512×512 | 30 | | 30 DETR | 1 | 1 | 0.98 | 0.54 | 0.99 | 1 | 0 | 0 |
| 512×512 | 30 | | 30 RetinaNet | 0.67 | 0.62 | 0.86 | 0.14 | 0.45 | 0.87 | 0.31 | 0.1 |
| 512×512 | 30 | | 130 YOLOv8 | 0.54 | 0.62 | 0.79 | 0.7 | 0.74 | 0.75 | 0.45 | 0 |
| 512×512 | 30 | | 130 DETR | 1 | 0.94 | 0.98 | 0 | 0.94 | 1 | 1 | 0.86 |
| 512×512 | 30 | | 130 RetinaNet | 0.84 | 0.71 | 0.86 | 0.6 | 0.81 | 0.89 | 0.88 | 0.22 |
| 512×512 | 30 | | 210 YOLOv8 | 0.31 | 0.72 | 0.83 | 0 | 0.26 | 0.78 | 0.8 | 0 |
| 512×512 | 30 | | 210 DETR | 1 | 0.94 | 0.98 | 0 | 0.9 | 0.96 | 1 | 0.99 |
| 512×512 | 30 | | 210 RetinaNet | 0.75 | 0.73 | 0.84 | 0.09 | 0.83 | 0.76 | 0.9 | 0.2 |
| 512×512 | 60 | | 30 YOLOv8 | 0 | 0 | 0 | 0 | 0 | 0.33 | 0 | 0 |
| 512×512 | 60 | | 30 DETR | 0.93 | 0.97 | 1 | 0 | 0 | 0.97 | 0 | 0 |
| 512×512 | 60 | | 30 RetinaNet | 0.51 | 0.54 | 0.73 | 0.07 | 0.07 | 0.42 | 0.27 | 0.08 |
| 512×512 | 60 | | 130 YOLOv8 | 0 | 0 | 0 | 0.26 | 0 | 0 | 0 | 0 |
| 512×512 | 60 | | 130 DETR | 0 | 0.93 | 1 | 0 | 0.65 | 0 | 0.92 | 0 |
| 512×512 | 60 | | 130 RetinaNet | 0.43 | 0.58 | 0.8 | 0.13 | 0.65 | 0.35 | 0.57 | 0.12 |
| 512×512 | 60 | | 210 YOLOv8 | 0 | 0 | 0 | 0 | 0.31 | 0 | 0 | 0 |
| 512×512 | 60 | | 210 DETR | 0 | 0.96 | 1 | 0 | 0 | 0 | 0.99 | 0.97 |
| 512×512 | 60 | | 210 RetinaNet | 0.29 | 0.58 | 0.81 | 0.08 | 0.52 | 0.31 | 0.55 | 0.07 |

*Рисунок 6.* Результаты экспериментов для изображений с разрешением 512×512

**Выводы**. На основе полученных данных из таблицы 1 можно сделать выводы о влиянии различных параметров изображений на эффективность классификации продуктов. Эксперимент проводился с учетом нескольких факторов, включая разрешение изображения, уровень шума и освещенность, которые имели заметное влияние на результаты работы моделей.

1 *Влияние разрешения изображения*. Модели показали улучшение результатов с повышением разрешения, что указывает на важность высококачественных изображений для точной классификации продуктов. Однако стоит отметить, что подача сверхвысококачественных изображений не всегда оправдана с точки зрения вычислительных ресурсов.

2 *Влияние шума*. Все модели показали снижение точности классификации при увеличении уровня шума в изображениях. Особенно это было заметно для модели YOLOv8, которая показала значительное падение точности при сильном уровне шума ($\sigma = 60$). Модели DETR и RetinaNet оказались более устойчивыми к шуму, сохраняя более высокие значения точности даже при сильных искажениях изображения. Это указывает на то, что некоторые модели могут быть более подвержены шуму, в то время как другие лучше справляются с его присутствием в данных.

3 *Влияние освещенности*. Освещенность изображений также оказала влияние на точность классификации, причем эффект был различен для каждой модели. Модели продемонстрировали разные уровни точности в зависимости от изменений освещенности, где в некоторых случаях повышение яркости способствовало улучшению точности (например, для модели DETR), в то время как другие модели (например, RetinaNet) показали стабильность даже при изменении яркости. Это может свидетельствовать о том,

что оптимизация под изменения освещенности также важна для достижения высоких показателей классификации, но должна быть адаптирована под конкретную модель.

Хотя основная цель эксперимента не заключалась в сравнении моделей, можно отметить, что модель DETR показала наибольшую устойчивость к изменениям освещенности и шума, достигая лучших результатов на части изображений с низким качеством. Модель YOLOv8, в свою очередь, продемонстрировала высокую точность на изображениях с высоким разрешением, но была более чувствительна к шуму.

## Список литературы

[1] Dodge S., Karam L. Understanding how image quality affects deep neural networks. 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). 2016; 1-6. DOI: 10.1109/QoMEX.2016.7498955.

[2] Mupparaju S., Thotakura S. R., Ch. V. R. Reddy. A Review on YOLOv8 and Its Advancements. Cryptology and Network Security with Machine Learning. 2024; 529-545. DOI: 10.1007/978-981-99-7962-2_39.

[3] Carion N., Massa F., Synnaeve G., Usunier N. End-to-End Object Detection with Transformers. In: Computer Vision – ECCV 2020, 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I. Lecture Notes in Computer Science. 2020; 213-229. DOI: 10.1007/978-3-030-58452-8_13.

[4] Al-Hasanat M., Alsafasfeh M., Alhasanat A., Althunibat S. RetinaNet-Based Approach for Object Detection and Distance Estimation in an Image. International Journal on Communications Antenna and Propagation (IRECAP). 2021; 11(1):19. DOI: 10.15866/irecap.v11i1.19341.

[5] Фролов, И. И. Эффективные библиотеки машинного обучения / И. И. Фролов // Технические средства защиты информации : тезисы докладов XVI Белорусско-российской научно – технической конференции, Минск, 5 июня 2018 г. – Минск: БГУИР, 2017. – С. 94.

## Авторский вклад

**Деркач Анжелика Валерьевна** – проведение исследования, анализ результатов исследования, формирование структуры статьи.

**Фролов Игорь Иванович** – руководство исследованием, рецензирование статьи.

# INVESTIGATION OF THE INFLUENCE OF INPUT IMAGE PARAMETERS ON PRODUCT CLASSIFICATION

***A.V. Dziarkach***
*Master's student of the Department of Electronic Computing Machines of BSUIR*

***I.I. Frolov***
*Associate Professor of the Department of Electronic Computing Machines of BSUIR, PhD of Technical Sciences, Associate Professor*

**Abstract.** A comprehensive analysis of the influence of input image parameters on the quality of food classification using different neural network models has been carried out. The main stages of image preparation are highlighted, including changing the resolution, adding Gaussian noise, and changing the average brightness. Experiments have been conducted with various parameters of the input images to assess their effect on classification accuracy.

**Keywords**: image processing, machine learning, neural networks, product classification, classification quality, input data parameters.

УДК 004.65:656.131.7

# ПРИМЕНЕНИЕ БОЛЬШИХ ДАННЫХ В СФЕРЕ ПРОКАТА АВТОМОБИЛЕЙ

***Е.М. Косарева***
*Ассистент кафедры проектирования информационно-компьютерных систем,*
*БГУИР*
kksrvvv@gmail.com

***А.И. Полуян***
*Студент факультета компьютерного проектирования*
*БГУИР*
toshapoluyan@gmail.com

***Е.М. Косарева***

*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием автоматизации систем распознавания потенциальных угроз в рамках «умного» города, организацией учебного и научно-исследовательского процессов в техническом университете.*

***А.И. Полуян***

*Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с исследованием возможностей развития облачных и серверных хранилищ в сфере услуг.*

**Аннотация.** Проведен анализ методов и алгоритмов применения больших данных в сфере проката автомобилей, способствующих оптимизации управления автопарком, динамическому ценообразованию и улучшению качества клиентского сервиса. Показано, что интеграция телематических данных, информации с IoT-устройств и аналитических систем на базе облачных технологий позволяет формировать прогностические модели и проводить оперативное управление инфраструктурой компании.

Проведена оценка эффективности использования облачных сервисов для обработки больших объёмов данных.

**Ключевые слова:** большие данные, прокат автомобилей, цифровая трансформация, IoT, телематика, облачные технологии, аналитика данных.

**Введение.** Цифровая трансформация охватила практически все отрасли экономики, и сфера проката автомобилей не является исключением. Большие данные (Big Data) стали фундаментальным инструментом для оптимизации бизнес-процессов, повышения качества обслуживания клиентов и эффективного управления автопарком. С каждым годом объём информации, собираемой с помощью IoT-устройств, мобильных приложений, телематических систем и социальных сетей, растёт экспоненциально. Современные компании проката автомобилей всё активнее внедряют аналитические системы для обработки этих данных, что позволяет им не только прогнозировать потребности клиентов, но и оперативно реагировать на изменения рыночной конъюнктуры.

**Анализ работы больших данных в сфере проката автомобилей.** Сфера проката автомобилей генерирует огромные объёмы информации, охватывающей как технические показатели состояния автопарка, так и поведение клиентов. Основными источниками данных являются:

– **телематические данные**: информация с GPS-трекеров, датчиков состояния автомобиля, диагностические отчёты, данные о маршрутах и пробеге.

– **данные бронирования и пользовательских приложений**: сведения о времени аренды, повторных заказах, региональных особенностях спроса.

– **социальные сети и отзывы**: фидбэк пользователей, оценки сервиса и комментарии, отражающие уровень удовлетворённости и потенциальные направления для улучшения качества обслуживания.

Рассмотрим ключевые направления применения аналитики больших данных в прокате автомобилей:

– **оптимизация управления автопарком:** анализ состояния транспортных средств в режиме реального времени позволяет выявлять потребности в профилактическом обслуживании, прогнозировать технические неисправности и снижать вероятность внеплановых простоев. На основе анализа телематических данных можно разработать алгоритмы предиктивного обслуживания, что приводит к снижению затрат и увеличению срока эксплуатации техники.

– **динамическое ценообразование и прогнозирование спроса:** системы, основанные на аналитике больших данных, учитывают сезонные колебания, региональные особенности спроса и даже погодные условия. Это позволяет устанавливать оптимальные цены, стимулирующие спрос в периоды низкой загрузки, и, наоборот, корректировать тарифы при всплеске активности.

– **персонализация сервисов и улучшение клиентского опыта**: сбор и обработка информации о предпочтениях клиентов помогает создавать индивидуальные предложения и программы лояльности. Анализ истории заказов и отзывов позволяет предлагать релевантные услуги, тем самым повышая удовлетворённость клиентов и их приверженность бренду.

Использование алгоритмов машинного обучения и статистической обработки данных позволяет не только анализировать исторические тенденции, но и строить прогнозы, что даёт компаниям конкурентное преимущество на рынке.

**Достижение высокой результативности с помощью облачного хранения.** Обработка больших объёмов информации требует серьёзной вычислительной мощности, масштабируемого хранилища и возможности проведения аналитических расчётов в режиме реального времени. Здесь на помощь приходят облачные технологии, которые обеспечивают:

– масштабируемость, облачные инфраструктуры, такие как Microsoft Azure, Amazon Web Services и Google Cloud Platform, позволяют динамически увеличивать вычислительные ресурсы по мере роста объёмов данных. Это избавляет компании от необходимости инвестировать в дорогостоящее оборудование и обслуживать собственные серверные фермы.

– эффективность и гибкость, использование облачных сервисов позволяет интегрировать аналитические платформы (например, Apache Hadoop, Apache Spark) с различными источниками данных. Это обеспечивает быстрый поток обработки информации, а также возможность параллельного анализа данных, что существенно ускоряет принятие управленческих решений.

– снижение затрат, модель оплаты по факту использования (pay-as-you-go) позволяет оптимизировать расходы компании, оплачивая только те ресурсы, которые действительно задействованы в момент обработки данных. Это особенно важно для сезонных бизнесов, где потребность в вычислительных ресурсах может существенно меняться.

Ниже приведена сравнительная таблица 1 характеристик локальных серверов и облачных решений:

Таблица 1. Процент потерянных сообщений

| Характеристика | Локальные серверы | Облачное хранение и сервера |
|---|---|---|
| Масштабируемость | Ограниченная, требует дополнительных инвестиций | Динамическая, масштабируется в режиме реального времени |

| | | |
|---|---|---|
| Стоимость | Высокие капитальные затраты | Оплата по факту использования |
| Обслуживание и обновления | Требуют собственных IT-ресурсов | Обслуживание включено в сервисный пакет |
| Интеграция аналитических платформ | Сложная интеграция, необходимость настройки | Широкий спектр интеграций и готовых решений |

Применение облачных технологий в сочетании с современными методами обработки больших данных позволяет создать эффективную инфраструктуру, готовую к оперативному реагированию на изменения в рыночной динамике и обеспечению высокого уровня обслуживания клиентов.

**Перспективы развития и вызовы.** Внедрение больших данных в сферу проката автомобилей открывает широкие перспективы для дальнейших инноваций. Среди них можно выделить:

– расширение возможностей персонализации, сбор данных о привычках и предпочтениях клиентов позволит разрабатывать гиперперсонализированные предложения, активнее использовать продвинутые алгоритмы машинного обучения и искусственного интеллекта для улучшения клиентского опыта.

– интеграция с другими цифровыми экосистемами, совмещение данных проката с информацией о городской инфраструктуре, дорожном движении и даже метеорологическими данными создаёт возможности для синергетического развития умных городов, где транспортные системы будут максимально автоматизированы и интегрированы.

Однако наряду с перспективами развитие технологий сопровождается и определёнными вызовами. Необходимость быстро адаптироваться к изменяющимся требованиям рынка, сложность интеграции с устаревшими системами и высокий порог входа в сферу высокотехнологичной аналитики – всё это требует от компаний постоянного обучения и изменения бизнес-моделей.

**Заключение.** Применение больших данных в сфере проката автомобилей демонстрирует, как современные технологии способны преобразить традиционные бизнес-процессы. Анализ телематических, клиентских и операционных данных позволяет повысить эффективность управления автопарком, оптимизировать логистику, внедрять динамическое ценообразование и значительно улучшать клиентский сервис. Облачные технологии и серверные платформы создают надёжную IT-инфраструктуру, которая способна обрабатывать огромные объёмы информации в режиме реального времени, обеспечивая скорость и точность аналитики.

Несмотря на существующие вызовы, связанные с безопасностью данных и необходимостью интеграции с устаревшими системами, преимущества внедрения Big Data-решений очевидны. Компании, способные эффективно использовать современные инструменты анализа, получают возможность принимать своевременные управленческие решения, адаптироваться к изменяющимся рыночным условиям и опережать конкурентов. В перспективе дальнейшее развитие технологий, интеграция с автономными системами и повышение персонализации услуг будут способствовать устойчивому развитию всей индустрии проката автомобилей.

## Список литературы

[1] Иванов И.И., Петров А.С. Применение больших данных в транспортной логистике. Журнал «Информационные технологии в сфере транспорта», 2020, №12(3). – с. 45-58.

[2] Сидоров В.А., Облачные технологии в аналитике больших данных. Издательство «Цифровой мир», 2019. – с 32-46.

[3] Provost F. Fawcett T., Data Science for Business: What You Need Know about Data Mining and Data-Analytic Thinking. O'Reilly Media, 2013. – 414 с.

**Авторский вклад**

**Косарева Екатерина Максимовна** – руководство исследованием и постановка задачи по изучению применения больших данных в сфере проката автомобилей.

**Полуян Антон Игоревич** – описание принципа работы больших данных, облачных и серверных хранилищ, анализ полученных результатов, перспективы развития применеия больших данных в сфере проката автомобилей.

# APPLICATION OF BIF DATA IN THE CAR RENTAL INDUSTRY

***E.M. Kosareva***
*Assistant of the department of information and computer design*

***A.I. Poluyan***
*Student of the Faculty of Computer Aided Design*

**Abstract.** An analysis of the methods and algorithms for applying Big Data in the car rental industry has been conducted, which contributes to optimizing fleet management, dynamic pricing, and enhancing the quality of customer service. It is demonstrated that integrating telematics data, information from IoT devices, and analytical systems based on cloud technologies enables the development of predictive models and facilitates real-time operational management of the company's infrastructure.

An assessment of the effectiveness of using cloud services to process large volumes of data has been carried out.

**Keywords**: Big Data, car rental, digital transformation, IoT, telematics, cloud technologies, data analytics.

УДК 621.391 + 621.395

# АКТУАЛЬНОСТЬ ПРОБЛЕМЫ УВЕЛИЧЕНИЯ ПРОПУСКНОЙ СПОСОБНОСТИ ВОЛС В СВЯЗИ С РОСТОМ ОБЪЕМА ПЕРЕДАВАЕМЫХ ДАННЫХ

***С. Тилевов***
*Заведующий кафедры «Телекоммуникационные системы»*
*Институт Телекоммуникаций и информатики Туркменистана,*
tilewow.tm@mail.ru

***С. Тилевов***
*Окончил Туркменский государственный университет. Область научных интересов связана с исследование пропускной способности волоконно-оптических линий связи*

**Аннотация:** Проблема увеличения пропускной способности ВОЛС является крайне актуальной в современном мире. Решение этой проблемы требует комплексного подхода, включающего инвестиции в модернизацию инфраструктуры, разработку новых технологий и оптимизацию сетевой архитектуры.
**Ключевые слова:** Оценка пропускной способности волоконно-оптической линии связи

**Введение.** В современном мире цифровые технологии играют ключевую роль во всех сферах жизни, что приводит к экспоненциальному росту объема передаваемых данных. Волоконно-оптические линии связи (ВОЛС) зарекомендовали себя как незаменимый элемент современной инфраструктуры связи, обеспечивая передачу огромных объемов данных с высокой скоростью и надежностью, поэтому ВОЛС являются основой современной телекоммуникационной инфраструктуры, и их пропускная способность должна постоянно увеличиваться, чтобы соответствовать растущим потребностям. Этот рост обусловлен целым рядом факторов, которые необходимо учитывать. Факторы, определяющие рост объема передаваемых данных.

**Анализ пропускной способности волоконно-оптических линий связи.** Все больше людей получают доступ к высокоскоростному интернету, что стимулирует потребление цифрового контента. Миллиарды устройств, подключенных к Интернету, генерируют колоссальные объемы данных, что оказывает огромное влияние на современную инфраструктуру связи. Этот феномен является ключевым аспектом развития Интернета вещей (IoT) и цифровой трансформации общества. Интернет вещей превращает наш мир в постоянно подключенную экосистему, где миллиарды устройств непрерывно взаимодействуют друг с другом и с облачными сервисами. Ключевой характеристикой многих IoT-устройств является их способность работать круглосуточно, генерируя данные непрерывно. Этот непрерывный поток данных создает как значительные вызовы, так и открывает широкие возможности для различных отраслей. От социальных сетей и потокового видео до Интернета вещей и научных исследований, мы постоянно создаем и собираем огромные объемы информацию, то есть, поток данных приводящиеся к накоплению колоссальных объемов информации, что создает как значительные вызовы, так и открывает широкие возможности. Баланс между использованием возможностей и

преодолением вызовов определяет успех нашего движения в цифровом будущем. Облачные сервисы становятся основой для хранения и обработки данных, что увеличивает трафик между центрами обработки данных и конечными пользователями. Потоковые платформы, такие как Netflix и YouTube, генерируют огромные объемы трафика, особенно в высоком разрешении. Недостаточная пропускная способность приводит к задержкам и снижению качества услуг. Невозможность передачи больших объемов данных сдерживает развитие инновационных технологий.

**Принципы достижения высокого качества.** Модернизация инфраструктуры требует значительных инвестиций, что может привести к росту цен на услуги. Проблема увеличения пропускной способности ВОЛС является крайне актуальной в современном мире. Решение этой проблемы требует комплексного подхода, включающего инвестиции в модернизацию инфраструктуры, разработку новых технологий и оптимизацию сетевой архитектуры. Использование новых типов оптических волокон открывает возможности для значительного увеличения пропускной способности ВОЛС. Однако существуют и вызовы, связанные с:

– технологией производства новых типов волокон требует высокой точности и контроля;

– необходимо обеспечить совместимость новых волокон с существующей инфраструктурой;

– стоимостью: новые типы волокон могут быть дороже стандартных волокон.

**Заключение.** Несмотря на эти вызовы, разработка и внедрение новых типов оптических волокон является ключевым направлением развития ВОЛС, позволяющим удовлетворить растущие потребности в передаче данных.

### Список литературы

[1] С.А. Иванов, С.Л. Кузнецов, А.С. Попов, А.А. Рыбаков. Оптические системы передачи информации. М.: Горячая линия-Телеком, 2015.

[2] А.В. Фомин, А.В. Голенков, А.В. Дворкович. Волоконно-оптические системы передачи. М.: Горячая линия-Телеком, 2012.

[3] А. С. Сигов, В. В. Смирнов. Оптоэлектроника и волоконная оптика. М.: Физматлит, 2018

## THE RELEVANCE OF THE PROBLEM OF INCREASING THE CAPACITY OF FIBER-OPTIC COMMUNICATION LINES IN CONNECTION WITH THE GROWTH OF THE VOLUME OF TRANSMITTED DATA

***S.Tilevov***

*Institute of Telecommunications and Informatics of Turkmenistan, Ashgabat Turkmenistan*

**Abstract:** In today's world, where digital technologies permeate all spheres of life, the volume of transmitted data is growing exponentially. The solution to this problem requires a comprehensive approach, including investments in infrastructure modernization, development of new technologies and optimization of network architecture.

**Keywords:** Estimation of bandwidth in fiber-optic communication lines

УДК 004.021:004.75

# АНАЛИЗ НЕЛИНЕЙНЫХ ПАРАМЕТРОВ ЭЛЕКТРОЭНЦЕФАЛОГРАММ ЖЕНЩИН ПРИ ПАТОЛОГИИ В ВИДЕ ДЕПРЕССИИ

***А. В. Сидоренко***
*Профессор кафедры физики и аэрокосмических технологий БГУ, доктор технических наук, профессор*
sidorenkoA@yandex.by

***Н.А. Солодухо***
*старший преподаватель кафедры телекоммуникаций и информационных технологий БГУ, магистр физико-математических наук*
mikita.saladukha@gmail.com

***П.П. Королевич***
*врач психиатр-нарколог, врач функциональной диагностики, научный сотрудник отдела психических и поведенческих расстройств ГУ «Республиканский научно-практический центр психического здоровья».*
p.korolevich@gmail.com

***М.М. Скугаревская***
*заместитель директора по научной работе ГУ «Республиканский научно-практический центр психического здоровья», врач психиатр-нарколог, доктор медицинских наук, профессор*
zamdirnauka@mentalhealth.by

***А. В. Сидоренко***

*Окончила Белорусский государственный университет. Область научных интересов связана с разработкой методов и алгоритмов для создания систем на основе динамического хаоса в применении к сфере защиты информации различного назначения, а также обработки и анализа биоэлектрических сигналов*

***Н. А. Солодухо***

*Окончил Белорусский государственный университет, магистратуру и аспирантуру. Проводит научные исследования в области работы головного мозга, обработки биоэлектрических сигналов, шумового излучения.*

***П. П. Королевич***

*Окончил Белорусский государственный медицинский университет. Высшая квалификационная категория. Окончил ординатуру по специальности «психиатрия-наркология», окончил аспирантуру. Специалист в области электросудорожной терапии, электроэнцефалографии.*

*Принимает участие в исследованиях особенностей биоэлектрической активности головного мозга пациентов с психическими и поведенческими расстройствами.*

***М. М. Скугаревская***

*Окончила Белорусский государственный медицинский университет. Сфера научных интересов: ранняя диагностика психических расстройств, лечение и профилактика психических и поведенческих расстройств.*

**Аннотация.** Проведены экспериментальные исследования электроэнцефалограмм пациентов женского пола с патологией в виде депрессии. Электроэнцефалограммы регистрировались в стандартных отведениях. Проведены обработка и анализ как линейных, так и нелинейных параметров электроэнцефалограмм, выраженных относительной спектральной плотностью мощности отдельных ритмических компонент мозга, выборочной и перестановочной энтропиями, экспоненты самоподобия. Установлено, что нелинейные параметры электроэнцефалограмм изменяются сильнее относительно фонового состояния, по сравнению с линейными, и обладают меньшей относительной погрешностью.

**Ключевые слова**: электроэнцефалограмм, человек, депрессия, нелинейные параметры, выборочная энтропия, перестановочная энтропия, экспонента самоподобия.

**Введение.** Широкое распространение информационных технологий среди различных сфер деятельности человека коснулось и медицины, среди которых значительное место занимает неврология. Проблема заключается в том, что анализ распространенности патологии в виде депрессии за последние годы выявил тенденцию к возрастанию и сохранению случаев этой патологии [1]. Точное определение этой патологии позволит своевременно проводить лечебные или профилактические действия, что положительно скажется на здоровье населения и является актуальной задачей в настоящее время. Одним из способов уточнения наличия депрессии может стать использование не только классических методов анализа электроэнцефалограмм пациента, но и нелинейных, включая метод анализа выборочной и перестановочной энтропий, а также флуктуационный анализ с исключенным трендом для анализа с использованием экспоненты самоподобия.

Целью работы является обработка и анализ электроэнцефалограмм пациентов женского пола с патологией в виде депрессии. В качестве параметров, применяемых при обработке и анализе электроэнцефалограмм, использовались спектральная плотность мощности различных ритмических компонент мозга, а также выборочная, перестановочная энтропии и экспонента самоподобия.

**Метод определения выборочной энтропии**. Электроэнцефалограмма представляется в виде последовательности чисел временного ряда u(j) из N точек {u(j), 1≤j≤N}. Выборочная энтропия SampEn временного ряда u(j) из N точек {u(j), 1≤j≤N} вычисляется следующим образом. Формируется N-m векторов

$$x_m(i) \text{ для } \{1 \le i \le N-m\}, \text{ где}$$
$$x_m(i) = \{u(i+k), 0 \le k \le m-1\} \quad (1)$$

является вектором из m точек от u(i) до u(i+m-1). Расстояние между двумя такими векторами $x_m(i)$ и $x_m(j)$ определяется как

$$d[x(i), x(j)] = \max\{|u(i+k) - u(j+k)| : 0 \le k \le m-1\}, \quad (2)$$

т.е. максимальное расстояние между их соответствующими скалярными компонентами.

Для вычисления выборочной энтропии SampEn требуется следующее. Определим $B_i^m(r)$ как деленное на (N-m-1) число векторов $x_m(j)$ на расстоянии r от $x_m(i)$, где j меняется в диапазоне [1, N-m], причем j≠i. Далее определяем

$$B^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} B_i^m(r) \quad (3)$$

Аналогичным образом определяем $A_i^m(r)$ как деленное на (N-m-1) число векторов $x_{m+1}(j)$ на расстоянии r от $x_{m+1}(i)$, где j меняется в диапазоне [1, N-m], причем j≠i. Далее определяем

$$A^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} A_i^m(r) \quad (4)$$

Таким образом, $B^m(r)$ представляет собой вероятность того, что две рассматриваемые последовательности будут совпадать на m точек. $A^m(r)$ это вероятность того, что две последовательности будут совпадать на m+1 анализируемых точек. Значение выборочной энтропии определяется:

$$SampEn(m, r, N) = -\ln\left(\frac{A^m(r)}{B^m(r)}\right) \quad (5)$$

Выборочная энтропия определяется как отрицательный натуральный логарифм вероятности того, что две похожие последовательности длиной m останутся похожими в следующей точке пространства задержанных координат. Таким образом, более низкое

значение выборочной энтропии свидетельствует о большем самоподобии во временной последовательности данных [2].

**Метод перестановочной энтропии.** Особенностью перестановочной энтропии H является то, что ее значение связано с характером изменения чисел в последовательности, представляющей сигнал электроэнцефалограммы [3]. Для расчета перестановочной энтропии временного ряда $\{x_i\}$ длиной N (т.е. 1<i<N), представляющего сигнал электроэнцефалограммы, необходимо:

1 Определим порядок n. Данный порядок определяет возможные варианты перестановок $\pi_j$ (j=1,...n!), которые строятся из чисел 1,...,n. Их графическое представление изображено на рисунке ниже (рисунок 1) для n=3.

2 Далее необходимо присвоить индексу i рассматриваемого временного ряда $\{x_i\}_{i=1,..,N}$ значение 1, а счетчик количества каждого варианта перестановок $z_j$ установить равным 0 для каждого $\pi_j$ .

3 Определить порядок следования чисел в последовательности $x_i,..., x_{i+n-1}$

*Рисунок 1.* Графическое представление возможных вариантов перестановок для трех рассматриваемых чисел

4 Сравнить порядок следования чисел в последовательности $x_i,..., x_{i+n-1}$ с существующими вариантами перестановки и увеличить счетчик соответствующего варианта перестановки на единицу.

5 Если выполняется условие $i \leq N-n$, следует увеличить i на единицу и перейти к шагу 3. Если выполняется условие i>N-n, необходимо перейти к следующему шагу.

6 Определить относительную частоту всех перестановок $p'_j$ по формуле

$$p'_j = \frac{z_j}{\sum z_k} \tag{6}$$

где $z_j$ или $z_k$ – счетчик количества j-ого или k-ого варианта перестановок соответственно. Суммирование $z_k$ в формуле (6) производится по всем возможным вариантам перестановок

7 Выбрать все $p'_j$ больше 0 и вычислить перестановочную энтропию PermEn:

$$PermEn = -\sum_{j=1}^{n!} p'_j \log_2(p'_j) \tag{7}$$

**Экспонента самоподобия с помощью метода флуктуационного анализа с исключенным трендом.** Экспонента самоподобия рассчитывается согласно алгоритму флуктуационного анализа с исключенным трендом (DFA). Электроэнцефалограмма представляется временным рядом v из N элементов. Экспонента самоподобия h некоторого временного ряда v(i), представляющего сигнал электроэнцефалограммы, где i изменяется от 1 до N, вычисляется с использованием метода флуктуационного анализа с исключенным трендом (DFA) следующим образом.

Определяется квадрат значения каждого элемента временного ряда:

$$x(i) = v^2(i), 1 \leq i \leq N \tag{8}$$

Далее от каждого элемента временного ряда x(i), рассчитанного по формуле (8), вычитается среднее значение всего временного ряда X, состоящего из элементов, рассчитанных по формуле (8), и производится суммирование:

$$y(k) = \sum_{i=1}^{k} (x(i) - < x >), 1 \le k \le N, \quad (9)$$

где <x> - это среднее значение временного ряда X в диапазоне [1, N]. Ряд Y, состоящий из рассчитанных по формуле (9) элементов, разбивается на сегменты длиной n (n временной масштаб наблюдения). Для каждого из этих сегментов определяется локальный тренд $y_n$ с помощью метода наименьших квадратов. Определяется среднеквадратичное отклонение между y(k) и $y_n(k)$ в диапазоне [1, N]:

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (y(k) - y_n(k))^2} \quad (10)$$

В дальнейшем вычисление F(n) по формуле (10) проводится для различных значений n из диапазона nЄ[4;N/4]. Финальным шагом является построение зависимости значения преобразования F(n) от n в двойном логарифмическом масштабе. Тангенс угла наклона данной зависимости log(F(n)) / log(n) и дает значение экспоненты самоподобия, т.к. выполняется соотношение $F(n) \sim n^h$[4].

**Методика проведения исследований.** Регистрация электроэнцефалограмм осуществлялась по схеме «10/20» с использованием электроэнцефалографа «Мицар-ЭЭГ-201» фирмы “Мицар”. Относительная погрешность измерения напряжения электроэнцефалографа составляет 10% [5]. Анализировались ЭЭГ следующих отведений: C4, F3, F4, Fp1, O1, O2, P4, T4, T5, T6. Выбирались те отведения, в которых наблюдалось наиболее значительное изменение параметра. Обработка и анализ электроэнцефалограмм проводились в разработанной авторами компьютерной программе[6]. Электроэнцефалограммы женщин обрабатывались в следующих режимах: здоровые; депрессия. Количество испытуемых составило 28 человек, причем 21,4% из них были с депрессией (количество женщин с депрессией составило 6 человек? количество здоровых женщин составило 22 человека)/ Погрешность рассчитывалась с использованием коэффициентов Стьюдента с доверительной вероятностью 0,9.

Выборочная и перестановочная энтропии оценивались с помощью одноименных методов. Экспонента самоподобия рассчитывалась в соответствии с методом флуктуационного анализа с исключенным трендом. Спектральная плотность мощности рассчитывалась с помощью быстрого преобразования Фурье.

**Результаты**. При наличии патологии в виде депрессии спектральная плотность мощности дельта-ритма в отведении O1 возрастала на 6,8% относительно спектральной плотности мощности тета-ритма у здоровых женщин (рисунок 2).

*Рисунок 2.* Гистограмма изменения спектральной плотности мощности дельта-ритма в отведении Fp1 при наличии патологии в виде депрессии

При наличии патологии в виде депрессии спектральная плотность мощности тета-ритма в отведении O1 падала на 5,7% относительно спектральной плотности мощности тета-ритма у здоровых женщин (рисунок 3).

*Рисунок 3*. Гистограмма изменения спектральной плотности мощности тета-ритма в отведении O1 при наличии патологии в виде депрессии

При наличии патологии в виде депрессии спектральная плотность мощности альфа-ритма в отведении T4 уменьшилась на 19,4% относительно значения у здоровых женщин, а в отведении T6 - на 19,5% относительно значения у здоровых женщин (рисунок 4).

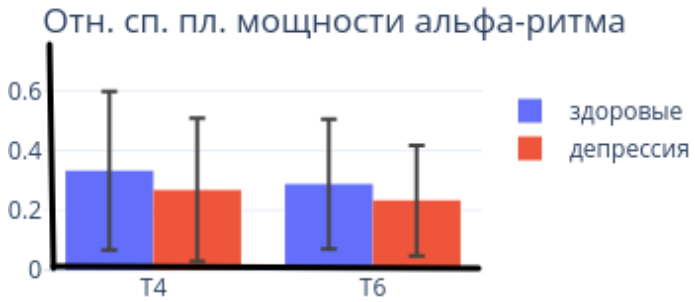



*Рисунок 4*. Гистограмма изменения спектральной плотности мощности альфа-ритма в отведениях T4, T6 при наличии патологии в виде депрессии

Спектральная плотность мощности бета-ритма в отведении F3 при наличии патологии в виде депрессии падала на 26,2% относительно значения у здоровых женщин (рисунок 5).

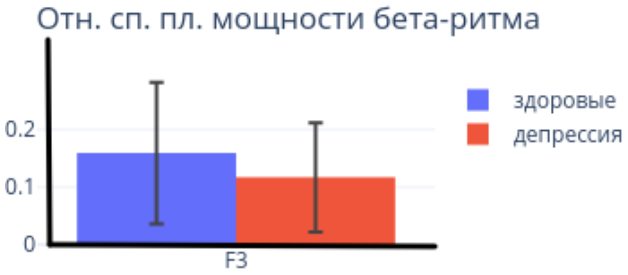



*Рисунок 5*. Гистограмма изменения спектральной плотности мощности бета-ритма в отведении F3 при наличии патологии в виде депрессии

При наличии патологии в виде депрессии спектральная плотность мощности гамма ритма в отведении C4 падала на 25,8% относительно значения у здоровых женщин, в отведении Fp1 - на 34%, в отведении T4 - на 29,9%, в отведении T6 - на 28,3% (рисунок 6).

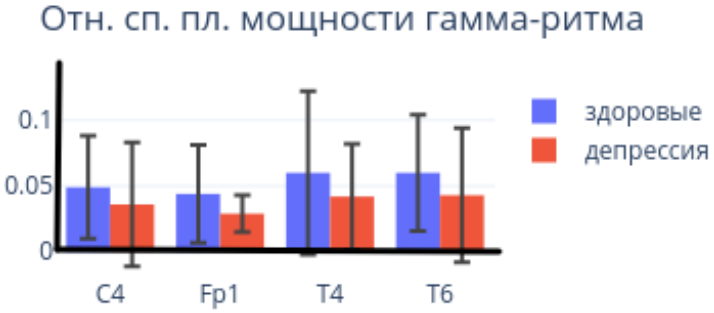



*Рисунок 6*. Гистограмма изменения спектральной плотности мощности гамма-ритма в отведениях C4, Fp1, T4, T6 при наличии патологии в виде депрессии

Выборочная энтропия в отведении F4 при наличии патологии в виде депрессии падала на 8,8% от значения выборочной энтропии в этом отведении у здоровых женщин, в отведении Fp1 - на 7,3%, в отведении O2 - на 5,1% от значения у здоровых женщин (рисунок 7).

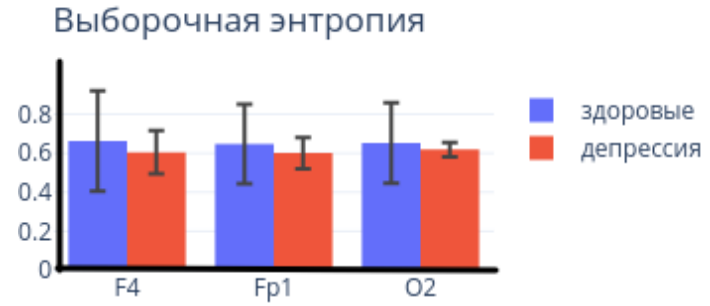



*Рисунок 7.* Гистограмма изменения выборочной энтропии в отведениях F4, Fp1, O2 при наличии патологии в виде депрессии

При наличии патологии в виде депрессии перестановочная энтропия в отведении Fp1 уменьшалась на 4,2%, а в отведении O2 - на 3,2% от соответствующего значения у здоровых женщин (рисунок 8).

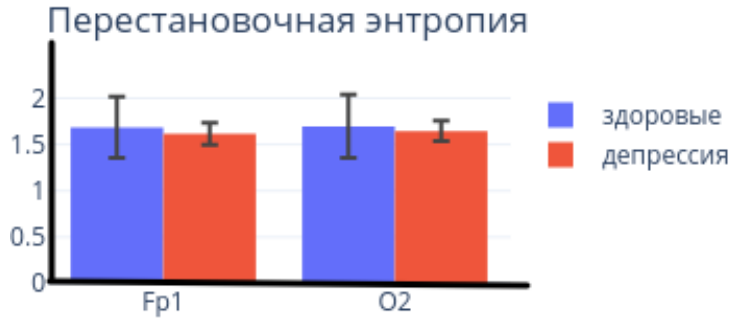



*Рисунок 8.* Гистограмма изменения перестановочной энтропии в отведениях Fp1, O2 при наличии патологии в виде депрессии

При наличии патологии в виде депрессии экспонента самоподобия в отведении C4 падает на 20,8%; в отведении F3 -на 25,6%; в отведении F4 - 17,2%; в отведении P4 - на 17,4%; в отведении T5 – на 23,3%; в отведении T6 – на 27% – от соответствующих значений экспоненты самоподобия у здоровых женщин (рисунок 9).

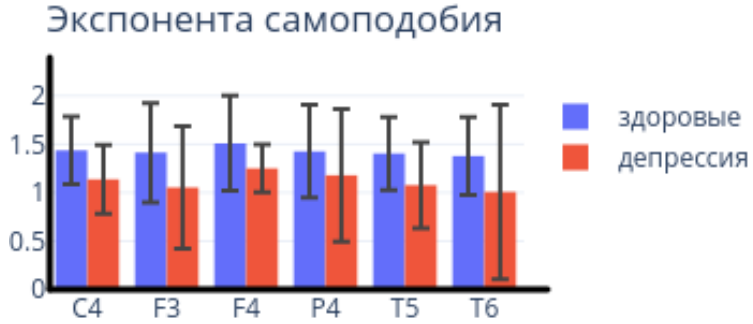



*Рисунок 9.* Гистограмма изменения экспоненты самоподобия в отведениях C4, F3, F4, P4, T5, T6 при наличии патологии в виде депрессии

**Заключение.** Проанализированы экспериментально полученные и обработанные линейные и нелинейные параметры электроэнцефалограмм женщин с патологией в виде депрессии. Проведен анализ следующих линейных параметров: спектральные плотности мощности дельта-, тета-, бета-, гамма-ритмов. Анализируемые нелинейные параметры электроэнцефалограмм представлены выборочной, перестановочной энтропиями, экспонентой самоподобия. Установлено, что нелинейные параметры при патологии

меняются сильнее, обладают меньшей относительной погрешностью по сравнению с линейными параметрами исследуемых электроэнцефалограмм. Тенденции изменения параметров электроэнцефалограмм в анализируемых отведениях при наличии патологии в виде депрессии сводятся к следующему. Спектральная плотность мощности дельта-ритма возрастала на 6,8% относительно спектральной плотности мощности дельта-ритма у здоровых женщин. Спектральная плотность мощности тета-ритма падала на 5,7% относительно спектральной плотности мощности тета-ритма у здоровых женщин. Спектральная плотность мощности альфа-ритма в отведении уменьшилась на 19,4%-19,5% относительно значения у здоровых женщин. Спектральная плотность мощности бета-ритма падала на 26,2% относительно значения у здоровых женщин. Спектральная плотность мощности гамма-ритма падала на 25,8%-34% относительно значения у здоровых женщин. Выборочная энтропия падала на 5,1%-8,8% относительно значения у здоровых женщин. Перестановочная энтропия уменьшилась на 3,2%-4,2% относительно значения у здоровых женщин. Экспонента самоподобия снизилась на 17,4%-25,6% относительно значения у здоровых женщин.

## Список литературы

[1] Darío Moreno-Agostino, Global trends in the prevalence and incidence of depression:a systematic review and meta-analysis / Journal of Affective Disorders - 2021. - Vol. 281. - P. 235-243.

[2] Richman J.S., Moorman J.R. Physiological time-series analysis using approximate entropy and sample entropy / American Journal of Physiology - Heart and Circulatory Physiology. – 2000. – Vol. 278, № 6. – P. 2039–2049.

[3] Bandt C., Pompe B. Permutation Entropy: A Natural Complexity Measure for Time Series / Physical Review Letters. - 2002. - Vol. 88, iss. 17 - P. 1-4.

[4] Golinska, A.K. Detrended Fluctuation Analysis (DFA) in biomedical signal processing: selected examples / Studies In Logic, Grammar And Rhetoric. – 2012. – Vol. 42, № 29. – P. 107-115.

[5] Комплексы аппаратно-программные электроэнцефалографические “Мицар-ЭЭГ” : государственный реестр средств измерений, регистрационный № 43721-10 Российская Федерация, ТУ 9441-001-52118320-2009 / Н.И. Ханов; ООО “Мицар” – № 43721-10; 28.12.2009 // Государственный реестр средств измерений – 2009.

[6] Компьютерная программа обработки и информационного анализа электроэнцефалограмм "jASCeeg" : свидетельство о добровольной регистрации и депонировании объекта авторского права / объекта смежных прав № 1577-КП / А.В. Сидоренко, Н.А. Солодухо – Опубл. 19.01.2023.

# ANALYSIS OF NONLINEAR PARAMETERS OF EEG IN WOMEN WITH DEPRESSION

***A. V. Sidorenko***
*Professor of the Department of Physics and Aerospace Technologies, BSU, Doctor of Science in Technical Sciences*

***M.A. Saladukha***
*Senior Lecturer of the Department of Telecommunications and Information Technology, BSU, Master of Physical and Mathematical Sciences*

***P. P. Korolevich***
*Psychiatrist-Narcologist, Functional Diagnostics Doctor, Researcher of Department of Mental and Behavioral Disorders of Republican Scientific and Practical Center for Mental Health*

***M. M. Skugarevskaya***
*Deputy Director for Scientific Work of Republican Scientific and Practical Center for Mental Health, Psychiatrist-Narcologist, Doctor of Science in Medical Sciences, Professor*

**Abstract**: Experimental studies of electroencephalograms (EEGs) of female patients with depression were conducted. EEGs were recorded in standard leads. Both linear and nonlinear parameters of EEGs, expressed as the relative spectral power density of individual rhythmic brain components, sample entropy, permutation entropy, and self-similarity exponent, were processed and analyzed. It was found that nonlinear EEG parameters change more significantly compared to the baseline state, compared to linear ones, and have a lower relative error.

**Keywords**: electroencephalogram, human, depression, nonlinear parameters, sample entropy, permutation entropy, self-similarity exponent.

УДК 004.896

# АСПЕКТЫ ПРИМЕНЕНИЯ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ ЭНЕРГОПОТРЕБЛЕНИЯ НА НЕФТЕПЕРЕРАБАТЫВАЮЩЕМ ЗАВОДЕ

***И.А. Бабкин***
*Аспирант кафедры математического обеспечения и стандартизации информационных технологий РТУ МИРЭА*
naliab@yandex.ru

***Е.И. Алаторцев***
*Профессор кафедры математического обеспечения и стандартизации информационных технологий РТУ МИРЭА, доктор технических наук*
alatortsev@mirea.ru

***И.А. Бабкин***

*Окончил Финансовый университет при Правительстве Российской Федерации. Область научных интересов связана с исследованием применения моделей машинного обучения в технологических процессах нефтеперерабатывающих заводов.*

***Е.И. Алаторцев***

*Окончил УВВТУ им. Б.Хмельницкого. Область научных интересов связана с исследованиями бизнес-процессов управления качеством в технологических процессах нефтеперерабатывающих заводов.*

**Аннотация.** В условиях цифровизации экономики России нефтегазовая отрасль должна соответствовать нарастающим технологическим требованиям, в том числе по оптимизации энергозатрат. Традиционные методы сокращения энергопотребления на нефтеперерабатывающих заводах ограничены в возможностях из-за низкого уровня автоматизации. Поэтому были рассмотрены современные подходы по цифровизации, связанные с использованием методов глубокого обучения, позволяющие значительно сократить энергозатраты за счет повышения на первом этапе оптимизации качества прогнозирования энергопотребления.

Внедрение нейросетевых моделей на нефтеперерабатывающих заводах и развитие технологий машинного обучения позволяет повысить уровень цифровизации и эффективности процесса оптимизации энергопотребления.

**Ключевые слова:** машинное обучение, глубокое обучение, нейронные сети, искусственный интеллект (ИИ), анализ данных, прогнозирование, энергопотребление, нефтепереработка, нефтегазовая отрасль, повышение эффективности.

**Введение.** Нефтеперерабатывающие заводы (НПЗ) играют важнейшую роль в удовлетворении глобальных энергетических потребностей и обеспечении экономического развития большинства стран, питая транспорт, промышленность, производство, строительство и многие другие сектора экономики. Процесс переработки нефтяного сырья состоит из последовательности сложных первичных и вторичных технологических процессов, таких как атмосферная и вакуумная дистилляция, коксование, гидрокрекинг и риформинг, для получения широкого спектра продуктов, включая бензин, керосин, дизельное топливо, мазут и другие.

Энергоемкость данного процесса составляет значительную часть операционных затрат и любое вмешательство в него может повлиять на качество выпускаемых нефтепродуктов [1]. Принимая во внимание масштабы этих предприятий, становится все

более актуальной задача исследования эффективных мер по снижению издержек производства, в том числе и энергозатрат. Кроме того, в условиях стремления к снижению выбросов углекислого газа и переходу к экологически безопасным технологиям НПЗ вынуждены искать инновационные решения, позволяющие одновременно повысить производительность и минимизировать негативное воздействие на окружающую среду.

Современным ответом на данные вызовы стала цифровизация нефтеперерабатывающей промышленности на всех ее этапах. К примеру, можно наблюдать успешную интеграцию технологии «распределенного реестра» (блокчейн), а именно смарт-контрактов, в цепочки поставок нефтепродуктов для прозрачного контроля за исполнением обязательств и эффективной цифровизации операционных процессов [2]. Также широко применяются облачные технологии и цифровые платформы для централизованного управления и координацией работы нефтехимических комплексов.

Особенно существенно на современных НПЗ отразилось внедрение решений «промышленного интернета вещей» (Industrial Internet of Things, IIoT), ключевого элемента Индустрии 4.0. Оснащение технологических установок на производстве множеством датчиков позволяет проводить непрерывный мониторинг и собирать большие массивы данных для их дальнейшего анализа. Поэтому открывается возможность применения комплексных интеллектуальных решений с целью повышения эффективности производства на НПЗ.

**Методы прогнозирования энергопотребления.** К стандартным методам оптимизации энергопотребления можно отнести физико-химическое моделирование, линейное программирование и экспертные системы. Анализ показал, что они имеют ряд ограничений, связанных со сложностью настройки и необходимостью постоянного ручного обновления моделей функционирования. Классические регрессионные методы, такие как линейная и полиномиальная регрессия, традиционно используются для задачи прогнозирования, но их применимость ограничена [3]. Связано это с тем, что данные методы предполагают наличие линейных зависимостей между переменными, что не всегда соотносится со сложными процессами производства на НПЗ. Регрессионные модели плохо адаптируются к нелинейным изменениям в потреблении энергии, вызванным отклонениями параметров сырья, окружающей среды и эксплуатационных условий от привычных значений. Такие особенности делают эти методы недостаточными для точного прогнозирования энергопотребления.

В научных работах, посвященных настоящему вопросу, подчеркиваются преимущества использования в данной задаче методов глубокого обучения, а именно искусственных нейронных сетей (ИНС) [4]. В отличие от обычных методов машинного обучения, инструменты глубокого обучения позволяют анализировать большие объемы данных, выявлять скрытые закономерности и адаптироваться к изменениям в технологических процессах на производстве в автоматическом режиме. Используя множество взаимосвязанных слоев нейронов, ИНС способны распознавать закономерности в данных, которые были бы невидимы для моделей машинного обучения, основанных на простых статистических закономерностях и предполагающих предсказуемый результат.

Одними из ключевых типов ИНС в задаче прогнозирования энергопотребления выделяют методы, основанные на обработке последовательностей данных, а не отдельных событий, называемые рекуррентные нейронные сети (Recurrent Neural Network, RNN). Такие виды ИНС учитывают зависимость текущих потребностей в топливе от ранее фиксированных состояний производственной системы, что позволяет значительно увеличить точность прогнозирования колебания в энергопотреблении. В частности, рассматривают сети с долгой краткосрочной памятью (Long Short-Term Memory, LSTM), способные сохранять информацию о долгосрочных зависимостях, не теряя ее в процессе обучения. Они могут демонстрировать высокую точность прогнозирования на краткосрочные периоды (до недели) с погрешностью менее 1% [5].

Дополнительно могут применяться сверточные нейронные сети (Convolutional Neural Network, CNN), позволяющие выявлять пространственные закономерности в данных. Применение такой архитектуры как «одномерная свертка» (1D-CNN) в сочетании со слоями RNN позволяет учитывать как временные, так и пространственные особенности данных с установок НПЗ, повышая точность прогнозирования.

Помимо этого, перспективными являются прогностические модели с использованием «обучения с подкреплением». Такие модели способны в реальном времени адаптироваться к изменениям на производственных системах, оптимизируя энергопотребление на установках на основе накопленного опыта. Использование глубокого обучения с подкреплением (Deep Reinforcement Learning, DRL) способно на основе динамической обратной связи корректировать параметры технологического процесса без нарушения работы производства. Также особенностью этого метода является возможность обучения на цифровом двойнике НПЗ для тестирования различных стратегий управления его энергопотреблением.

Совмещая масштабируемость моделей глубокого обучения, использующие большие наборы данных, с более простыми архитектурами обычных методов машинного обучения, можно добиться оптимальной работы прогностических алгоритмов. Например, гибридные модели, объединяющие градиентный бустинг деревьев решений (XGBoost) со слоями RNN или CNN, демонстрируют высокую устойчивость к шуму в данных, позволяют учитывать только самые значимые признаки, не упуская при этом сложные временные закономерности и адаптируясь к изменяющимся условиям. Это приводит к уменьшению размерности данных и снижению возможности переобучения модели. Синергия отличающихся между собой подходов к прогнозированию обеспечивает вариативность регулировки параметров модели машинного обучения и возможность использовать преимущества каждого из подходов [6].

**Вызовы и перспективы внедрения глубокого обучения на НПЗ.** Помимо очевидных преимуществ глубокое обучение также характеризуется рядом сложностей, которые могут проявиться при его использовании на промышленных производствах. Они могут обусловлены как техническим, так и экономическими и организационными факторами. Можно выделить следующие вызовы в данной области:

*1 Сложность интерпретации обучаемых моделей.* Решения, получаемые внедрением методов глубокого обучения, не обладают достаточной объяснимостью из-за своей архитектуры. Они представляют из себя «черные ящики», выдающие желаемый результат, но не способные предоставить прозрачную логику своего выбора. Операторам технологических установок и управленческому персоналу НПЗ будет затруднительно проводить диагностику ошибок и поиск причин возникновения отклонений в прогнозах. Это создает препятствия для внедрения таких решений в промышленную среду.

*2 Дефицит качественных данных для обучения.* Ключевым аспектом обучения моделей машинного обучения являются данные. Эффективность обучаемых моделей существенно зависит от качества и объема поступаемых им данных. Данные, получаемые с технологических установок НПЗ, отличаются сильной разнородностью и могут содержать шумы и искажения. Для решения этой проблемы необходимо прибегнуть к ряду процедур по очистке данных и их нормализации. Помимо этого, накопленных на НПЗ данных может не хватать для качественного обучения моделей. В таком случае можно воспользоваться синтетическими наборами данных, сгенерированных на основе существующих, что, в свою очередь, негативно отразится на степени обобщенности модели.

*3 Потенциальная уязвимость информационной безопасности.* Автоматизация производственных процессов на НПЗ и интеграция IIoT-решений открывает возможность для потенциальных кибератак на предприятия. В связи со стратегическим значением предприятий нефтегазовой отрасли внедрение в них систем управления, основанных на глубоком обучении, возможно только при обеспечении необходимого уровня защиты ИТ-

инфраструктуры и данных.

*4 Экономические затраты при внедрении.* Внедрение решения на основе ИНС потребует капитальных затрат из-за развертывания высокопроизводительных вычислительных ресурсов, привлечения соответствующих специалистов, сопровождения и обслуживания системы. Сюда же включены возможные риски ошибочной работы обученной модели на реальных данные, последствия которой могут быть как несущественными, так и угрожающими остановкой НПЗ.

*5 Интеграция с устаревшими системами.* Несовместимость предлагаемой модели с развернутыми на НПЗ системами управления (например, SCADA, ERP) потребует дополнительной разработки методов и протоколов их взаимодействия, поскольку они в этом ограничены производителями.

Несмотря на рассмотренные сложности, потенциальная польза от внедрения глубокого обучения на НПЗ указывает на перспективность такого решения. Современные подходы к разработке ИНС позволяют адаптировать модели к изменяющимся условиям эксплуатации, распределять вычисления на физические и облачные ресурсы, развертывать их в безопасной виртуальной среде, а также проверять их работу на цифровых двойниках НПЗ (при их наличии).

Помимо повышения эффективности прогнозирования энергопотребления модели глубокого обучения способны при необходимости без дополнительных затрат осуществлять контроль и прогноз состояния технологического оборудования на НПЗ, оптимизацию плана их технического обслуживания. Кроме того, ИНС могут использоваться для обнаружения предаварийных ситуаций на производстве и выброса вредных веществ.

**Заключение.** Интеллектуализация производственных процессов необходима в условиях продолжающейся цифровизации экономики России. Среди ряда сквозных технологий машинное обучение и искусственный интеллект представляют собой перспективное направление цифровизации нефтегазовой отрасли, способное существенно повысить эффективность производства на НПЗ. Использование различных архитектур ИНС позволяет учитывать различные сложные зависимости в данных, адаптироваться к изменяющимся условиям, а их комбинации могут использовать преимущества нескольких подходов. Такие системы способны не только повышать точность прогнозирования энергопотребления на НПЗ, но и способствуют оптимизации других производственных процессов, снижению выброса вредных веществ, улучшению контроля за состоянием оборудования и предотвращению аварийных ситуаций.

Дальнейшее развитие технологий цифровых двойников и облачных вычислений может существенно облегчить внедрение интеллектуальных решений на НПЗ. В целом, сочетание передовых методов машинного обучения с традиционными подходами к оптимизации производства, несмотря на сложности их внедрения, открывает новые возможности для повышения эффективности энергопотребления и экологичности нефтеперерабатывающей промышленности.

## Список литературы

[1] Алаторцев Е.И. К вопросу о методическом обеспечении контроля и сохранения качества нефти и нефтепродуктов // Нефтепереработка и нефтехимия. Научно-технические достижения и передовой опыт. – 2021. – № 10. – С. 33-36.

[2] Терешкин С.И., Малышевский П.В., Белова И.В., Бабкин В.А., Бабкин И.А. Смарт-контракты и цифровая платформа для нефтяного рынка // Экономика. Информатика. – 2024. – Т. 51, № 1. – С. 211-220. – DOI 10.52575/2712-746X-2024-51-1-211-220.

[3] Li J., Guo Y., Zhang X., Fu Z. Using Hybrid Machine Learning Methods to Predict and Improve the Energy Consumption Efficiency in Oil and Gas Fields // Mobile Information Systems. – 2021. – Vol. 2021. – P. 1-7. – DOI 10.1155/2021/5729630.

[4] R Azmi P. A., Yusoff M., Mohd Sallehud-Din M.T. A Review of Predictive Analytics Models in the Oil and Gas Industries // Sensors. – 2024. – Vol. 24, No. 12. – P. 4013. – DOI 10.3390/s24124013.

[5] Клюев Р.В., Моргоева А.Д., Гаврина О.А., Босиков И.И., Моргоев И.Д. Прогнозирование планового потребления электроэнергии для объединенной энергосистемы с помощью машинного обучения // Записки Горного института. – 2023. – Т. 261. – С. 392-402.

[6] Mathumitha R., Rathika P., Manimala K. Intelligent deep learning techniques for energy consumption forecasting in smart buildings: a review // Artificial Intelligence Review. – 2024. – Vol. 57, No. 2. – P. 35. – DOI 10.1007/s10462-023-10660-8.

## Авторский вклад

**Бабкин Илья Андреевич** – анализ существующих методов прогнозирования потребления энергоресурсов на предприятиях нефтеперерабатывающей промышленности, исследование и описание вызовов и перспектив внедрения методов глубокого обучения на производстве.

**Алаторцев Евгений Иванович** – постановка научной проблемы, описание современного состояния производства нефтепродуктов, подбор источников, анализ полученных результатов и формирование заключения.

# ASPECTS OF APPLYING DEEP LEARNING METHODS FOR ENERGY CONSUMPTION OPTIMIZATION AT AN OIL REFINERY

***I.A. Babkin***
*Postgraduate student of the Department of Mathematical Support and Standardization of Information Technologies RTU MIREA*

***E.I. Alatortsev***
*Professor of the Department of Mathematical Support and Standardization of Information Technologies RTU MIREA, Doctor of Technical Sciences*

**Abstract.** In the context of digitalization of the Russian economy, the oil and gas industry must meet the growing technological requirements, including energy consumption optimization. Traditional methods of reducing energy consumption at oil refineries are limited in their capabilities due to the low level of automation. Therefore, modern digitalization approaches related to the use of deep learning methods were considered, which allow to significantly reduce energy costs by improving the quality of energy consumption forecasting at the first stage of optimization.

The introduction of neural network models at refineries and the development of machine learning technologies can increase the level of digitalization and efficiency of the energy consumption optimization process.

**Keywords**: machine learning, deep learning, neural networks, artificial intelligence (AI), data analysis, forecasting, energy consumption, oil refining, oil and gas industry, efficiency improvement.

УДК 004.021:004.832.2

# МОДЕЛИРОВАНИЕ ПЕРСОНАЛИЗИРОВАННОЙ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ НА ОСНОВЕ ГЛУБОКОГО ОБУЧЕНИЯ

***М.Д. Михальков***
*магистрант кафедры ПОИТ, ФКСиС, БГУИР*
solution.m@mail.ru

***С.Н. Нестеренков***
*кандидат технических наук, доцент кафедры ПОИТ, ФКСиС, БГУИР*
s.nesterenkov@bsuir.by

***М.Д. Михальков***

*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с анализом и разработкой моделей и алгоритмов построения интеллектуальных систем поддержки принятия решений.*

***С.Н. Нестеренков***

*Окончил аспирантуру Белорусского государственного университета информатики и радиоэлектроники (2013), в 2017 защитил диссертацию на соискание ученой степени кандидата технических наук по специальности 05.13.01 - «Системный анализ, управление и обработка информации». Является доцентом кафедры ПОИТ.*

**Аннотация.** В рамках данного исследования было проведено моделирование рекомендательной системы, предназначенной для платформ электронной коммерции. В процессе моделирования были рассмотрены различные архитектурные решения, формализованы основные компоненты системы и разработаны математические модели, описывающие работу рекомендательных алгоритмов.

Особое внимание было уделено выбору и анализу методов рекомендаций, включая гибридные модели и глубокое обучение. В данной статье было представлено концептуальное моделирование системы, описание ее логической и функциональной структуры, выбор формализованных методов и алгоритмов, а также анализ возможных сценариев работы.

**Ключевые слова:** Рекомендательные системы, глубокое обучение, DLRM, матричная факторизация, гибридная фильтрация.

**Введение.** В статьях [1, 2] был представлен обзор моделей и методов построения интеллектуальных систем поддержки принятия решений в области электронной коммерции. В данном же исследовании будет рассмотрено описание модели рекомендательной системы, включая основные подходы и методы, используемые для создания эффективных рекомендаций. Подразумевается, что в последующем будет смоделирована общая система, которая представляет собой рекомендательную систему, интегрированную с интеллектуальной системой поддержки принятия решений.

Глубокое обучение – это современное решение многих проблем машинного обучения, которое превосходит альтернативные методы. Последние тенденции включают применение методов глубокого обучения в рекомендательных системах. Многие крупные компании, такие как *AirBnB, Facebook, Google, Home Depot, LinkedIn* и *Pinterest*, делятся своим опытом использования глубокого обучения для рекомендательных систем.

Далее будет рассмотрена архитектура моделей на основе глубокого обучения.

**Описание базовой модели рекомендательной системы на основе глубокого обучения.** Эмбеддинг-слои представляют категории с плотными векторами. Эта техника популярна в *NLP* для встраивания слов с плотным представлением. Слова с похожим значением имеют похожий эмбеддинг-вектор. Эту же технику можно применить к рекомендательным системам. Самая тривиальная рекомендательная система основана на пользователях и объектах-продуктах (айтемах). Имеются идентификаторы пользователей и идентификаторы айтемов. Слова – это пользователи и айтемы, поэтому в данном случае рассматривается две эмбеддинг-таблицы (рисунок 1).

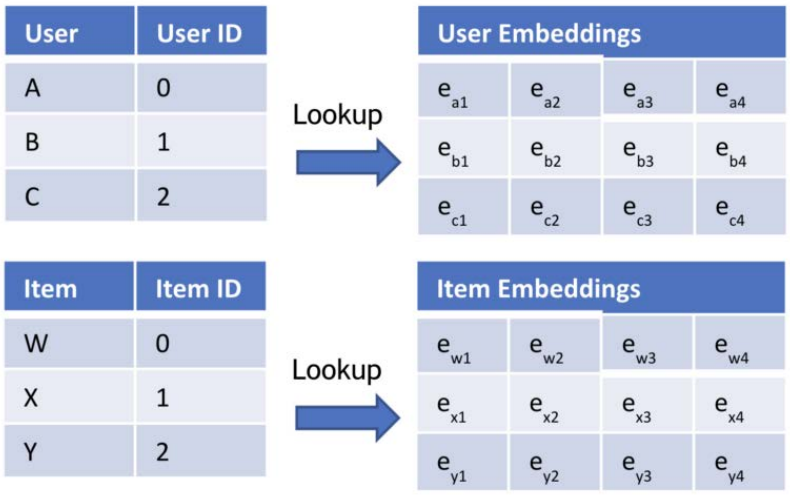



*Рисунок 1.* Эмбеддинг-таблицы пользователей и айтемов

Вычисление *dot*-умножения между эмбеддингом пользователя и эмбеддингом айтема позволяет получить окончательный результат – оценку релевантности данного айтема для пользователя. Этот подход широко применяется в нейросетевых рекомендательных системах, так как он эффективно моделирует скрытые предпочтения пользователя.

Для получения вероятности взаимодействия данный результат можно пропустить через сигмоидную функцию активации, которая преобразует значение в диапазон от 0 до 1. Это особенно полезно в бинарных задачах предсказания, таких как определение вероятности клика или покупки.

Дополнительно, для улучшения качества рекомендаций можно применять нормализацию эмбеддингов, регуляризацию или обучение с учетом отрицательных примеров. На рисунке 2 представлена схема этого процесса.

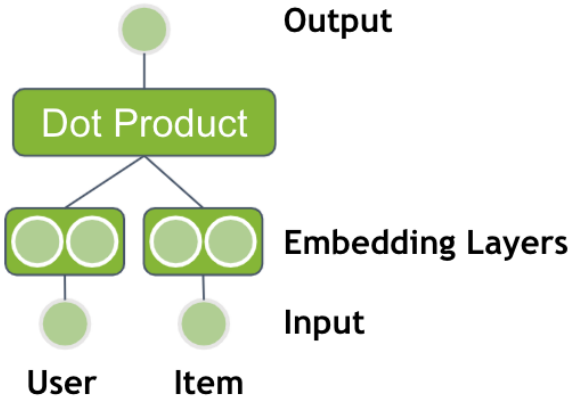



*Рисунок 2.* Нейронная сеть с двумя эмбеддинг-таблицами и *dot*-продуктом на выходе

Этот метод эквивалентен матричной факторизации или методу наименьших чередующихся квадратов (*Alternating Least Squares, ALS*).

Однако производительность нейронных сетей основана на более глубоких архитектурах с несколькими нелинейными слоями. Можно расширить предыдущую модель, подавая выход эмбеддинг-слоев через несколько полносвязных слоев с активациями *ReLU*.

Выбор дизайна заключается в том, как объединить два эмбеддинг-вектора. Можно либо объединить только эмбеддинг-векторы, либо поэлементно умножить векторы друг на друга, по аналогии с *dot*-продуктом. За выводом следуют несколько скрытых слоев (рисунок 3).

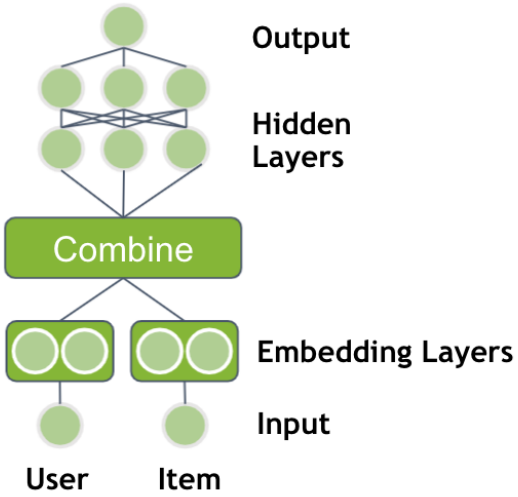



*Рисунок 3.* Нейронная сеть с несколькими полносвязными слоями

До сих пор в качестве входных данных использовались только идентификатор пользователя и идентификатор продукта, но зачастую в распоряжении имеется гораздо больше информации. Дополнительной информацией о пользователе может быть пол, возраст, город (адрес), время с момента последнего посещения или кредитная карта, использованная для оплаты. У продукта-товара обычно есть бренд, цена, категории или количество, проданное за последние 7 дней. Эта сопутствующая информация может помочь модели более успешно выполнять обобщение. Модифицировав нейронную сеть, можно использовать дополнительные характеристики в качестве входных данных (рисунок 4).

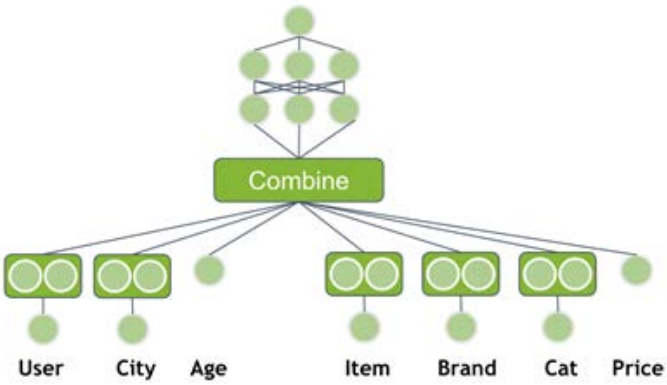



*Рисунок 4.* Нейронная сеть с метаинформацией и несколькими полносвязными слоями

Базовой моделью для разрабатываемой рекомендательной системы на основе глубокого обучения была выбрана *DLRM* от *Facebook* [3]. Она имеет структуру, аналогичную архитектуре нейронной сети с метаданными (рисунок 4), но обладает некоторыми специфическими отличиями (рисунок 5).

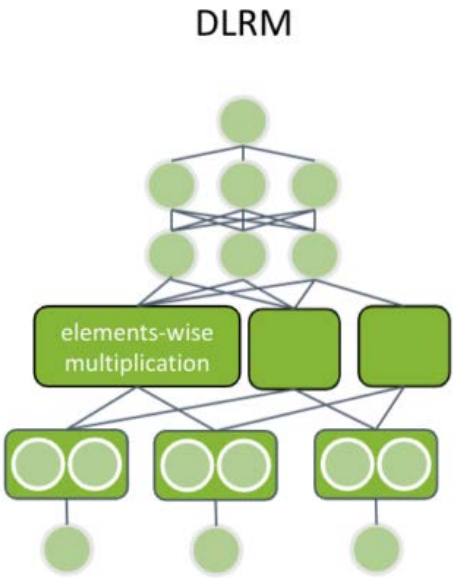



*Рисунок 5.* Архитектура *DLRM*

Важным аспектом исследования является рассмотрение рабочего процесса модели *DLRM*. Модель использует встраивания для обработки разреженных признаков, которые представляют категориальные данные, и многоуровневые перцептроны (*MLP*) для обработки плотных признаков. Эти признаки взаимодействуют с помощью статистических методов, предложенных в модели. В конечном итоге она находит вероятность события, выполняя пост-обработку взаимодействий с другим *MLP* (рисунок 6).

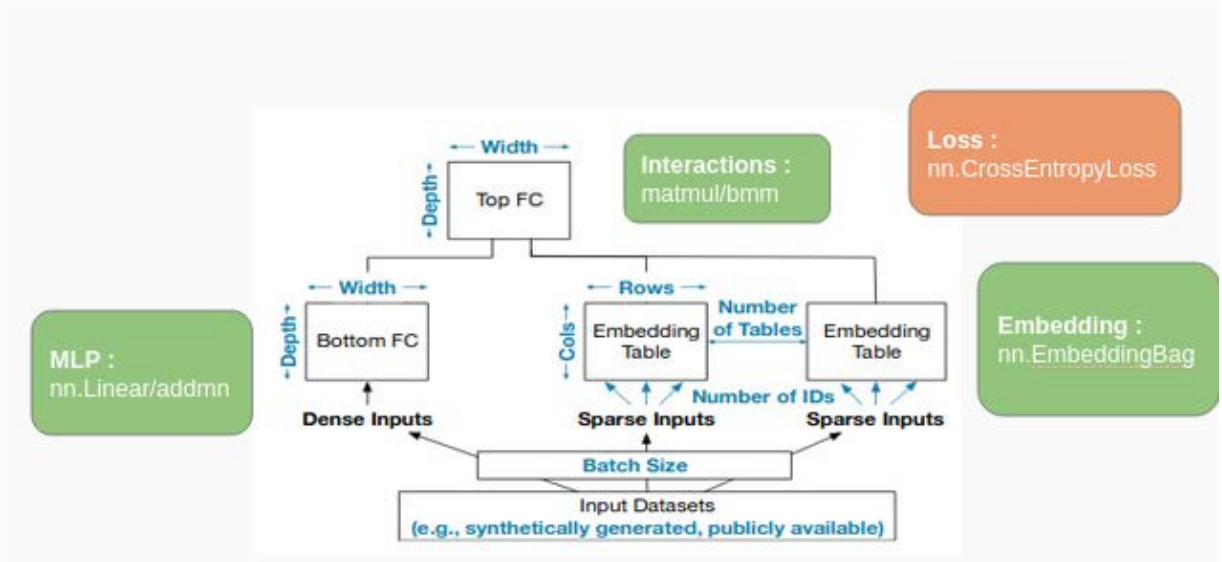



*Рисунок 6.* Рабочий процесс модели *DLRM*

В архитектуре модели можно выделить следующие основные аспекты: встраивания, матрица факторизации, машины факторизации и многоуровневый перцептрон (*MLP*).

Встраивание – это отображение концепций, объектов или элементов в векторное пространство. В контексте нейронных сетей встраивания – это низкоразмерные, обучаемые,

непрерывные векторные представления дискретных переменных. Они позволяют моделям машинного обучения выявлять скрытые закономерности и семантические связи между объектами.

Они широко применяются в областях, где необходимо преобразовать дискретные категории в числовые представления для последующего анализа и обработки. В рекомендательных системах встраивания используются для представления пользователей и товаров в общем пространстве, что позволяет находить скрытые предпочтения и предлагать более точные рекомендации.

В чем стоит преимущество использования встраивания вместо других вариантов, таких как списки разреженных элементов? Встраивания уменьшает размерность категориальных переменных и осмысленно представляет категории в абстрактном пространстве. Можно измерять расстояние между встраиваниями более осмысленным способом. Встраивания элементов представляют разреженные признаки в некотором абстрактном пространстве, релевантном для модели, в то время как целые числа представляют упорядочивание входных данных. Вектор встраивания проецирует пространство $n$-мерных элементов в $d$-мерные векторы встраивания, где $n \gg d$.

Пример встраивания представлен на рисунке 7.

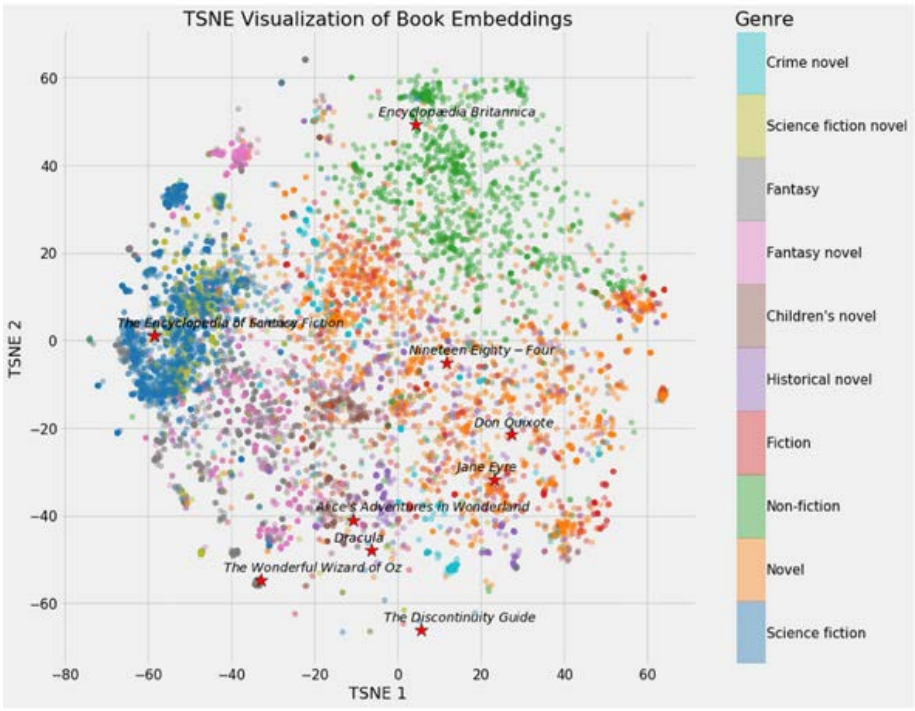



*Рисунок 7.* Визуализация встраивания

Матрица факторизации – это техника, которая принадлежит к классу алгоритмов совместной фильтрации, используемых в системах рекомендаций. Она позволяет выявлять скрытые паттерны взаимодействий пользователей с элементами, что делает рекомендации более точными и персонализированными.

Алгоритмы матричной факторизации работают, разлагая матрицу взаимодействий пользователей и элементов на произведение двух матриц меньшей размерности (рисунок 8). Эти матрицы представляют собой эмбеддинги пользователей и элементов в общем латентном пространстве признаков. Взаимодействие между пользователем и элементом можно оценить с помощью скалярного произведения их эмбеддингов.

Несмотря на свою эффективность, матричная факторизация имеет ограничения, например, сложность работы с разреженными и динамически изменяющимися данными.

Для их преодоления современные рекомендательные системы часто дополняют матричные разложения методами глубокого обучения и гибридными моделями.

Машины факторизации (*FM*) – это улучшенная версия матричной факторизации. На самом деле *FM* можно рассматривать как обобщение нескольких классических методов машинного обучения. При определенных параметрах они могут вести себя как линейные модели, матричная факторизация или полиномиальные модели, что делает их гибким и мощным инструментом для анализа данных. Машины факторизации предназначены для эффективного захвата взаимодействий между признаками в высокоразмерных разреженных данных. Они способны оценивать взаимодействия в разреженных условиях, поскольку они нарушают независимость взаимодействий между параметрами, факторизуя их. Кроме того, машины факторизации включают взаимодействия второго порядка в линейную модель с категориальными данными, определяя модель в следующем виде, рисунок 9.

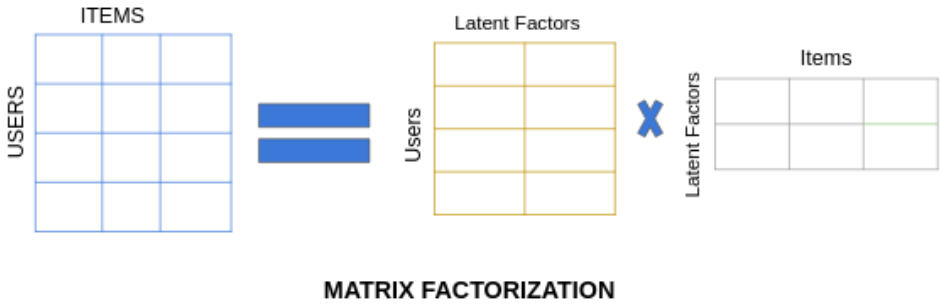



*Рисунок 8.* Визуализация встраивания

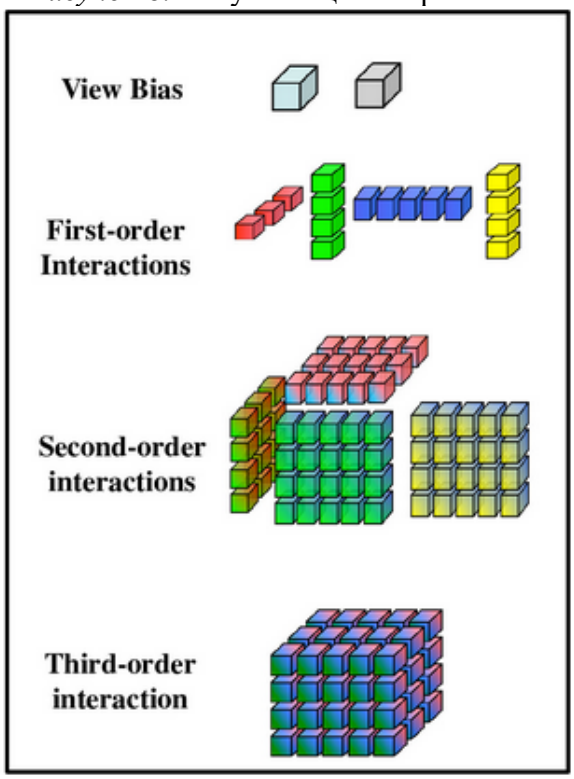



*Рисунок 9.* Различные порядки матриц взаимодействия

*FM* факторизуют матрицу взаимодействий второго порядка в ее латентные факторы (или вектора встраивания), как в матричной факторизации, что более эффективно обрабатывает разреженные данные. Немаловажно, что они значительно снижают

сложность взаимодействий второго порядка, захватывая только взаимодействия между парами различных векторов встраивания, что дает линейную вычислительную сложность.

Многоуровневый перцептрон (*MLP*) – это класс нейронных сетей с прямой связью, состоящих как минимум из трех слоев узлов: входного слоя, скрытого слоя и выходного слоя. Все узлы, кроме входных, являются нейронами, которые используют нелинейные активационные функции.

Эти методы были использованы для захвата более сложных взаимодействий. *MLP* с достаточной глубиной и шириной могут подгонять данные с произвольной точностью.

**Описание улучшенной модели персонализированной рекомендательной системы.** Моделирование рекомендательной системы с использованием глубокого обучения имеет в своей основе решение задачи предсказания предпочтений пользователей на основе их истории взаимодействий с системой. В данной статье модель *DLRM* будет модифицирована для решения задач с использованием регуляризации, оптимизации вычислений, гибридной фильтрации и учета дополнительных признаков пользователей и товаров. Все эти улучшения направлены на повышение эффективности персонализированных рекомендаций, а также адаптации работы рекомендательной системы.

Первое улучшение модели *DLRM* состоит в стабилизации процесса обучения. Для этого используются методы *Dropout* и *Batch Normalization (BatchNorm)*. Это методы регуляризации, которые помогают улучшить обобщающие способности модели.

*Dropout* заключается в случайном исключении некоторых нейронов на каждом шаге обучения, что предотвращает переобучение, улучшает обобщение и помогает бороться с избыточной адаптацией модели к данным. Математически, *Dropout* можно представить как операцию, которая умножает выход нейрона на случайную бинарную маску $m$ размерности $n$, где $n$ – количество нейронов, формула (1):

$$y^{drop} = m \odot y, \tag{1}$$

где $y$ – выход нейрона; $m$ – маска; $\odot$ – операция покомпонентного умножения.

*Batch Normalization* (*BatchNorm*) нормализует выходы слоев нейронной сети на каждом шаге обучения. Она вычисляет среднее и стандартное отклонение на батче данных и использует их для нормализации каждого слоя, формула (2):

$$\hat{x} = \frac{x - \mu_B}{\sigma_B}, \tag{2}$$

где $\mu_B$ – среднее по батчу; $\sigma_B$ – стандартное отклонение; $x$ – входные данные слоя.

После нормализации данные могут быть масштабированы с помощью обучаемых параметров $\gamma$ и $\beta$, формула (3):

$$y = \gamma \hat{x} + \beta. \tag{3}$$

Это помогает ускорить обучение и повышает стабильность, уменьшив проблему с исчезающими или взрывающимися градиентами.

В контексте рекомендательных систем краеугольным камнем является вопрос оптимизации времени предсказаний. Для этих целей будет использован *Quantization* – это метод преобразования числовых значений в модели с более высокой точностью в представление с меньшей точностью (например, 32-битное представление заменяется 8-битным). В математическом выражении для квантования можно записать следующее преобразование, формула (4):

$$Q(x) = round\left(\frac{x}{\Delta}\right) * \Delta, \quad (4)$$

где Δ – шаг квантования; *x* – исходное значение; *round(·)* – операция округления.

Это позволяет уменьшить размер модели и ускорить вычисления, не значительно ухудшая точность предсказания. В контексте рекомендательных систем квантование может быть использовано для ускорения инференса, особенно при обработке больших данных и работе в реальном времени.

Оптимизация модели производится через *TorchScript* и *TensorRT*.

*TorchScript* позволяет превращать модель, написанную в *PyTorch*, в статическую форму, которую можно оптимизировать и выполнять без *Python*. В этом случае модель представляется как вычислительный граф, который можно оптимизировать для быстрого выполнения.

Математически, *TorchScript* преобразует исходный динамический граф *PyTorch* в статический граф вычислений *G(θ)*, где $\theta$ – параметры модели. Преобразование в статический граф позволяет использовать такие оптимизации, как удаление лишних операций, слияние слоев и других техник, которые ускоряют инференс.

*TensorRT* представляет собой библиотеку для оптимизации выполнения нейронных сетей на графических процессорах (*GPU*) от *NVIDIA*. *TensorRT* использует оптимизации, такие как слияние слоев, использование пониженной точности (*FP16, INT8*), аппаратная оптимизация для *GPU*. Это позволяет ускорить выполнение операций свертки, умножения матриц и других типичных операций в нейронных сетях.

Для улучшения точности рекомендаций важным аспектом является добавление дополнительных признаков, таких как история покупок или поведение на сайте. Это позволяет дополнительно персонализировать рекомендации, учитывая не только текущие предпочтения пользователя, но и его поведение в прошлом.

Категориальные эмбеддинги для пользователя и товара позволяют эффективно представлять их характеристики в виде низкоразмерных векторов, что значительно сокращает размер пространства признаков.

Вместо классической коллаборативной фильтрации, которая реализуется в базовой модели *DLRM*, в улучшенной модели персонализированной рекомендательной системы будет использоваться гибридная фильтрация в совокупности с матричной факторизацией. Матричная факторизация является одним из ключевых методов и заключается в разложении матрицы взаимодействий пользователей и товаров *R* на две низкоразмерные матрицы, формула (5):

$$R \approx P * Q^T, \quad (5)$$

где *P* – матрица скрытых факторов пользователей; *Q* – матрица скрытых факторов товаров.

Эти матрицы обучаются таким образом, чтобы минимизировать ошибку предсказания, формула (6):

$$\min_{P,Q} \sum_{i,j} (r_{ij} - P_i Q_j^T)^2, \quad (6)$$

где $r_{ij}$ – оценка взаимодействия; $P_i$ и $Q_j$ – скрытые факторы для пользователя *i* и товара *j*.

**Заключение.** Таким образом, математическое описание улучшенной модели персонализированной рекомендательной системы на основе *DLRM* включает в себя широкий набор методов и техник, таких как регуляризация (*Dropout, BatchNorm*), оптимизация вычислений (*Quantization, TorchScript, TensorRT*), использование эмбеддингов для представления категориальных признаков и гибридная фильтрация с

матричной факторизацией. Эти методы в совокупности обеспечивают высокую точность, скорость и стабильность работы модели, что особенно важно при обработке больших и сложных данных в реальных задачах рекомендательных систем.

## Список литературы

[1] Михальков, М. Д. Преимущества и недостатки эмпирических моделей интеллектуальных систем поддержки принятия решений в области электронной коммерции = Advantages and disadvantages of empirical models of intelligent decision support systems in the field of e-commerce / М. Д. Михальков // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей X Международной научно-практической конференции, Минск, 13 марта 2024 г. : в 2 ч. Ч. 2 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2024. – С. 201–207.

[2] Михальков, М. Д. Модели и методы построения интеллектуальных систем поддержки принятия решений / М. Д. Михальков // Информационные технологии и системы 2023 (ИТС 2023) = Information Technologies and Systems 2023 (ITS 2023) : материалы Международной научной конференции, Минск, 22 ноября 2023 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск : БГУИР, 2023. – С. 67–68.

[3] Keith Bonawitz, Hubert Eichner, Wolfgang Grieskamp, Dzmitry Huba, Alex Ingerman, Vladimir Ivanov, Chloe Kiddon, Jakub Konecny, Stefano Mazzocchi, Brendan McMahan, Timon Van Overveldt, David Petrou, Daniel Ramage, and Jason Roselander. Towards federated learning at scale: System design. In Proc. 2nd Conference on Systems and Machine Learning (SysML), 2019.

## Авторский вклад

**Михальков Максим Дмитриевич** – моделирование рекомендательной системы, предназначенной для платформ электронной коммерции, рассмотрение различных архитектурных решений, формализация основных компонентов системы и разработка математических моделей, описывающих работу рекомендательных алгоритмов.

**Нестеренков Сергей Николаевич** – постановка задачи исследования, анализ полученных результатов.

# MODELING A PERSONALISED RECOMMENDATION SYSTEM BASED ON DEEP LEARNING

***M.D. Mikhalkov***
*PhD student at the Department of Information Technology Software, Faculty of Computer Systems and Networks, BSUIR*

***S.N. Nesterenkov***
*PhD of Technical Sciences, Associate Professor at the Department of Information Technology Software, Faculty of Computer Systems and Networks, BSUIR*

**Abstract.** In this article, a modeling of a recommender system designed for e-commerce platforms was conducted. During the modeling process, various architectural solutions were considered, the main components of the system were formalized, and mathematical models describing the operation of recommender algorithms were developed.

Particular attention was paid to the selection and analysis of recommendation methods, including hybrid models and deep learning. This article presented a conceptual modeling of the system, a description of its logical and functional structure, a selection of formalized methods and algorithms, and an analysis of possible operating scenarios.

**Keywords**: Recommender systems, deep learning, DLRM, matrix factorization, hybrid filtering.

УДК 004.822+89

# СТАТИСТИЧЕСКИЙ АНАЛИЗ И МОДЕЛИРОВАНИЕ СВОЙСТВ НАНОМАТЕРИАЛОВ МЕТОДАМИ LARGE LANGUAGE MODELS, AGENTIC AI & MACHINE LEARNING

***Н.А. Шиманский***
*Соискатель кафедры физики твердого тела и нанотехнологий физического факультета БГУ, системный архитектор компании Andersen Lab (ПВТ)*
nikita.shymanski@gmail.com

***А.В. Баглов***
*Старший научный сотрудник НИЛ энергоэффективных материалов и технологий физического факультета БГУ*
baglov@bsu.by

***Л.С. Хорошко***
*Ведущий научный сотрудник НИЛ энергоэффективных материалов и технологий физического факультета БГУ, канд. физ.-мат. наук, доцент*
khoroshko@bsu.by

***Н.А. Шиманский***
*Окончил Белорусский государственный университет. Занимается проектированием и разработкой IT-решений в области бизнеса и науки. Область научных интересов – разработка программных средств для оптимизации решения прикладных задач в области наноматериаловедения с применением Machine Learning & Generative AI.*

***А.В. Баглов***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Проводит научные исследования в рамках компьютерного моделирования электронной структуры перспективных материалов.*

***Л.С. Хорошко***
*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой наноразмерных катализаторов фотостимулированных реакций и управляемым синтезом наноматериалов.*

**Аннотация**: В данной работе рассматривается возможность комбинированного использования генеративного машинного обучения и классических нейронных сетей в качестве методов работы с большими данными для предиктивного анализа электронных свойств полупроводниковых наноматериалов, не ограничивая при этом общность данного подхода для иных кристаллических материалов и гетероструктур.

**Ключевые слова**: машинное обучение, нейронные сети, большие языковые модели, большие данные, наноматериалы, статистический анализ, предсказательный анализ.

**Введение.** Успешное развитие современного материаловедения, в особенности, в области наноматериалов, практически невозможно без использования вычислительных средств и методов компьютерной автоматизации. Стремительное развитие нейронных сетей и программных интерфейсов для них (*Discriminative AI, Generative AI*) позволяет анализировать в кратчайшие сроки значительные объемы экспериментальных данных и не только исследовать, но и с высокой достоверностью предсказывать интересующие свойства низкоразмерных материалов, например, стехиометрический состав, электронное строение,

влияние дефектов на структуру и свойства и др. Для оптимизации подобных задач, сокращения времени обработки больших массивов данных, автоматизации анализа полученных результатов и ряда других операций могут быть задействованы современные инструменты *Big Data, Machine Learning, Neural Networks* и *Advanced Analytics.* Использование таких подходов в области материаловедения предоставляет исследователям ряд новых возможностей, в частности, генерации новых экспериментальных моделей наноструктур с предиктивным анализом их электронных свойств, значительно сокращая временные затраты на получение результата по сравнению с традиционными «ручными» режимами моделирования, в которых значительная часть времени может быть затрачена на построение экспериментальной структурной модели и выбор и подготовку входных данных для вычислений. В данной работе предложен подход для оптимизации решения специальной исследовательской задачи – анализа структуры и свойств выбранного материала, который может быть использован для изучения свойств объемных и сверхтонких наноструктур из практически любых неаморфных материалов.

**Постановка задачи.** Эффективность вычислительных экспериментов для исследования и определения структур и электронных свойств наноматериалов в значительной мере обусловлена качеством программной реализации и используемыми вычислительными моделями компьютерного эксперимента. Хорошо зарекомендовали себя, например, такие общедоступные программные комплексы как *VESTA* (*Visualisation for Electronic and Structural Analysis*) и *OpenMX* (*Open source package for Material eXplorer*) [1–3]. Первый применяется для построения пространственных структур и визуализации их электронных свойств, второй – для определения структурных и электронных свойств материалов с использованием теории функционала плотности, теории псевдопотенциала и базиса численных атомно-центрированных орбиталей. С применением результатов моделирования в *OpenMX* можно оценить возможность использования новых или сложных в получении материалов в наноэлектронике и других приложениях, что сокращает затраты и время на проведение реальных физических экспериментов. Практическая трудность использования компьютерного эксперимента, например, в случае полупроводниковых материалов, обусловлена широким разнообразием параметров, характеризующих их морфологические свойства (такие как симметрия кристаллической ячейки, взаимное расположение кристаллографических плоскостей и пространственная ориентация интерфейса в случае двумерных материалов, взаимодействие и деформация слоев для слоистых структур, разнообразие возможных дефектов, особенно, для композитных составов, и др.). Современные вычислительные средства позволяют достаточно быстро проводить масштабные моделирования наноструктур с помощью программных комплексов, однако, подготовка входных данных для вычислений и описание всех возможных вариаций для моделирования в ручном режиме может занимать значительное время, превышающее многократно затраченное на сам вычислительный эксперимент. Для решения описанных проблем требуются более углубленные, комплексные подходы, сочетающие в себе возможности оптимизации и ускорения процессов сбора данных, моделирования и анализа с помощью инструментов *Big Data, Machine Learning & Generative AI.*

Авторами данной работы для оценки и анализа результатов рентгенодифракционного анализа материалов с возможностью прогнозирования свойств наноструктур ранее было применено предиктивное машинное обучение [4, 5]. Данный подход включал в себя создание нейронной сети и её обучение с помощью эталонных образцов дифрактограмм наноструктур с описанными свойствами. Такая нейросеть смогла бы предложить вероятностное определение их характеристик с определенной степенью точности, зависящей от количества циклов обучения и степени подробности обучающего материала, в случае запроса. Однако, для прогнозирования электронных свойств реализация данного подхода оказалась весьма трудоемкой, что обусловлено наличием

неочевидных зависимостей электронных свойств от структурной конфигурации в материалах с понижением размерности [6].

Подход, основанный на использовании больших языковых моделей (*LLM – Large Language Models*) с применением генерации дополненного поиска (*RAG – Retrieval-Augmented Generation*) ранее применялся авторами для попыток экспресс-анализа больших объемов данных для определения ключевых характеристик наноструктур, опубликованных в открытых источниках. Была реализована специальная методология “обогащения” контекста в виде наполнения специализированных баз знаний (*AI Knowledge Bases*) и, соответственно, расширения области знаний фундаментальной модели (в данном решении используется фундаментальная модель *Anthropic Claude 3.5 Sonnet*). Затем при обращении пользовательского запроса в модель происходила сверка с *RAG*-контекстом и, с учетом специализированного контекста из векторной базы данных, осуществлялась генерация ответа пользователю, при этом релевантный поиск по сотням гигабайт векторных данных языковая модель занимал несколько секунд. При использовании описанного подхода пользователь получает ответ в виде наиболее вероятных словосочетаний и предложений, которые описывают интересующие характеристики наноструктур [7].

**Методология и архитектура.** Ряд преимуществ описанных подходов, тем не менее, не может решить одного из трудоемких вопросов, стоящих перед исследователями в области материаловедения, заключающегося в поиске наиболее вероятно подходящих параметров кристаллической структуры материала, комбинация которых обеспечит, во-первых, сходимость расчета, а во-вторых требуемое значение ширины запрещенной зоны исследуемого материала. В рамках данной работы реализуется комбинированный подход, использующий *Generative AI* / *LLM* для накопления, структурирования и классификации открытых данных о наноструктурах, а также классическое машинное обучение (*Discriminative ML*), основанное на известных классифицирующих алгоритмах (*XGBoost, LightGBM, Convolutional Neural Network, LTSM – Long-Term Short Memory*). Следует отметить, что применяется продвинутый подход *LLM Agentic AI* (т. наз. агент), который, в свою очередь, основан на использовании описанной выше фундаментальной модели *LLM* вместе с обогащением *RAG*. При этом агент позволяет создавать сложные сценарии поведения чат-бота, хранение и анализ истории диалогов с пользователем, автоматизацию программных действий, таких, как вызовы *REST API* методов, сохранение данных в *SQL*, трансформация собранных данных в виде таблиц, и др. Эти действия описываются с помощью *NLP* (*Natural Language Processing*), фактически, «человеческим языком», вместо написания сложного программного кода, традиционно основанного на циклах, условном выполнении, и т.д. (рисунок 1).

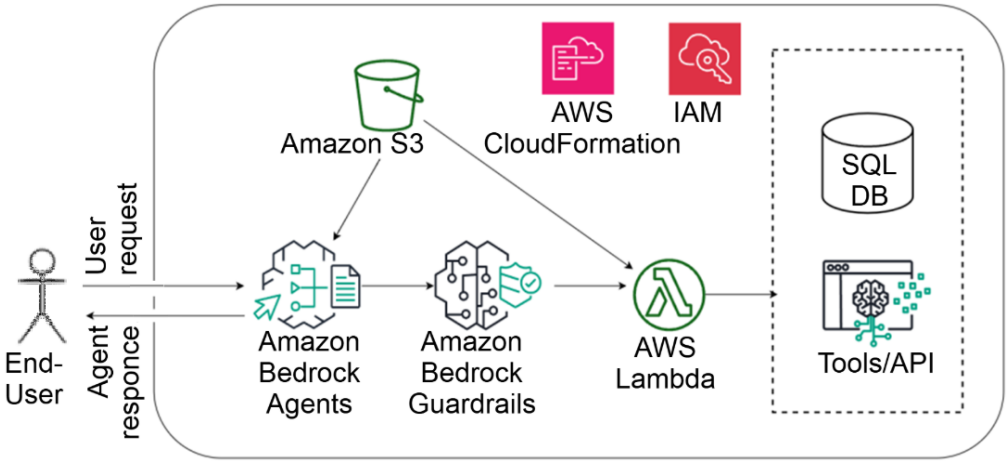



*Рисунок 1*. Архитектура реализации *Agentic AI* на основе облачного сервиса *AWS* с использованием *Bedrock Agents* и *AWS Lambda functions*

В рамках данной работы также разработан пользовательский интерфейс, который позволяет с помощью *LLM*-инструкций (промпт-инжиниринг) и агентов пополнять базу знаний в виде текстовых данных произвольного формата, подходящих под заданную тематику изучения свойств наноструктур. С помощью интерфейса реализуется возможность вести диалог с *AI*-чат-ботом, производить выборки данных по интересующим материалам с возможностью задания нужных критериев (симметрия кристаллической решетки, стехиометрический состав, ширина запрещенной зоны и др.), а также сохранять их в структурированном виде для последующего развития предиктивной нейронной сети.

Описываемая нейронная сеть была создана на основе облачного сервиса *AWS SageMaker*, который является полноценной средой разработки нейронных сетей и моделей с широкими встроенными возможностями по обработке и классификации тренировочных данных (*SageMaker Data Wrangler*), и автоматизацией подготовки, обучения и разворачивания *ML* моделей для вероятностных классификаций (*SageMaker Pipelines*). В представляемом решении данные, получаемые с помощью *Agentic AI* в виде *csv*-файлов проходят обработку в сервисе *SageMaker Data Wrangler*, которая включает классификацию исходных данных, фильтрацию заведомо непригодных данных, обогащение и трансформацию (*k-nearest neighbors, dropout, oversampling*). Подготовленные данные позволяют далее запускать тренировочные процессы для нейронной сети. На рисунке 2 представлена обобщенная схема работы созданной нейронной сети, предоставленная сервисом *AWS SageMaker*.

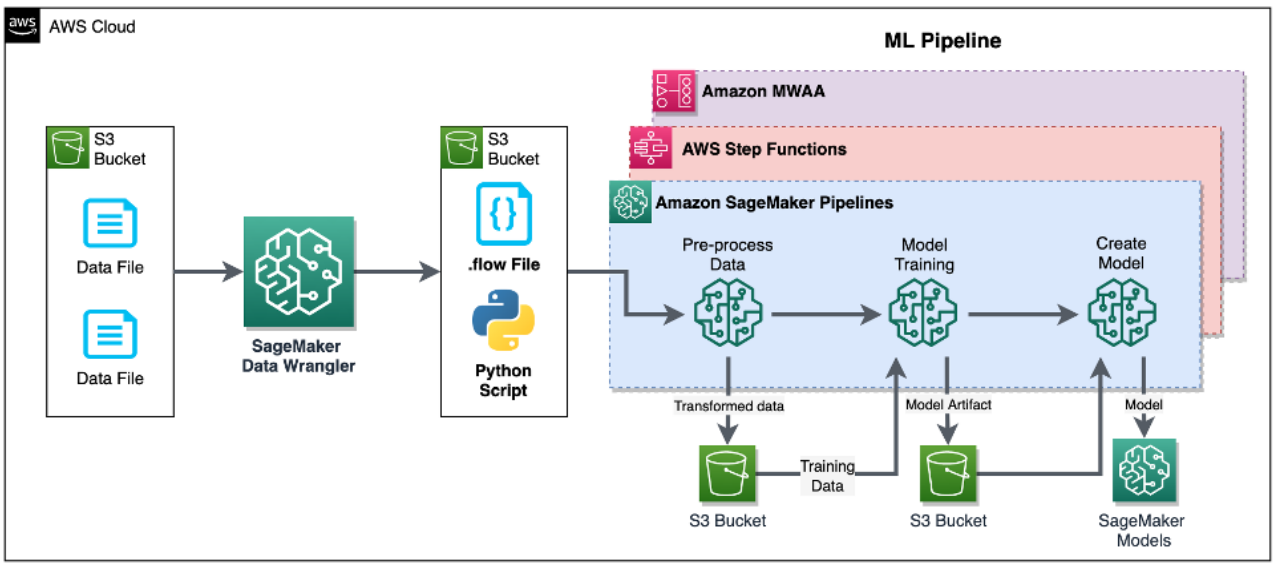



*Рисунок 2*. Автоматизация подготовки и обработки данных, тренировки нейронной сети и доставки предиктивного сервиса с помощью *SageMaker Data Wrangler* и *SageMaker Pipelines* (скриншот рабочего окна конструктора сервиса)

Следует отметить, что классификация атрибутов нейронной модели в виде точных чисел зачастую демонстрирует плохую сходимость (*ML convergence*) и, как следствие, низкую точность предсказаний свойств исследуемых материалов. В данной работе для нивелирования этой проблемы применяли специальный подход группирования численных значений по диапазонам (*numerical binning*), который на порядки уменьшает количество различных уникальных атрибутов модели за счет группирования близких по значению численных свойств. Такой метод позволяет значительно улучшить сходимость тренировки нейронной сети и качество ее предсказаний, но в то же время может быть ухудшена либо потеряна связь диапазона с исходным точным значением численного параметра вследствие нелинейных сложных зависимостей свойств внутри конкретного диапазона чисел (например, с шириной запрещенной зоны), что в свою очередь, также потребует корректировки.

Разработанное программное решение (использован язык *Python*) способно создавать большие массивы файлов с псевдо-описанием различных возможных конфигураций наноструктур, представляющих потенциальный научный интерес. Данные файлы используются в качестве входных данных программного комплекса *OpenMX*, который на их основе определяет существенные свойства материала (в данном исследовании – ширина запрещенной зоны). Поскольку реальное время расчета тысяч всевозможных комбинаций наноструктур с использованием *OpenMX* занимает значительный период времени, данные массивы потенциально интересных конфигураций предварительно проходят предиктивную фильтрацию в сервисе *SageMaker Inference*, который на основе обученной нейронной модели классифицирует входные данные и отсеивает, по его мнению, заведомо непригодные для расчета. Таким образом выборка конфигураций, перспективных для непосредственного расчета, может быть сокращена в десятки раз. Разработанный подход был проверен на совокупности наноструктур на основе широкозонного сегнетоэлектрика титаната бария ($BaTiO_3$). Выборочные контрольные расчеты отобранных сервисом структур показали, что в подавляющем количестве случаев классификация подходящих и неподходящих структур совпадает с диапазонами реальных расчетов ширины запрещенной зоны в пакете *OpenMX*. Таким образом, предварительная выборка способна значительно ускорить проведение подобных исследований и сократить время реального потребления вычислительных ресурсов.

**Заключение.** Прогрессивные вычислительные подходы могут и должны внести значительный вклад в развитие современного материаловедения. Применение нейронных сетей (*Generative AI, LLM*) в качестве больших языковых моделей для сбора, классификации и обобщения свойств наноструктур и наноматериалов в сочетании со статистическим машинным обучением демонстрирует высокий потенциал применения, в особенности, для повышения уровня автоматизации ряда исследовательских задач при значительном сокращении временных и трудовых затрат на проведение вычислительных экспериментов. Предложенный в данной работе подход предусматривает непрерывное дополнение и развитие базы знаний *LLM* модели в рамках специализированного научного контекста (в рассматриваемом случае – в области наноматериаловедения), а также развитие предиктивной нейронной сети для предварительного статистического анализа моделируемых наноматериалов, что позволит значительно снизить временные и аппаратные затраты на моделирование с помощью автоматизированной классификации перспективных конфигураций наноструктур. Данный подход при подборе соответствующей базы знаний может быть расширен также для предиктивного анализа свойств более сложных наноструктур, в частности, бислоев, Ван-дер-Ваальсовских структур, гетеросистем различного назначения, сильнолегированных материалов и т.д.

## Список литературы

[1] Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures / T. Ozaki // Phys. Rev. B. 2003. Vol. 67. P. 155108.

[2] Ozaki, T. Numerical atomic basis orbitals from H to Kr / T. Ozaki, H. Kino // Phys. Rev. B: Condens. Matter Mater. Phys. 2004. Vol. 69. P. 195113.

[3] Ozaki, T. Efficient projector expansion for the ab initio LCAO method / T. Ozaki, H. Kino // Phys. Rev. B. 2005. Vol. 72. P. 045121.

[4] Шиманский, Н.А. Автоматизация обработки результатов исследования структуры и свойств наноматериалов / Н.А. Шиманский, А.В. Баглов, Л.С. Хорошко // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей IX Международной научно-практической конференции, Минск, 17–18 мая 2023 г. : в 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2023. – С. 296-300.

[5] Шиманский, Н. А. Автоматизация обработки результатов исследования структуры материалов / Н.А. Шиманский, А.В. Баглов, Л.С. Хорошко // Information Tehnologies and Systems 2023 (ITS 2023) : материалы международной научной конференции, Минск, Беларусь, 22 ноября / ред. Л. Ю. Шилин [и др.]. – Минск : БГУИР, 2023. – С. 207.

[6] Шиманский, Н.А. Автоматизация обработки результатов исследования структуры наноматериалов с использованием методов BIG DATA & MACHINE LEARNING / Н.А. Шиманский, А.В. Баглов // Математические методы и компьютерное моделирование в ФКС. – Гродно: ГрГУ, 2024. – С. 159.

[7] Шиманский, Н.А. Экспресс-анализ структурных и электронных свойств наноматериалов методами Big Data, Large Language Models & Generative AI / Н.А. Шиманский, А.В. Баглов, Л.С.Хорошко // Компьютерное проектирование в электронике = Electronic Design Automation :сб. трудов Междунар. науч.-практ. конф. (Республика Беларусь, г. Минск, 28 ноября 2024 г.) / редкол. : В. Р. Стемпицкий [и др.]. – Минск : БГУИР, 2024. – С. 100–103.

[8] Nemeth, M. The Comparison of Machine-Learning Methods XGBoost and LightGBM to Predict Energy Development / M. Nemeth, D. Borkin, G. Michalconok // In: Computational Statistics and Mathematical Modeling Methods in Intelligent Systems. CoMeSySo 2019. Advances in Intelligent Systems and Computing [Silhavy, R., Silhavy, P., Prokopova, Z. (eds)] / Springer Cham. 2019. Vol 1047. P.208–215.

## Авторский вклад

**Шиманский Никита Андреевич** – концептуализация и разработка решения, описание принципов работы агента, работа с облачными сервисами, написание и адаптация программного кода, тестирование решения, формирование структуры статьи.

**Баглов Алексей Викторович** – постановка задачи исследования, описание принципа работы решения, анализ и верификация результатов, тестирование решения, обеспечение проверки результатов в пакете *OpenMX*.

**Хорошко Людмила Сергеевна** – постановка задачи исследования, концептуализация решения, определение структуры исследования, тестирование решения, анализ и верификация результатов, формирование структуры статьи, общее руководство и менеджмент проекта.

# STATISTICAL ANALYSIS AND MODELING OF NANOMATERIALS PROPERTIES USING LARGE LANGUAGE MODELS, AGENTIC AI & MACHINE LEARNING

***N.A. Shymanski***
*Postgraduate of Department of Solid State Physics and Nanotechnologies, Faculty of Physics, BSU;*
*Solutions Architect of Andersen Lab (HTP)*

***A.V. Baglov***
*Senior researcher in the R&D Lab of Energy-effective materials and technologies, Faculty of Physics, BSU*

***L.S. Khoroshko***
*Lead researcher in the R&D Lab of Energy-effective materials and technologies, Faculty of Physics, BSU*

**Abstract.** This paper investigates the possibility of the combination of generative machine learning and classical neural networks as methods for working with Big Data for the predictive analysis of the electronic properties of semiconductor nanomaterials, without limiting the generality of this approach to other crystalline materials and heterostructures.

**Keywords**: machine learning, neural networks, inference, large language models, agentic ai, big data, nanomaterials, predictive analysis.

УДК 004.65

# МЕТОДИЧЕСКИЕ ПОДХОДЫ К ОЦЕНКЕ ЭФФЕКТИВНОСТИ ВНЕДРЕНИЯ ТЕХНОЛОГИЙ BIG DATA

***В.С. Князькова***
*Доцент кафедры менеджмента БГУИР,*
*кандидат экономических наук, доцент*
knyazkova@bsuir.by

***В.С. Князькова***
*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием проблем цифровизации и цифровой трансформации, оценки их экономической эффективности.*

**Аннотация.** Современные информационные технологии позволяют получать, хранить, анализировать большое число данных. Использование Big Data открывает широкие возможности для бизнеса, науки, государственного управления и других сфер. Исследованы основные вызовы использования технологий Big Data. Изучены основные статьи затрат, проанализированы возможные результаты от внедрения технологий Big Data, что следует учитывать при проведении комплексной оценки экономической эффективности внедрения.

**Ключевые слова:** Big Data, экономическая эффективность, совокупная стоимость владения.

**Введение.** Современные информационные системы позволяют аккумулировать огромные массивы разнородных данных, объемы которых увеличиваются ежегодно в несколько раз. Высокий уровень использования мобильных устройств, развитие сетевых и других цифровых технологий, включая Интернет-вещей, искусственный интеллект, мобильные сети пятого поколения, создают предпосылки для формирования единого киберфизического пространства, в котором данные играют важную роль. Навыки применения существующих (и разработки новых) технологий и инструментов обработки и анализа данных наряду с возможностью получения релевантной информации становится важнейшим конкурентным преимуществом субъектов хозяйствования, позволяющих им существенно повысить эффективность своей деятельности и точность прогнозирования. По некоторым данным в настоящее время анализируется менее 1% накапливаемых данных, в связи с чем потенциал развития «экономики данных» представляется огромным.

Вместе с тем на наш взгляд в научной литературе существует определенный пробел в исследованиях, посвященных оценке экономического эффекта и эффективности внедрения технологий Big Data.

**Проблемы и вызовы при использовании технологий Big Data.** Использование технологий Big Data, несмотря на их огромный потенциал, сопряжено с рядом проблем и вызовов [1]. Их можно условно сгруппировать по следующим направлениям – технические, организационные, этические, юридические, экологические, вызовы информационной безопасности, вызовы, связанные с аналитикой данных, вызовы масштабируемости.

*Технические вызовы:*

– объем данных: большие объемы данных требуют мощных вычислительных ресурсов и специализированных систем хранения;

– скорость обработки данных: часть данных поступает в режиме реального времени (например, данные с датчиков IoT), что требует быстрой обработки;

– сложность интеграции: данные могут поступать из разных источников и иметь разный формат (например, структурированные / неструктурированные / полуструктурированные), что усложняет их интеграцию и анализ;

– качество данных: данные могут быть неполными, устаревшими или содержать ошибки, что снижает точность анализа.

*Организационные вызовы:*

– нехватка квалифицированных специалистов: работа с Big Data требует знаний в области data science, машинного обучения, статистики и программирования. Нехватка таких специалистов является серьезной проблемой;

– высокая стоимость внедрения: создание инфраструктуры для хранения и обработки больших данных требует значительных инвестиций в оборудование, программное обеспечение и обучение персонала;

– сопротивление изменениям: внедрение Big Data-решений может встретить сопротивление со стороны сотрудников, особенно если они не готовы к изменениям в рабочих процессах.

*Этические вызовы:*

– конфиденциальность данных: сбор и анализ больших объемов данных могут нарушать приватность пользователей. Например, данные из социальных сетей или медицинские записи могут быть использованы неэтично;

– дискриминация и предвзятость: алгоритмы, обученные на исторических данных, могут воспроизводить существующие предрассудки и дискриминацию (например, в кредитном скоринге или найме сотрудников);

– прозрачность: пользователи часто не понимают, как их данные собираются, анализируются и используются, что вызывает вопросы о прозрачности процессов.

*Юридические вызовы:*

– соблюдение законодательства: в разных странах действуют строгие законы о защите данных и несоблюдение этих законов может привести к штрафам и репутационным потерям;

– права на данные: возникают вопросы о том, кому принадлежат данные, особенно когда они собираются из нескольких источников и/или используются третьими сторонами.

*Вызовы информационной безопасности:*

– утечка данных: большие объемы данных привлекают хакеров, что увеличивает риск утечек и кибератак;

– защита данных: обеспечение безопасности данных на всех этапах (сбор, хранение, обработка, передача) требует значительных усилий и ресурсов.

*Вызовы, связанные с аналитикой данных:*

– перегруженность информацией: большие объемы данных могут привести к тому, что важная информация будет потеряна среди "шума";

– неточность прогнозов: если данные низкого качества или алгоритмы анализа не настроены правильно, результаты могут быть неточными или вводящими в заблуждение;

– интерпретация результатов: даже при наличии точных данных их интерпретация может быть сложной, особенно если аналитики не обладают достаточной экспертизой в предметной области.

*Экологические вызовы:*

– энергопотребление: центры обработки данных потребляют огромное количество энергии, что негативно влияет на окружающую среду;

– утилизация оборудования: частая замена серверов и другого оборудования создает проблему электронных отходов.

**Экономический инструментарий оценки эффективности внедрения технологий Big Data.** Полагаем, что управление вышеназванными вызовами будет более успешным, если внедрение технологий Big Data будет сопровождаться оценкой экономической эффективности данного мероприятия.

В экономической науке разработаны четкие методологические критерии оценки экономической эффективности принимаемых решений. В основе экономического подхода к оценке эффективности лежат следующие принципы: принцип проектных затрат; принцип приростных затрат; принцип альтернативных затрат; принцип невозвратных издержек; принцип денежных потоков. Данные принципы основываются на определении упущенной выгоды при вложении денег в конкретный проект, а также на соизмеримости денежных средства, полученных и уплаченных в разное время.

На практике для оценки вышеуказанных аспектов используются такие показатели, как чистая приведённая (текущая) стоимость, срок окупаемости (простой и динамический), внутренняя норма рентабельности, индекс доходности и т.п. По сути все эти показатели так или иначе связаны с денежным потоком – разницей между совокупными результатами и затратами (в контексте данной статьи – от внедрения технологии Big Data). Базовым показателем здесь является чистая приведенная стоимость (Net Present Value – NPV), которая определяется как разница между приведенными к началу проекта притоками денежных средств и инвестициями [2]:

$$NPV = \sum_{t=0}^{n} S_t(1 / (1 + r))^t - \sum_{t=0}^{n} I_t(1 / (1 + r))^t = \sum_{t=0}^{n} C_t(1 / (1 + r))^t,$$

где $S_t$ – поток денежных средств проекта в $t$-м году (приток денежных средств);

$I_t$ – инвестиции (отток денежных средств) в t-м году;

$C_t$ – денежный поток, с учетом знака.

$r$ – ставка дисконтирования.

Размер инвестиций, которые требуются для внедрения технологии Big Data, собственно, как и любого технологического / программного решения, целесообразно рассчитать при помощи методики совокупной стоимости владения (ССВ) информационной системой. В соответствии с ней в совокупные затраты на любую технологию включаются не только затраты на ее приобретение (например, цена пакета программных продуктов), а также т.н. капитальные затраты (стоимость системы защиты информации, устройств передачи данных и линий связи, стоимость технической инфраструктуры, необходимой для надежного функционирования системы, пр.). Также необходимо учесть т.н. текущие или операционные затраты: затраты на оплату труда сотрудников, занятых непосредственно сопровождением, обслуживанием и управлением информационной системой, сетями и устройствами хранения данных в процессе жизненного цикла используемых технологий / программных продуктов, стоимость обучения и повышения квалификации указанных сотрудников, услуги электросвязи, затраты на аренду помещения для серверной комнаты (при необходимости), обслуживание данного помещения, пр. [3]. Причем вышеуказанные затраты рассчитываются на весь период использования технологии.

Приток денежных средств обеспечивается засчет получения эффекта. При оценке эффективности реализации запланированных мероприятий принято различать следующие виды эффектов: экономический эффект, научно-технический эффект, социальный эффект, экологический эффект [4–5]. *Экономический эффект* как правило рассматривается с точки зрения повышения эффективности финансовой деятельности субъекта хозяйствования, рационального использования ресурсов, увеличения добавленной стоимости, динамике прибыли. *Научно-технический эффект* выражается в приращении новых знаний; создании условий для разработки и внедрения перспективных технологий, а также их масштабирования. *Экологический эффект* главным образом заключается в поддержке

процессов цифровой трансформации, которые, в свою очередь, способствуют снижению негативного экологического воздействия на окружающую среду, в частности, в результате сокращения натуральных объемов выбросов загрязняющих веществ, включая и парниковые газы. *Социальный эффект* проявляется в повышении качества жизни людей, что подразумевает: появление новых нематериальных ценностей (культурных, этических, эстетических); улучшение жилищных и культурно-бытовых условий; улучшение здоровья; улучшение условий труда, снижение производственного травматизма, снижение рисков для здоровья и жизни персонала в результате улучшений условий производства; создание условий для самореализации, развития человеческого потенциала, освоения новых знаний и компетенций.

**Заключение.** Big Data представляет собой мощный инструмент для бизнеса, науки и общества. Успешное использование технологий Big Data требует комплексного подхода, включающего технические, организационные и этические аспекты. Кроме того, важно проводить экономическую оценку эффективности внедрения данных технологий. Процесс внедрения Big Data сложен, он требует тщательного планирования, оценку необходимых инвестиций, а также зачастую и изменений в организационной культуре. Однако при правильном подходе компании могут получить значительные преимущества, такие как улучшение принятия решений, оптимизация процессов и создание новых продуктов на основе данных. Ключевым фактором успеха является четкое понимание бизнес-целей и готовность к изменениям.

**Список литературы**

[1] Беляцкая, Т. Н. Управление электронной экономикой / Т. Н. Беляцкая // Наука и инновации. – 2018. – № 5. – С. 49–55.

[2] Баев, Л.А. Основы анализа инвестиционных проектов: учебное пособие / Л.А. Баев. – Челябинск: Издательский центр ЮУрГУ, 2017. – 287 с.

[3] Методические рекомендации по расчету совокупной стоимости владения информационной системой для государственных органов и организаций [Электронный ресурс]. – Режим доступа: https://becloud.by/upload/iblock/5f4/Metodika_opredelenia_TCO.pdf. – Дата доступа: 01.03.2025.

[4] Бучаева, С.А. Виды экономических эффектов и особенности их оценки для инноваций / С.А. Бучаева, М.М. Гаджиев // π-Economy. – 2012. – № 4 (151). – С. 225–229.

[5] Князькова, В. С. Оценка эффективности повышения цифровой грамотности персонала в условиях цифровой трансформации / В. С. Князькова // Цифровая трансформация. – 2023. – Т. 29, № 1. – С. 23–31.

# METHODOLOGICAL APPROACHES TO ASSESSING THE EFFICIENCY OF IMPLEMENTING BIG DATA TECHNOLOGIES

***V.S. Knyazkova***
*Associate Professor of the Department of Management of BSUIR,*
*PhD of Economic Sciences, Associate Professor*

**Abstract.** Modern information technologies allow receiving, storing, and analyzing a large amount of data. The use of Big Data opens up wide opportunities for business, science, public administration, and other areas. The main challenges of using Big Data technologies are studied. The main cost items are studied, the possible results from the implementation of Big Data technologies are analyzed, which should be taken into account when conducting a comprehensive assessment of the economic efficiency of implementation.

**Keywords**: Big Data, economic efficiency, total cost of ownership.

УДК 658.8:004.9

# ОЦЕНКА ЭФФЕКТИВНОСТИ ГЕНЕРАТИВНОЙ ПОЛИТИКИ ПРИ ОПТИМИЗАЦИИ ПРОЕКТНЫХ РЕШЕНИЙ

***В.Ф. Алексеев***
*Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент*
alexvikt.minsk@gmail.com

***Д.В. Лихачевский***
*Декан факультета компьютерного проектирования БГУИР, кандидат технических наук, доцент*
likhachevskyd@bsuir.by

***Г.А. Пискун***
*Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент*
piskunbsuir@gmail.com

***В.Ф. Алексеев***

*Окончил Минский радиотехнический институт. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.*

***Д.В. Лихаческий***

*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием проблем радиочастотной идентификации объектов, организацией учебного и научно-исследовательского процессов в техническом университете.*

***Г.А. Пискун***

*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.*

**Аннотация.** В статье рассматривается применение генеративной политики как эффективного инструмента для оптимизации проектных решений в различных областях деятельности. Показано, что генеративная политика, как многообещающий подход, предполагает использование алгоритмических решений для генерации и анализа альтернативных вариантов действий, что делает его особенно ценным в контексте оптимизации проектных решений. Рассмотрены ключевые методы и технологии, которые используются для реализации генеративной политики. Описана оценка влияния генеративной политики на оптимизацию проектных решений сложных систем, а также выявление эффективных подходов, способствующих достижению поставленных целей.

Статья делает вывод о том, что генеративная политика представляет собой перспективный подход для повышения эффективности проектных решений, способный активно влиять на процесс инновационного развития в различных отраслях. Рекомендуются направления для дальнейших исследований, направленных на оптимизацию методов генеративного проектирования и их адаптацию к специфическим требованиям различных секторов экономики страны.

**Ключевые слова:** генеративная политика, алгоритмические решения, оптимизации проектных решений, машинное обучение, искусственный интеллект, ключевые индикаторы эффективности

**Введение.** В условиях глобализации и стремительного развития технологий организации сталкиваются с необходимостью быстрого и эффективного принятия решений в рамках проектной деятельности. Генеративная политика, как многообещающий подход,

предполагает использование алгоритмических решений для генерации и анализа альтернативных вариантов действий, что делает его особенно ценным в контексте оптимизации проектных решений [1, 2]. Этот подход не только позволяет существенно расширить пространственный диапазон возможностей, но и обеспечивает более гибкий и адаптивный процесс разработки проектов.

Актуальность данной темы подчеркивается работами исследователей, таких как [3], которые поднимают вопросы о том, как генеративные методы могут быть интегрированы в существующие системы управления проектами. В свою очередь, зарубежные работы, например, исследования [4], показывают значимость применения машинного обучения и искусственного интеллекта в контексте генеративной политики, что открывает новые горизонты для оптимизации и инноваций.

Цель настоящей статьи – оценить эффективность генеративной политики в реализации проектных решений, основываясь на существующих методах анализа и практических примерах. Авторы ставили перед собой задачу не только рассмотреть методологические аспекты, но и проанализировать успешные кейсы применения генеративной политики как на отечественном, так и на международном уровне.

Статья будет полезна как для ученых и практиков, стремящихся понять преимущества генеративной политики, так и для организаций, интересующихся внедрением современных подходов в свою проектную деятельность.

**Ключевые аспекты генеративной политики.** Генеративная политика – это подход в управлении и принятии решений, который сосредотачивается на создании, анализе и оптимизации множества вариантов и сценариев действий. В отличие от традиционных методов проектирования, которые часто основываются на заранее определенных параметрах и фиксированных алгоритмах, генеративная политика использует алгоритмические средства, модели и средства анализа данных для генерации новых, нестандартных решений. Главная цель данного подхода – оптимизация проектных решений путем создания и оценки альтернативных вариантов, что позволяет находить наиболее эффективные пути достижения целей.

Существует несколько ключевых методов и технологий, которые используются для реализации генеративной политики:

– алгоритмы оптимизации – к ним можно отнести генетические алгоритмы (хорошо справляются с задачами, где пространство решений велико и сложно, например, в задачах оптимизации проектных решений), алгоритмы муравьиной колонии (позволяют параллельно исследовать множество решений, что ускоряет процесс поиска) и алгоритмы роя частиц (хорошо подходят для поиска оптимальных решений в многомерных пространствах и позволяют множеству решений одновременно исследовать пространство решений, что делает их эффективными для сложных задач). Эти методы способны исследовать пространство решений и выявлять наилучшие варианты, учитывая множественные критерии и ограничения;

– сценарный анализ – сценарное моделирование предполагает создание различных сценариев на основе неопределенных переменных и факторов. Он способствует принятию более гибких проектных решений, основанных на прогнозировании и анализе возможных будущих состояний проектируемой системы;

– машинное обучение и искусственный интеллект – представляет интеграцию методов машинного обучения в генеративные модели и позволяет анализировать большие объемы данных, находя скрытые закономерности. Эти технологии могут выявлять паттерны и адаптироваться к изменяющимся условиям, что делает процесс генерации решений более эффективным и информированным;

– динамическое программирование – этот метод позволяет решать сложные задачи, разбивая их на меньшие подзадачи и запоминая результаты для последующего использования. Динамическое программирование является полезным инструментом в

контексте генеративной политики, поскольку оно позволяет оптимизировать решения на каждом этапе;

– интерактивные модели и симуляции – это мощные инструменты, которые позволяют визуализировать, анализировать и прогнозировать поведение сложных систем в различных областях. Создание интерактивных симуляций и моделей позволяет участникам проектной деятельности экспериментировать с различными параметрами и условиями эксплуатации. Это может помочь в генерации новых идей и улучшении процесса принятия решений за счет анализа реакций системы на изменения;

– краудсорсинг идей – вовлечение широкого круга участников в процессе генерации идей и предложений может значительно улучшить качество принимаемых проектных решений. Краудсорсинг позволяет использовать разнообразие мнений и знаний, что способствует более широкой генерации альтернативных решений.

Проектное решение – это комплект документов, необходимых для осуществления задач с последующим устранением обнаруженных недостатков и несоответствий. Авторы рассматривают его как конкретный план действий, который разрабатывается для достижения определенных целей в рамках проекта и включает в себя описание задач, ресурсов, сроков, ответственных лиц, методов для реализации проекта. Проектное решение является неотъемлемой частью управления проектами и позволяет структурировать процесс, минимизировать риски и оптимизировать использование ресурсов.

В контексте генеративной политики проектные решения играют важную роль в разработке и внедрении стратегий, направленных на создание условий для инноваций, саморегуляции и активного участия проектировщиков в процессах принятия решений [5].

Можно выделить следующие основные аспекты роли проектных решений в генеративной политике:

– стратегическое планирование – проектные решения помогают формулировать четкие цели и задачи, определять ключевые показатели эффективности и разрабатывать стратегии для достижения желаемого результата;

– усиление акцента на инновациях – в генеративной политике акцент делается на генерацию новых идей и подходов. Проектные решения могут служить платформой для внедрения инновационных инициатив, возможности тестирования новых идей и подходов в рамках ограниченного времени и ресурсов, заранее выявить потенциальные риски и разработать стратегии для их минимизации;

– гибкость и адаптивность – проектные решения могут быть адаптированы в зависимости от меняющихся условий. Это позволяет быстро реагировать на вызовы и использовать новые возможности;

– оценка и коррекция – после реализации проектных решений проводится мониторинг и оценка их эффективности, что позволяет выявить успешные практики и области для улучшения, а также вносить коррективы в процесс генеративной политики.

Таким образом, проектные решения являются ключевыми инструментами, позволяющими реализовывать генеративную политику на практике, обеспечивая структурированный и целенаправленный подход к решению сложных задач.

**Методология оценки эффективности генеративной политики при оптимизации проектных решений.** В условиях быстро развивающихся технологий и растущих требований к эффективности работы информационных систем становится все более актуальной необходимость в систематической оценке генеративной политики. Эта политика направлена на создание условий для инновационного развития и активного участия всех заинтересованных сторон в процессе проектирования и реализации программных решений. Методы оценки позволяют не только анализировать результаты, но и адаптировать проектные решения в соответствии с изменяющимися условиями.

Основной целью оценки является определение влияния генеративной политики на оптимизацию проектных решений технических систем, а также выявление эффективных

подходов, способствующих достижению поставленных целей. Задачи, которые необходимо решить в этом процессе, могут включать:

Определение ключевых индикаторы эффективности (*Key Performance Indicators* – KPI) для различных аспектов разработки информационных систем – это важный шаг в управлении проектами и оценке их успеха. KPI помогают измерять прогресс, эффективность и качество работы команды, а также соответствие целям проекта. Авторы предлагают рассмотреть следующие основные аспекты, которые следует учитывать при определении KPI (рисунок 1).

*Рисунок 1.* Основные аспекты ключевых индикаторов эффективности

*1. Определение целей проекта*

– ясность целей: KPI должны быть связаны с конкретными целями проекта, такими как улучшение производительности, снижение затрат или повышение удовлетворенности пользователей;

– SMART-критерии: цели должны быть конкретными, измеримыми, достижимыми, актуальными и ограниченными во времени.

*2. Качество разработки*

– количество ошибок: измерение числа дефектов на этапе тестирования и после развертывания системы;

– уровень удовлетворенности пользователей: оценка качества системы через опросы и отзывы пользователей.

*3. Производительность команды*

– скорость разработки: измерение времени, необходимого для завершения определенных этапов или задач в рамках проекта (например, скорость выполнения спринтов в Agile);

– процент завершенных задач: количество выполненных задач по сравнению с запланированными.

*4. Управление ресурсами*

– использование бюджета: сравнение фактических затрат с запланированными, чтобы определить, насколько эффективно используются ресурсы;

– загрузка команды: оценка распределения задач между членами команды и их загруженности.

*5. Технические аспекты*

– время отклика системы: измерение времени, необходимого для обработки запросов пользователей;

– надежность системы: количество сбоев или времени простоя системы в течение определенного периода.

*6. Удовлетворенность пользователей*

– NPS (*Net Promoter Score*): оценка готовности пользователей рекомендовать систему другим;

– клиентская поддержка: время, необходимое для решения запросов пользователей и уровень удовлетворенности клиентской поддержкой.

*7. Инновации и улучшения*

– внедрение новых функций: количество новых функций или улучшений, добавленных в систему за определенный период.

– обратная связь от пользователей: оценка предложений и отзывов пользователей для дальнейшего улучшения системы.

*8. Обучение и развитие команды*

– количество часов обучения: измерение времени, потраченного командой на обучение и повышение квалификации.

– уровень вовлеченности: оценка уровня вовлеченности команды в процесс разработки и улучшения системы.

При определении KPI важно учитывать сбалансированность (рисунок 2).

*Рисунок 2*. Показатели сбалансированности KPI

Определение и мониторинг KPI для разработки информационных систем помогает командам оставаться на правильном пути, выявлять проблемы на ранних стадиях и вносить необходимые коррективы в процесс разработки. Это, в свою очередь, способствует

успешной реализации проектов и повышению общей эффективности.

Эффективная генеративная политика в управлении проектами требует всестороннего анализа, включающего как количественные, так и качественные методы оценки. Эти методы обеспечивают глубокое понимание и объективное измерение результатов, что позволяет лучше оптимизировать проектные решения [5, 6].

Количественные методы анализа основываются на статистических данных и числовых показателях, что позволяет создать объективную картину эффективности генеративной политики. Основные из них включают:

– статистический анализ – использование описательной и инференциальной статистики для определения закономерностей и тенденций в данных, связанных с выполнением проектов. Это может включать анализ прибыльности, сроков реализации и удовлетворенности клиентов;

– методы многокритериальной оптимизации – использование алгоритмов, таких как AHP (*Analytic Hierarchy Process*) или TOPSIS (*Technique for Order Preference by Similarity to Ideal Solution*), для оценки и выбора наилучших решений на основе нескольких критериев. Это позволяет учесть различные аспекты проектной деятельности и эффективно расставить приоритеты;

– моделирование и симуляции – применение математических и компьютерных моделей для прогнозирования результатов внедрения генеративной политики, а также сценарного моделирования, позволяющего оценить риски и возможности, связанные с различными решениями;

– анализ временных рядов – изучение изменений показателей эффективности во времени, что позволяет выявить тренды и циклы, а также оценить устойчивость генеративной политики к внешним воздействиям.

Качественные методы анализа фокусируются на глубинном понимании процессов и факторов, влияющих на эффективность генеративной политики. К числу таких методов можно отнести:

– кейс-методы – изучение конкретных случаев успешного (или неуспешного) применения генеративной политики помогает выявить лучшие практики и возможные проблемы. Такой подход позволяет анализировать контекст и факторы, влияющие на результаты;

– интервью и опросы – сбор мнений экспертов и участников проектов через полуструктурированные интервью или анкетирование помогает получить ценные инсайты о восприятии и восприятии генеративной политики различными стейкхолдерами;

– фокус-группы – проведение обсуждений в малых группах позволяет глубже понять взгляды и ожидания участников проекта относительно генеративной политики, а также выявить возможные барьеры и предложения по улучшению;

– системный анализ – рассматривает проект как часть более широкой системы, что позволяет оценить влияние генерируемых решений на смежные процессы и сферы деятельности. Этот метод также подчеркивает важность комплексного подхода к оптимизации проектных решений.

Применение как количественных, так и качественных методов позволяет обеспечить комплексный подход к анализу эффективности генеративной политики. Совмещение этих методов дает возможность не только получить точные данные, но и понять контекст, в котором они были собраны. Это, в свою очередь, способствует более обоснованному принятию решений и оптимизации проектных решений, что является ключевым фактором для достижения успеха в условиях динамично развивающегося рынка.

**Сравнение генеративной политики с другими подходами к оптимизации проектных решений.** Генеративная политика представляет собой инновационный подход к оптимизации проектных решений, основанный на создании и исследовании множества возможных альтернатив. Сравним генеративную политику с другими традиционными и

современными подходами к управлению проектами, такими как классический проектный менеджмент, адаптивное управление проектами и использование искусственного интеллекта (ИИ).

*Классический проектный менеджмент*

Классический проектный менеджмент, согласно методологии «водопад», фокусируется на четком определении всех этапов проекта заранее, включая планирование, исполнение и завершение [8]. Этот подход имеет свои достоинства, такие как структурированность и предсказуемость, однако он часто не учитывает изменчивость условий и потребностей рынка.

Сравнение:

– гибкость – генеративная политика позволяет адаптироваться к изменениям в реальном времени, в то время как классический подход требует жесткого следования первоначальному плану [9];

– итеративность – генеративная политика предусматривает итеративную проработку альтернатив, что способствует поиску оптимальных решений. В отличие от этого, классический проектный менеджмент склонен к линейному процессу [10].

*Адаптивное управление проектами*

Адаптивное управление проектами, основанное на принципах Agile, подразумевает гибкость и итеративный подход к разработке. В отличие от генеративной политики, Agile фокусируется на регулярном обновлении требований и тесном сотрудничестве с заказчиками [11].

Сравнение:

– фокус на альтернативе – генеративная политика исследует множество альтернатив и их взаимодействие, в то время как Agile чаще всего использует заранее определенные итерации и обновления [12];

– количество решений – генеративная политика может оценивать и предлагать большее количество возможных альтернатив, что является ее основным преимуществом, тогда как Agile ограничивается обычно несколькими итерациями и версиями продукта.

*Искусственный интеллект*

В последние годы искусственный интеллект и машинное обучение начинают активно использоваться в управлении проектами для улучшения процессов принятия решений, прогнозирования рисков и автоматизации рутинных задач [13].

Сравнение:

– анализ данных – генеративная политика фокусируется на генерации и оценке альтернатив, тогда как ИИ интегрирует большие объемы данных для автоматического анализа и предоставления рекомендаций [14].

– человеческий фактор – генеративная политика требует активного участия разработчиков в процессе принятия решений, что может приводить к более креативным и адаптивным решениям, тогда как ИИ может снизить влияние человеческого участия, что в некоторых случаях может привести к утрате контекстной информации.

Сравнение генеративной политики с другими подходами к оптимизации проектных решений демонстрирует ряд ее уникальных преимуществ, включая гибкость, итеративность и способность к поиску множества альтернатив. Традиционные методы управления проектами, такие как классический проектный менеджмент и Agile, а также новые технологии, такие как ИИ, имеют свои сильные и слабые стороны. Однако именно генеративная политика предлагает интегративный подход, позволяющий использовать лучшие практики и технологии для повышения эффективности проектного управления.

Таким образом, данное сравнение подчеркивает уникальные аспекты генеративной политики и ее значение в контексте современных подходов к управлению проектами.

**Направления для дальнейших исследований в области оптимизации методов генеративного проектирования.** Авторы предлагают следующие направления, позволяющие выполнить оптимизацию методов генеративного проектирования [16–20]:

– исследования, направленные на усовершенствование алгоритмов машинного обучения для повышения точности генеративного дизайна, а также на разработку адаптивных моделей, способных самообучаться на основе успешных решений из прошлых проектов;

– изучение методов генеративного проектирования, позволяющих учитывать специфические требования природоохранных, строительных и производственных стандартов, что поможет обеспечить соответствие конечных решений законодательным нормам и экологическим требованиям;

– расширение исследований в области междисциплинарных методов, объединяющих архитекторов разработки программного обеспечения, инженеров, дизайнеров интерфейсов и специалистов по данным, для совместной работы над генеративными проектами, что может привести к более интегрированным и инновационным решениям;

– оптимизация генеративных алгоритмов для более глубокого анализа и учета пользовательских предпочтений и поведения, что позволит создавать более персонализированные и востребованные решения;

– исследование экономических моделей, позволяющих оценить стоимость внедрения технологий генеративного проектирования для разных отраслей и их влияние на прибыльность бизнеса, включая расчет затрат на внедрение и потенциальную выгоду;

– разработка программного обеспечения для генеративного проектирования – создание и оптимизация интуитивно понятных инструментов и платформ, которые позволят пользователям без глубоких технических знаний эффективно применять генеративные методы в своей практике;

– разработка общепринятых стандартов и методик для применения генеративного проектирования в разных отраслях, что повысит его внедрение и использование;

– объединение усилий академических и исследовательских институтов, а также бизнеса для создания совместных программ по изучению генеративного проектирования, что поможет обмену знаниями и практическим опытом между различными отраслями.

Эти направления исследований помогут улучшить методы генеративного проектирования и сделать их более адаптивными к специфическим требованиям и вызовам, стоящим перед различными секторами экономики страны, способствуя таким образом их инновационному развитию и конкурентоспособности.

**Заключение.** Исследования эффективности генеративной политики как одного из перспективных механизмов оптимизации проектных решений показывает его значительный потенциал в улучшении процессов разработки и проектирования программного обеспечения. С помощью алгоритмических методов и искусственного интеллекта удается не только ускорить создание альтернативных решений, но и повысить их качество и экономическую целесообразность.

Анализ результатов внедрения генеративной политики в реальных проектах подтверждает, что такая политика способствует снижению затрат и времени на разработку, а также увеличению инновационности и адаптивности конечных продуктов к требованиям современного рынка. Тем не менее, важно учитывать существующие вызовы и ограничения, такие как необходимость в междисциплинарных знаниях и сложности в разработке и адаптации алгоритмов.

Перспективы дальнейших исследований в этой области открывают новые возможности для оптимизации методов генеративного проектирования. Это включает в себя необходимость разработки адаптивных моделей, учет различных факторов внешней среды, а также создание удобных инструментариев для практического применения. В конечном счете, интеграция генеративной политики в процессы проектирования сможет не

только повысить эффективность решений, но и играющую ключевую роль в устойчивом развитии и инновационном прогрессе в различных отраслях экономики.

Таким образом, генеративная политика представляет собой важный шаг к более рациональному и инновационному подходу в процессе проектирования, способствуя созданию более эффективных и адаптивных решений, отвечающих современным требованиям.

**Список литературы**

[1] Борисов, И. А. Генеративные методы в проектном управлении / А. И. Борисов // Управление проектами. – 2020. – № 12(3). – С. 45-59.

[2] Smith, J. Machine Learning Applications for Generative Decision-Making / J. Smith, T. Johnson // Journal of Business Intelligence. – 2021. – № 15(2). – PP. 88-99.

[3] Ковалев, В. Н. Инновационные подходы к оптимизации проектных решений / В. Н. Ковалев, А. Е. Дьяченко // Вестник науки и образования. – 2019. – № 8. С. 119-128.

[4] Miller, R. Generative Policy Optimization in Project Management / R. Miller, A. Sokolov // International Journal of Project Management. – 2022. – № 40(1). С. 23-36.

[5] Демарко Т. Человеческий фактор: успешные проекты и команды, 2-е издание. – пер. с англ. / Том Демарко, Тимоти Листер – Спб: Символ-плюс, 2005. – 256 с.

[6] Law, Averill M. Simulation modeling and analysis / Averill M. Law, President Averill M. Law & Associates, Inc. – Tucson, Arizona, USA, www.averill-law.com.

[7] Field, A. P. Discovering Statistics Using IBM SPSS Statistics. SAGE Publications / A. P. Field – 2013. – № 5. – PP. 179-185.

[8] A Guide to the Project Management Body of Knowledge (PMBOK Guide) – 6th Edition. – Project Management Institute, 2017. – 756 p.

[9] Wysocki R. K. Effective Project Management: Traditional, Agile, Extreme / R. K. Wysocki – Wiley – 2014. – 770 p.

[10] Meredith, J.R. Project Management: A Managerial Approach, 10th ed. / J. R Meredith, S. J. Mantel S.J. – Wiley, 2018. – 520 p.

[11] Beck, K., et al. (2001). Manifesto for Agile Software Development. Available at: URL: https://agilemanifesto.org/.

[12] Highsmith, J. Agile Project Management: Creating Innovative Products. Addison-Wesley. https://www.researchgate.net/publication/234809670_Agile_Project_Management_Creating_Innovative_Products (Дата обращения: 10.03.2025).

[13] Serrador, P. Does Agile work? A quantitative analysis of agile project success / P. Serrador, J. K. Pinto, // International Journal of Project Management. – 2015. – №33(5), PP. 1040-1051.

[14] Mello, R. A., et al. Artificial Intelligence for Project Management / R. A. Mello // PM World Journal. 2019. –№ 8(8) URL: https://www.researchgate.net/publication/379960592 (Дата обращения: 10.03.2025).

[15] Моделирование и оптимальное проектирование технических систем : учебно-методическое пособие / В. Ф. Алексеев, Д. В. Лихачевский, Г. А. Пискун, В. В. Шаталова. − Минск : БГУИР, 2024. – 99 с.

[16] Программные инновационные платформы информационных систем. Лабораторный практикум : пособие / В. Ф. Алексеев, Д. В. Лихачевский, Г. А. Пискун, В. В. Шаталова. − Минск : БГУИР, 2024. – 158 с.

[17] Алексеев, В. Ф. Подходы к разработке программного средства для оценки кадровых рисков компании / В. Ф. Алексеев, Д. В. Лихачевский // Цифровая среда: технологии и перспективы : сборник материалов II Международной научно-практической конференции, Брест, 31 октября–1 ноября 2024 г. / Брестский государственный технический университет ; редкол.: Н. Н. Шалобыта (гл. ред.) [и др.]. – Брест, 2024. – С. 212–221.

[18] Алексеев, В. Ф. Обеспечение работоспособности программного средства по автоматизации документооборота дипломного проектирования / В. Ф. Алексеев, А. Ю. Писарчик, А. Д. Сыс // Фундаментальные и прикладные научные исследования: актуальные вопросы, достижения и инновации : сборник статей LII Международной научно-практической конференции, Пенза, 15 января 2022 г.: в 2 ч. Ч. 1. / Международный центр научного сотрудничества «Наука и просвещение» ; редкол.: Г. Ю. Гуляев [и др.]. – Пенза, 2022. – С. 100–102.

[19] Оценка качества передачи информации в системе диспетчеризации на базе MQTT-архитектуры / В. Ф. Алексеев [и др.] // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сборник научный статей VIII Международной научно-практической конференции, Минск, 11-12 мая 2022 года / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2022. – С. 483–488.

[20] Малиновская, В. В. Архитектура больших данных / В. В. Малиновская, В. Ф. Алексеев // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: VII Международная научно-практическая конференция [Электронный ресурс] : сборник материалов VII Международной научно-

практической конференции, Минск, 19-20 мая 2021 года / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2021. – С. 222–226.

### Авторский вклад

**Алексеев Виктор Федорович** – постановка задачи исследования, разработка методологии оценки эффективности генеративной политики при оптимизации проектных решений, руководство исследованием по оценке эффективности генеративной политики при оптимизации возможных проектных решений.

**Лихаческий Дмитрий Викторович** – постановка задачи исследования, сравнение генеративной политики с другими подходами к оптимизации проектных решений, разработка направлений для дальнейших исследований в области оптимизации методов генеративного проектирования.

**Пискун Геннадий Адамович** – основные аспекты, которые следует учитывать при определении KPI, определение и мониторинг KPI для разработки информационных систем, разработка перспектив дальнейших исследований в области генеративной политики.

## EVALUATION OF THE EFFECTIVENESS OF GENERATIVE POLICY IN OPTIMIZING DESIGN DECISIONS

***V.F. Alekseev***
*Associate Professor, Department of Information Computer Systems Design, PhD of Technical sciences, Associate Professor*

***D.V. Likhachevsky***
*Dean of the Faculty of Computer Design of BSUIR, PhD of Technical Sciences, Associate Professor*

***G.A. Piskun***
*Associate Professor of the Department of Design of Information and Computer Systems of BSUIR, PhD of Technical Sciences, Associate Professor*

**Abstract.** The article discusses the use of generative policy as an effective tool for optimizing design decisions in various fields of activity. It is shown that generative policy, as a promising approach, involves the use of algorithmic solutions for generating and analyzing alternative courses of action, which makes it especially valuable in the context of optimizing design decisions. Key methods and technologies used to implement generative policy are considered. An assessment of the impact of generative policy on the optimization of design decisions for complex systems is described, as well as the identification of effective approaches that contribute to achieving the goals. The article concludes that generative policy is a promising approach to improving the efficiency of design decisions, capable of actively influencing the process of innovative development in various industries. Directions for further research aimed at optimizing generative design methods and adapting them to the specific requirements of various sectors of the country's economy are recommended.

**Keywords**: generative policy, algorithmic solutions, design decision optimization, machine learning, artificial intelligence, key performance indicators.

УДК 004.8 (075)

# РОЛЬ KDD В РАЗВИТИИ ИНСТРУМЕНТОВ АНАЛИЗА И УПРАВЛЕНИЯ ДАННЫМИ

***Ф.М. Алимова***
*Доцент кафедры Конвергенция цифровых технологий ТУИТ имении Мухаммада ал Хоразмий, доктор философии технических наук, PhD*
tuit.alimova@gmail.com

***Ф.М. Алимова***
*Окончила Ташкентский университет информационных технологий. Область научных интересов связана с исследованием проблем оптимизации бизнес процессов, построения информационных систем, организацией учебного и научно-исследовательского процессов в техническом университете.*

**Аннотация.** В статье рассматривается роль процесса извлечения знаний из данных (KDD) в анализе и управлении образовательными данными. Выявлены особенности методов интеллектуального анализа данных (Data Mining), такие как кластеризация, классификация, прогнозирование и анализ ассоциаций. Освещаются анализ успеваемости учащихся, моделирование образовательных данных и интеграцию аналитических методов в учебные процессы. Рассматриваются преимущества и ограничения различных методов KDD, а также их практическое применение для оптимизации образовательных систем. В статье подчеркивается важность автоматизированных методологий анализа данных для выявления полезных знаний и принятия обоснованных решений в сфере образования.

**Ключевые слова:** Data Mining, big data, анализа данных, знания.

**Введение.** В сфере образования большие объемы данных собираются в системах управления базами данных и хранилищах данных из всех вовлеченных областей. Анализ данных стал важным инструментом для сбора знаний из учебных баз данных. Основные функции интеллектуального анализа данных, которые должны быть выполнены, включают в себя характеристику и описание, ассоциацию, классификацию, прогнозирование, кластеризацию и анализ эволюции. Было показано, что в последние годы наблюдается быстрый рост применения интеллектуального анализа данных в контексте образовательных процессов и учреждений. Этот обзор раскрывает прогрессивные применения и существующие пробелы, выявленные в контексте интеллектуального анализа данных в образовательном сфере.

В последнее время большое внимание уделяется важному аспекту инновации-большие данные. Учитывая важность сферы образования, в настоящее время наблюдается тенденция к изучению роли больших данных в этом секторе. На данный момент проводиться множество исследований, посвящённых применению больших данных в различных областях для разных целей. Однако в сфере образования до сих пор нет всестороннего обзора больших данных.

На сегодняшний день проводимые исследования охватывают четыре основные темы, связанные с большими данными в образовании: поведение и успеваемость учащихся,

моделирование и хранилища образовательных данных, совершенствование образовательной системы и интеграция больших данных в учебную программу. Большинство исследований в области образования, связанных с большими данными, были сосредоточены на поведении и успеваемости учащихся.

Большие объемы данных в образовательных базах данных, которые содержат большое количество записей со множеством атрибутов, которые необходимо одновременно исследовать для обнаружения полезной информации и знаний, делают ручной анализ непрактичным. Все эти факторы указывают на необходимость интеллектуальных и автоматизированных методологий анализа данных, которые могли бы обнаруживать полезные знания из данных. Поэтому обнаружение знаний в базах данных (KDD) и интеллектуальный анализ данных (DM) стали чрезвычайно важными инструментами в реализации цели интеллектуального и автоматизированного анализа данных. Интеллектуальный анализ данных является особым шагом в процессе KDD, включающим применение определенных алгоритмов для извлечения шаблонов (моделей) из данных. Дополнительные шаги в процессе KDD, такие как подготовка данных, очистка данных, выбор данных, включение соответствующих предварительных знаний и правильная интерпретация результатов интеллектуального анализа, гарантируют, что полезные знания будут извлечены из данных.

KDD включает в себя теории, алгоритмы и методы из пересечения нескольких областей исследований, включая технологию баз данных, машинное обучение, статистику, искусственный интеллект, системы, основанные на знаниях, и визуализацию данных.

KDD отличается от других областей, таких как машинное обучение или искусственный интеллект или смежных областей тем, что эти области предоставляют конкретные инструменты добычи данных, которые могут использоваться на различных этапах процесса KDD. В последнее время, с ростом технологии добычи данных, исследователи и практики в различных аспектах производства и логистики начали применять эту технологию для поиска скрытых связей или закономерностей, которые могли бы использоваться для оснащения их систем новыми знаниями. Ранние приложения добычи данных в основном применялись к финансовым приложениям, например, Чжан и Чжоу (2004) описали добычу данных в контексте финансовых приложений как с технической, так и с прикладной точки зрения. Главным преимуществом интеллектуального анализа данных перед другими экспериментальными методами является то, что необходимые данные для анализа могут быть собраны во время нормальной работы изучаемого производственного процесса. Поэтому, как правило, нет необходимости специально выделять машины или процессы для сбора данных.

Большинство исследований в области образования, связанных с большими данными, были сосредоточены на поведении и успеваемости учащихся. Это исследование служит руководством для будущих исследований и освещает новые идеи и направления для успешного использования больших данных в образовании.

Методы интеллектуального анализа даёт возможность выполнять систематизацию данных по критериям количества и качества. При обработке массивов информации наиболее часто используется следующие способы:

- кластерный анализ (иерархический и неиерархический);
- байесовские сети;
- поиск явных и неявных ассоциаций;
- линейная регрессия;
- представление сведений в визуальной форме;
- эволюционное программирование;
- генетические алгоритмы.

Каждый из перечисленных методов применяется для выборки конкретных данных, построения прогностической модели или очистки БД от ошибочных сведений.

Хан и Камбер (2001) отметили, что вид знаний, которые необходимо добыть, определяет функции добычи данных, которые необходимо выполнить. Возможные виды знаний включают описание концепций (характеристику и дискриминацию), ассоциацию, классификацию, кластеризацию и прогнозирование.

KDD– это процесс получения из данных знаний, в виде зависимости правил моделей позволяющих моделировать и прогнозировать различные процессы.

Технологии KDD описывают не конкретный алгоритм обработки данных или математический аппарат, а последовательность действий которую необходимо выполнить с данными для построения модели и получения знаний.

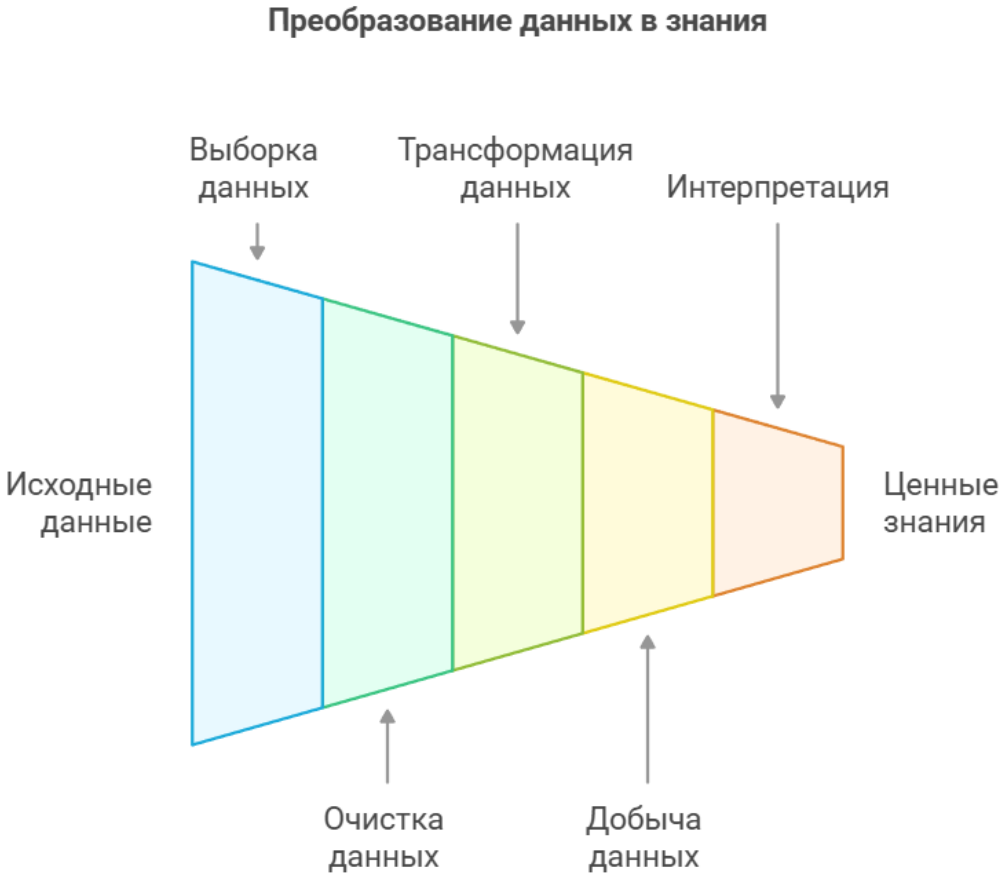



*Рисунок 1.* Процесс KDD в сфере образования

Процесс KDD является интерактивным и итеративным процессом которая предполагает динамическое взаимодействие и повторение действий для достижения наилучшего результата включающим более или менее следующие этапы.

1 Понимание области: Это включает в себя соответствующие предыдущие знания, связанные с применением образования и целевой целью.

2 Сбор целевых данных: Это включает в себя сбор необработанных данных, выбор набора данных и фокусирование на наборе переменных, влияющих на процесс образования.

3 Очистка, предварительная обработка и преобразование данных: Это включает предварительную обработку данных, такую как удаление шума, замена пропущенных значений и очистка данных. Данные консолидируются в формы, подходящие для добычи данных.

4 Интеграция данных: Это включает интеграцию нескольких образовательных и производственных гетерогенных источников данных.

5 Выбор функций добычи данных: На основе типа требуемых знаний необходимо выполнить различные функции добычи данных (кластеризация, классификация, прогнозирование, ассоциация, регрессия, суммирование и т. д.) для получения модели.

6 Выбор подходящего алгоритма интеллектуального анализа данных: это включает выбор методов для выполнения желаемой функции поиска закономерностей в данных.

7 Интеллектуальный анализ данных: это включает поиск закономерностей, представляющих интерес, в определенной репрезентативной форме или наборе таких представлений.

8 Интерпретация и визуализация: это включает интерпретацию и визуализацию закономерностей для получения новых знаний.

9 Внедрение обнаруженных знаний: обнаруженные знания включаются в систему производительности области образования. Получается обратная связь, и знания могут быть дополнительно изменены на основе обратной связи.

10 Хранение, повторное использование и интеграция знаний в образовательную систему: Это включает хранение обнаруженных знаний для будущего повторного использования и возможной интеграции в образовательной системе

Алгоритмы поиска бывают двух типов: поиск параметров по заданной модели и поиск моделей в пространстве моделей. Поиск наилучших параметров часто сводится к задаче оптимизации. Алгоритмы добычи данных, как правило, опираются на относительно простые методы оптимизации, хотя в принципе используются и более сложные методы оптимизации. Проблемы с локальными минимумами являются обычными и решаются обычным способом (например, множественные случайные перезапуски и поиск нескольких моделей). Поиск в пространстве моделей обычно выполняется жадным способом. Важным моментом является то, что каждый метод обычно подходит для некоторых задач лучше, чем для других. Например, классификаторы деревьев решений могут быть очень полезны для поиска структуры в многомерных пространствах, а также полезны в задачах со смешанными непрерывными и категориальными данными (поскольку методы деревьев не требуют метрик расстояний). Однако деревья классификации с одномерными пороговыми границами решений могут не подходить для задач, где истинные границы решений являются нелинейными многомерными функциями. Таким образом, не существует универсально наилучшего метода интеллектуального анализа данных. Выбор конкретного алгоритма для конкретного приложения — это своего рода искусство. На практике большая часть усилий приложения может быть направлена на правильную формулировку проблемы (постановку правильного вопроса), а не на оптимизацию алгоритмических деталей конкретного метода добычи данных.

Высокоуровневые цели добычи данных, как правило, являются предиктивными, описательными или комбинацией предиктивных и описательных. Чисто предиктивная цель фокусируется на точности предиктивной способности. Чисто описательная цель фокусируется на понимании основного процесса генерации данных — тонкое, но важное различие. При прогнозировании пользователю может быть все равно, отражает ли модель реальность, если она имеет предиктивную силу. Описательная модель, с другой стороны, интерпретируется как отражение реальности. На практике, большинство приложений KDD требуют некоторой степени как предиктивного, так и описательного моделирования.

Основные исследовательские и прикладные проблемы для KDD включают:

1 Массивные наборы данных и высокую размерность. Многогигабайтные базы данных с миллионами записей и большим количеством полей (атрибутов и переменных) являются обычным явлением. Эти наборы данных создают комбинаторно взрывные пространства поиска для индукции модели и увеличивают вероятность того, что алгоритм добычи данных найдет ложные шаблоны, которые в целом не являются допустимыми. Возможные решения включают очень эффективные алгоритмы, выборку, методы аппроксимации, массивно-параллельную обработку, методы снижения размерности и включение предыдущих знаний.

2 Взаимодействие с пользователем и предыдущие знания. Аналитик обычно не является экспертом KDD, а лицом, ответственным за осмысление данных с использованием

доступных методов KDD. Поскольку процесс KDD по определению интерактивен и итеративен, сложно обеспечить высокопроизводительную, быстродействующую среду, которая также помогает пользователям в правильном выборе и сопоставлении соответствующих инструментов и методов для достижения их целей. Необходимо больше внимания уделять взаимодействию человека и компьютера и меньше — полной автоматизации — с целью поддержки как опытных, так и начинающих пользователей. Многие текущие методы и инструменты KDD не являются по-настоящему интерактивными и не позволяют легко включать предыдущие знания о проблеме, за исключением простых способов. Использование знаний о предметной области важно на всех этапах процесса KDD. Например, байесовские подходы используют априорные вероятности по данным и распределениям как один из способов кодирования предыдущих знаний. Другие используют возможности дедуктивной базы данных для обнаружения знаний, которые затем используются для руководства поиском по добыче данных.

3 Интеграция. Отдельная система обнаружения может быть не очень полезной. Типичные проблемы интеграции включают интеграцию с СУБД (например, через интерфейс запросов), интеграцию с электронными таблицами и инструментами визуализации, и размещение показаний датчиков в реальном времени.

Высоко интерактивные среды человек-компьютер, как описано в процессе KDD, допускают как компьютерное обнаружение с помощью человека, так и компьютерное обнаружение с помощью человека. Разработка инструментов для визуализации, интерпретации и анализа обнаруженных шаблонов имеет первостепенное значение. Такие интерактивные среды могут обеспечить практические решения многих реальных проблем гораздо быстрее, чем люди или компьютеры, работающие независимо. Существует потенциальная возможность и проблема разработки методов интеграции инструментов OLAP сообщества баз данных и инструментов интеллектуального анализа данных машинного обучения и статистических сообществ.

4 Нестандартные, мультимедийные и объектно-ориентированные данные. Значительной тенденцией является то, что базы данных содержат не только числовые данные, но и большие объемы нестандартных и мультимедийных данных. Нестандартные типы данных включают нечисловые, нетекстовые, геометрические и графические данные, а также нестационарные, временные, пространственные и реляционные данные, а также смесь категориальных и числовых полей в данных. Мультимедийные данные включают многоязычный текст в свободной форме, а также оцифрованные изображения, видео, речевые и аудиоданные.

Эти типы данных в значительной степени выходят за рамки текущей технологии KDD.

**Заключение.** Несмотря на быстрый рост, область KDD все еще находится в зачаточном состоянии. Необходимо преодолеть множество проблем, но некоторые успехи были достигнуты. Поскольку потенциальная отдача от приложений KDD высока, на рынке наблюдается спешка с предложением продуктов и услуг. Большой проблемой, стоящей перед областью, является то, как избежать ложных ожиданий, преследующих другие зарождающиеся (и связанные) технологии. Исследователи и практики в этой области обязаны обеспечить, чтобы потенциальный вклад KDD не был преувеличен и чтобы пользователи понимали истинную природу вклада вместе с его ограничениями. Фундаментальные проблемы, лежащие в основе этой области, остаются нерешенными. Например, основные проблемы статистического вывода и открытия остаются такими же сложными и затруднительными, какими они были всегда. Овладение искусством анализа и способностью человеческого мозга синтезировать новые знания из данных по-прежнему непревзойденно ни одной машиной. Однако объемы данных, которые необходимо проанализировать, делают машины необходимостью. Эта ниша для использования машин в качестве вспомогательного средства для анализа и надежда на то, что огромные наборы

данных содержат крупицы ценных знаний, стимулируют интерес и исследования в этой области. Объединяя набор различных областей, KDD создает плодородную почву для роста новых инструментов для управления, анализа и, в конечном итоге, получения превосходства над потоком данных, с которым сталкивается современное общество. Тот факт, что область движима сильными социальными и экономическими потребностями, является стимулом для ее дальнейшего роста. Проверка реальности реальных приложений будет действовать как фильтр, отсеивающий хорошие теории и методы от менее полезных.

## Список литературы

[1] A.Abdullaye, F.M. Alimova, U.Giyosov. Control of knowledge in the test. “XXI аср ва технология соҳасидаги устувор йўналишлар”, VII Халқаро илмий конференцияси бўйича мақолалар тўплами, 2021. – 477 с.

[2] F.M. Alimova. Data mining in customer relationship managemen. TATU xabarlari, 2014. – 11 с.

[3] Usmanov Jonibek Turdikulovich, Pulatova Ziyoda Mahmudjonovna, Naim Nodira Abdujalolovna, Makhbuba Kushmanova Abdunabiyevna, Alimova Fotima Muratovna. Developing a model for comprehensive and similarity analysis of systems. International Journal of Mechatronics and Applied Mechanics, 2022, Issue 11c.

# THE ROLE OF KDD IN THE DEVELOPMENT OF DATA ANALYSIS AND MANAGEMENT TOOLS

***F.M. Alimova***
*Associate Professor, Department of Convergence of digital technologies, PhD of Technical sciences, Associate Professor*

**Abstract.** The article discusses the role of knowledge-driven data mining (KDD) in analyzing and managing educational data. The article highlights the features of data mining methods, such as clustering, classification, forecasting, and association analysis. It covers the analysis of student performance, educational data modeling, and the integration of analytical methods into educational processes. The advantages and limitations of various KDD methods are discussed, as well as their practical application for optimizing educational systems. The article emphasizes the importance of automated data mining methodologies for discovering useful knowledge and making informed decisions in the field of education.

**Keywords**: Data Mining, big data, data analysis, knowledge.

УДК 004.623

# ОПТИМИЗАЦИЯ ВРЕМЕНИ СБОРА ДАННЫХ О ТРАНЗАКЦИЯХ В СЕТИ SOLANA

***А.Г. Бокун***
*Ассистент кафедры информатики, БГУИР.*
a.bokun@bsuir.by

***А.В. Герчик***
*Ассистент кафедры информатики, БГУИР.*
a.gerchik@bsuir.by

***А.Г. Бокун***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с информатикой, теоретической и прикладной криптографией, а также технологиями шифрования и защиты информации.*

***А.В. Герчик***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием роли и перспективы интеграций технологии дополненной реальности в различных отраслях современного мира.*

**Аннотация.** В статье исследована работа одной из составляющих систем валидатора сети Solana – модуля сбора и трансляции данных о прошедших транзакциях geyser. Во введении обозначена важность и актуальность вопроса сбора данных в высокопроизводительных сетях. В основной части статьи приведен эксперимент, позволяющий больше узнать о тонкостях работы стандартного решения. Практическая ценность полученных результатов оценивается в заключении.

**Ключевые слова:** Solana, блокчейн, криптография, geyser, валидация, транзакции.

**Введение.** Последние разработки в области прикладной криптографии, блокчейн-сетей и децентрализованных финансов получили новый импульс для активного развития ввиду роста рынка криптовалют и институционального признания токенов с наибольшей капитализацией (BTC и ETH) через создание ETF, выпускающего публично торгуемые ценные бумаги, которые отражают движение цен на фьючерсные контракты целевой криптовалюты. Это показывает заинтересованность традиционного финансового капитала в работе со средствами децентрализованной торговли и накопления ликвидности и делает более актуальными и экономически обоснованными исследования в этой области.

Актуальной задачей является разработка инструментов сбора и анализа большого объема данных о прохождении транзакций в блокчейн-сетях. Агрегация подобной информации решает сразу несколько важных задач: анализ движения активов на рынке, оценка и мониторинг работоспособности сети и облегчение процесса разработки и внедрения смарт-контрактов.

**Блокчейн Solana.** В работе рассматривается блокчейн-сеть Solana. Она обладает крайне высокой скоростью работы: актуальный TPS (transaction per second) основной сети (mainnet) составляет 4000, а время создания блока – 0.4 секунды. Данные показатели делают сеть крайне чувствительной к любым внедрениям в процесс валидации транзакций, так как это может значительно повлиять на ее TPS [1].

Для сбора информации о проходящих в сети транзакциях валидаторы Solana имеют поддержку плагина geyser. Данный плагин предоставляет разработчикам интерфейс

взаимодействия с валидатором в части получения данных об интересующих сущностях сети. На данный момент geyser может отправлять данные об аккаунтах, транзакциях, слотах, блоках и записях. Для дальнейшей работы необходимо четко определить следующие термины, опуская некоторые подробности реализации сети.

– блок – непрерывный набор записей в списке валидатора, покрываемый голосованием. Лидирующий валидатор производит не более одного блока за слот.

– запись – часть блока, содержащая в себе информацию об обработанной транзакции. В блоке должно быть как минимум 64 записи.

– слот – период времени, в течение которого каждый лидирующий валидатор принимает транзакции и создает блок (на данный момент равен 0.4 секунды).

– транзакция – одна или несколько инструкций, подписанных клиентом с помощью одной или нескольких пар ключей и выполняемых атомарно лишь с двумя возможными исходами: успех или неудача [2].

**Постановка вопроса.** Уже было отмечено, что для ретрансляции информации о сущностях сети валидаторы используют интерфейс geyser к которому могут подключаться внешние программы с помощью extern “C” метода с особым именем _create_plugin(), как это делает, например, solana-accountsdb-plugin-postgres – реализация доступа к geyser-интерфейсу с записью информации в базу данных PostgreSQL. Однако не для всех разработчиков может быть приемлема работа только лишь со стандартным интерфейсом, в таком случае ставится вопрос: в какой момент наиболее эффективно и безопасно проводить сбор данных о прошедшей через валидатор транзакции? Было принято четыре возможной гипотезы, отвечающей на этот вопрос:

- после выпуска блока (окончания слота), которому принадлежит транзакция;
- после добавления транзакции в запись и закрытия записи;
- после добавления транзакции в запись;
- перед добавлением транзакции в запись.

Также было принято за аксиому, что наиболее оптимальным и безопасным для работоспособности сети будет тот вариант, который был реализован разработчиками в стандартном интерфейсе.

**Эксперимент.** Для проведения исследования локально был развернут тестовый валидатор и установлен плагин solana-accountsdb-plugin-postgres для сохранения данных в PostgreSQL. К сожалению данная реализация плагина (как и любая другая из существующих на сегодняшний день) не собирает данные о записях, поэтому с его помощью получится проверить только первую гипотезу. Плагин создает в базе данных отношения с большим количеством полей, однако в рамках эксперимента большинство из них не имеет значения. Ключевыми для эксперимента являются поля:

– для транзакции: slot (номер слота локальной сети, в который попала транзакция), updatet_on (время изменения записи о транзакции) и index (номер транзакции в блоке);

– для блока: slot (номер слота локальной сети, в который попал блок), updatet_on (время изменения записи о блоке).

Кроме этого в исследовании будет использоваться собственное программное обеспечение, предназначенное для генерации потока транзакций на заданном валидаторе. Также стоит отметить, что название поля updatet_on указывает на то, что временная отметка будет меняться при изменении записи плагином. Чтобы избежать искажения данных было принято решение создать дополнительные отношения и триггеры, которые при добавлении новых кортежей в отношения transaction и block будет копировать ключевые для эксперимента поля в новые отношения. Это будет гарантировать, что временные метки не были изменены в процессе работы плагина и валидатора.

Для первого этапа эксперимента будет сгенерирован поток из десяти транзакций, которые должны будут попасть на один слот (то есть в один блок).

Все транзакции попали на слот 14404 локального валидатора. Блок, принадлежащий этому слоту был создан в HH:MM:00.892. В таблице 1 будет представлена информация о транзакциях (10 сгенерированных транзакций + 1 голосующая, создаваемая валидатором автоматически), принадлежащих слоту 14404.

Таблица 1. Информация о транзакциях слота 14404.

| Индекс | Время создания | Разница с временем создания блока (с) |
|---|---|---|
| 0 | HH:MM:00.493 | 0.399 |
| 1 | HH:MM:00.550 | 0.342 |
| 2 | HH:MM:00.563 | 0.329 |
| 3 | HH:MM:00.563 | 0.329 |
| 4 | HH:MM:00.569 | 0.323 |
| 5 | HH:MM:00.587 | 0.305 |
| 6 | HH:MM:00.593 | 0.299 |
| 7 | HH:MM:00.593 | 0.299 |
| 8 | HH:MM:00.600 | 0.292 |
| 9 | HH:MM:00.600 | 0.292 |
| 10 | HH:MM:00.610 | 0.282 |

Данные свидетельствуют, что все транзакции попали в блокчейн и получили индексы внутри блока корректно. Также видно, что между добавлением информации о последней транзакции и добавлением информации о блоке прошло 0.282 секунды.

Предварительные результаты эксперимента говорят о несостоятельности первой гипотезы. Второй этап предполагает проведение исследования на большем количестве слотов, при большем потоке сгенерированных транзакций. В данном случае была сгенерирована нагрузка из 1000 транзакций, они попали на слоты локальной сети 14602-14608. Ввиду этого для визуализации результат будет принята сквозная индексация транзакций между блоками исследуемых слотов. На рисунке 1 представлен результат второго этапа исследования в виде графика зависимости времени добавления сущности сети в базу данных от ее сквозного индекса.

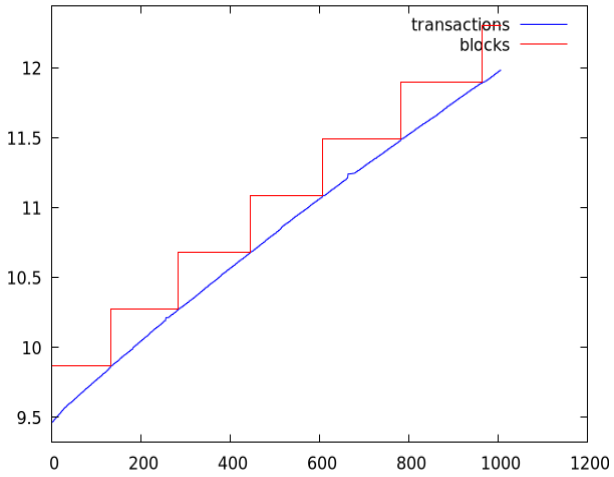



*Рисунок 1*. График зависимости времени добавления сущностей сети (транзакций и блоков) в базу данных от ее сквозного индекса

Точки графика, относящегося к транзакциям всегда принадлежат площади фигуры, находящейся под графиком, относящемся к блокам. Точки времени добавления блоков максимально приближаются к графику транзакций, однако не достигают его. На рисунке 2 показан участок приближения точки добавления блока, принадлежащего слоту 14602, к графику транзакций.

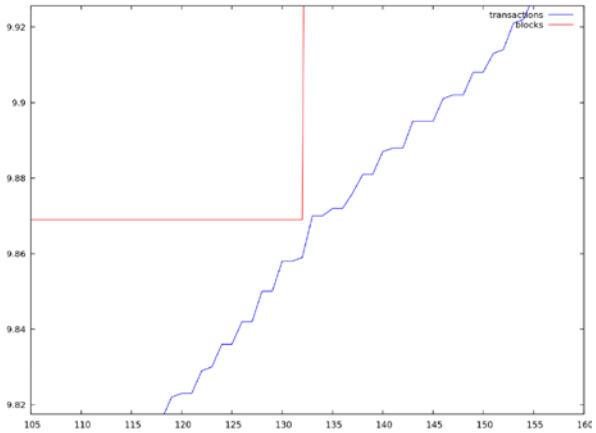



*Рисунок 2*. Участок приближения точки добавления блока, принадлежащего слоту 14602, к графику транзакций

Полученные экспериментальные данные позволяют опровергнуть первую гипотезу.

Ранее было отмечено, что на сегодняшний день нет открытых плагинов, которые были бы заинтересованы в сборе данных о записях через интерфейс geyser. В связи с этим необходимо модифицировать интерфейс таким образом, чтобы он вызывал методы оповещения о записях даже когда нет вызывающего плагина, которому нужно отправить данные. Для этого необходимо модифицировать файлы *geyser_plugin_interface.rs* и *geyser_plugin_manager.rs*. В первом случае надо заменить возвращаемое значение функции *transaction_notifications_enabled* на *true*, эта функция оповестит интерфейс, что есть существуют интересанты получения данных о записях. Во втором случае необходимо изменить функцию *entry_notifications_enabled* таким образом, чтобы она во всех возможных ветках своего выполнения возвращала *true*, это позволит эмулировать поведение внешнего плагина, запрашивающего данные.

Далее необходимо добавить в начало функции *notify_entry*, принадлежащей типажу *EntryNotifier* в имплементации этого типажа для структуры *EntryNotifierImpl* код, представленный в листинге 1.

```
let now = Local::now();
let mut file = OpenOptions::new().append(true).open("entry.txt").unwrap();
file.write_all(
    String::from(format!(
        "slot: {}, index: {}, entry (num_transactions): {}, hash: {}, timestamp:
{}\n",
        slot,
        index,
        entry.num_transactions,
        entry.hash.to_string(),
        now.format("%Y-%m-%d %H:%M:%S%.9f"),)).as_bytes(),);
```

*Листинг 1*. Вставка в функцию *notify_entry*

Код данной вставки будет получать входящий поток оповещений о записях и сохранять информацию о записях, а именно:
– слот, которому принадлежит запись;
– индекс записи в блоке;
– количество транзакций в записи;
– хэш записи;
– временная отметка;
– построчно в указанный файл.

Скорость предложенного способа сбора данных будет значительно отличаться от скорости отправки информации в базу данных плагином, так как здесь исключаются накладные расходы на пересылку данных по сети, но появляются издержки работы с файлами. Для сохранения корректности измерений (в первую очередь временных) было принято решение на текущем этапе эксперимента собирать данные о транзакциях аналогичным образом (сохранять данные в файл, эмулируя наличие плагина-интересанта). В таблице 2 будет представлена информация о транзакциях и записях, которые содержат в себе эти транзакции, все данные принадлежащих слоту 14704.

Таблица 2. Информация о транзакциях и записях

| Индекс транзакции | Время добавления транзакции (T) | Индекс записи | Время добавления записи (E) | Разница (T - E, с) |
|---|---|---|---|---|
| 0 | HH:MM:40.961892434 | 1 | HH:MM:40.961703902 | 0.000188532 |
| 1 | HH:MM:41.276283850 | 52 | HH:MM:41.275962329 | 0.000321521 |
| 2 | HH:MM:41.288145746 | 54 | HH:MM:41.288069011 | 0.000076735 |
| 3 | HH:MM:41.288608823 | 56 | HH:MM:41.288494233 | 0.00011459 |
| 4 | HH:MM:41.295746669 | 58 | HH:MM:41.295608933 | 0.000137736 |
| 5 | HH:MM:41.296041318 | 59 | HH:MM:41.295960419 | 0.000080899 |
| 6 | HH:MM:41.312644970 | 62 | HH:MM:41.312569852 | 0.000075118 |
| 7 | HH:MM:41.313114525 | 63 | HH:MM:41.313040279 | 0.000074246 |
| 8 | HH:MM:41.320464516 | 66 | HH:MM:41.320357877 | 0.000106639 |

| | | | | |
|---|---|---|---|---|
| 9 | HH:MM:41.326665713 | 68 | HH:MM:41.326576463 | 0.00008925 |
| 10 | HH:MM:41.331976224 | 69 | HH:MM:41.331941086 | 0.00003513 |

Полученные данные демонстрируют, что информацию о записи интерфейс отдает раньше, чем информацию о транзакциях, которые этой записи принадлежат. То есть сначала происходит окончательное формирование и закрытие записи (вывод сделан на основе того, что в число получаемой информации о записи входят данные, которые могут быть доступны только после окончательного формирования, например количество транзакций в записи или хэш), отправка оповещения о записи, а после этого отправка оповещений о всех транзакциях.

Для проведения второго этапа исследования также был сгенерирован поток из 1000 транзакций, которые попали на слоты локальной сети 20468-20475. Стоит отметить, что по правилам сети любой блок должен содержать в себе как минимум 64 записи, в связи с этим валидатор на каждом слоте, помимо прочих, генерирует 64 пустых записи. В целях повышения соответствия данных эти записи были исключены из исследовательской выборки. На рисунке 3 представлен результат второго этапа исследования в виде графика (с использованием сквозного индексирования).

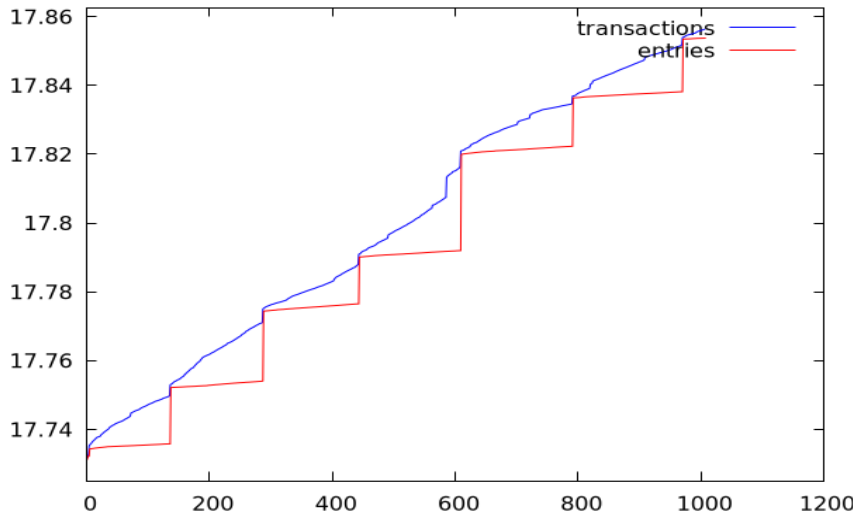



*Рисунок 3*. График зависимости времени добавления сущностей сети (транзакций и записей) в файл от ее сквозного индекса

Ступенчатый вид графика записей объясняется исключением из выборки пустых записей. В целом, полученные данные позволяют подтвердить сделанное ранее предположение о времени отправки оповещения о транзакции. Как и в случае с графиком на рисунке 1 видны точки потенциального касания графиков, однако ни в одной из точек касания не происходит. На рисунке 4 показан участок приближения точки добавления нулевой транзакции, принадлежащей слоту 20473, к графику записей.

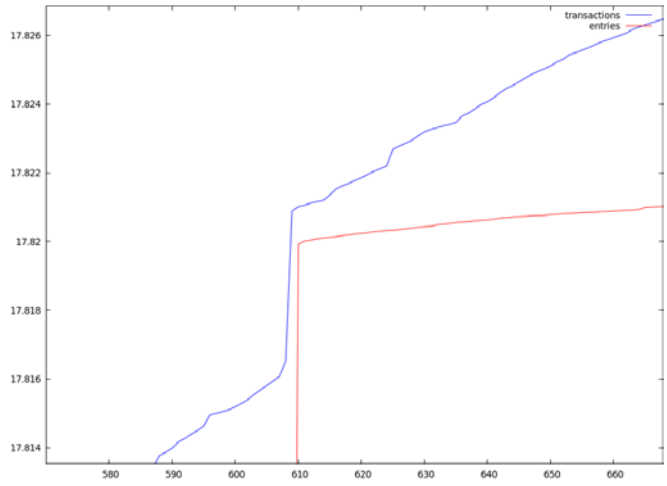



*Рисунок 4*. Участок приближения точки добавления нулевой транзакции, принадлежащей слоту 20473, к графику записей

Таким образом было подтверждено, что информация о транзакциях отправляется интерфейсом geyser после закрытия записи, которой эта транзакция принадлежит.

**Вывод.** На основе практического исследования ответом на вопрос, который был поставлен в начале статьи, является гипотеза номер 2. Информация о транзакциях интерфейсом валидатора Solana geyser отправляется после добавления транзакции в запись и закрытия записи. Полученный результат будет ценен при проектировании прокси-узлов для валидаторов и инструментов анализа сети Solana.

## Список литературы

[1] Mishra D.P. и др. Solana blockchain technology: a review // International Journal of Informatics and Communication Technology (IJ-ICT). Institute of Advanced Engineering and Science. – 2024. – Т. 13, № 2. – С. 197.

[2] Qiu J. A Hybrid Data Decomposition and Deep Learning Approach for Solana Price Prediction Incorporating Market Factors // Advances in Economics, Management and Political Sciences. EWA Publishing. – 2025. – Т. 162, № 1. – С. 6–14.

## Авторский вклад

**Бокун Артем Геннадьевич** – составление плана эксперимента, сбор и анализ данных о транзакциях на локальном валидаторе с использованием специального программного обеспечения, интерпретация и визуализация результатов исследования.

**Герчик Артем Вадимович** – проведение этапов исследования с использованием программного обеспечения, предназначенного для генерации потока транзакций, сбор данных, анализ полученных результатов, формирование структуры статьи.

# OPTIMIZING TRANSACTION DATA COLLECTION TIME IN THE SOLANA NETWORK

***A.G. Bokun***
*Assistant, Department of Computer Science, BSUIR.*

***A.V. Gerchik***
*Assistant, Department of Computer Science, BSUIR.*

**Abstract.** The paper studies the work of one of the components of the Solana network validator systems – geyser, the module for collecting and broadcasting data on past transactions. In the introduction the importance and relevance of the issue of data collection in high-performance networks is outlined. The main part of the article contains an experiment that allows us to learn more about the subtleties of the standard solution. The practical value of the obtained results is evaluated in the conclusion.

**Keywords**: Solana, blockchain, cryptography, geyser, validation, transactions.

УДК 004.413:621.396.969.3

# ПРИЗНАКИ РАДИОЛОКАЦИОННОГО РАСПОЗНАВАНИЯ БАЛЛИСТИЧЕСКИХ ОБЪЕКТОВ

***В. Л. Кожуховский***
*Инженер-программист 1 категории компании ОАО «КБ Радар», магистр*
koozhihooyski@gmail.com

***А. С. Солонар***
*Старший научный содрудник компании ОАО «КБ Радар», кандидат технических наук, доцент*
andssnew@yandex.ru

***А. Ю. Брынчикова***
*Инженер-программист 2 категории компании ОАО «КБ Радар»*
brin4ikova@yandex.by

***В. Л. Кожуховский***

*Окончил Белорусский Государственный Университет. Область научных интересов связана с разработкой алгоритмов обнаружения опасных баллистических объектов на внеатмосферном участке полета.*

***А. С. Солонар***

*Окончил Военную Академию Республики Беларусь. Область научных интересов связанна с автоматизацией процесса обработки радиолокационной информации.*

***А. Ю. Брынчикова***

*Окончила Белорусский Государственный Университет Информатики и Радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов выделения радиолокационных признаков для распознавания баллистических объектов разного класса.*

**Аннотация.** Выполнен анализ существующих методов и признаков радиолокационного распознавания баллистических объектов. Разработано ПО позволяющее проводить моделирование алгоритмов распознавания основанных на радиолокационных признаках баллистических объектов.

**Ключевые слова:** Псевдо-моменты Цернике, «значащие» отсчеты, коэффициент корреляции, дисперсия.

**Введение.** В современных условиях развития аэрокосмических технологий и средств космического нападения остро возникает вопрос выделения опасных объектов из группы наблюдаемых объектов на внеатмосферном участке траектории их полета. Эта задача решается при помощи различных средств наблюдения, в том числе, при помощи многофункциональных радиолокационных станций дальнего наблюдения. В этих станциях космические объекты наблюдаются на дальностях от 1000 до 10000 км. В качестве инструментария разделения объектов по дальномерному портрету используется многочастотный зондирующий сигнал с эквивалентной шириной спектра от 300 до 1000 МГц. Повышение разрешающей способности до 10-ков сантиметров повышает качество различения опасных объектов на фоне других объектов, однако, высокая степень подобия пространственного движения и габаритные размеры объектов затрудняют решение задачи распознавания. Этот факт позволяет говорить об необходимости выделения группы радиолокационных признаков баллистических объектов для эффективного их обнаружения на дальних рубежах обороны.

**Выбор признаков распознавания баллистических объектов.** Эффективность радиолокационного распознавания баллистических объектов (БО) во многом определяется выбором информативных признаков, позволяющих с высокой точностью дифференцировать цели. В рамках данного исследования разработаны и реализованы признаки, обеспечивающие эффективное разделение БО на классы «Боеголовка (RW)», «Ложная цель (Decoy)» и «Сфера (Sphere)».

Процесс выбора признаков базировался на анализе статистических свойств параметров радиолокационных портретов и их динамики изменения. Основной целью являлось выделение параметров, обладающих максимальной контрастностью при минимальной чувствительности к шумам наблюдения и вариативности исходных данных.

По результатам анализа известной литературы [1, 7, 10] и результатов математического моделирования динамики полета БО и выделения их портретов был выбран следующий набор радиолокационных признаков:

1 Диапазон разброса оценок ширины корреляционных функций «значащих» отсчетов дальномерного портрета буфера размером 32 элемента (обозначение – D_width).

2 Среднее значение ширины корреляционных функций (ширины энергетических спектров) «значащих» отсчетов дальномерного портрета в окне 32 интервала наблюдения (обозначение – width).

3 Среднее значение дисперсии «значащих» отсчетов дальномерного портрета в окне 32 интервала наблюдения (обозначение – disper).

4 Коэффициент корреляции текущего дальномерного портрета с его усредненным значением в буфере размером 32 элемента (обозначение – covar).

5 Угол в многомерном пространстве отсчетов дальномерного портрета между точками текущего дальномерного портрета с его усредненным значением в буфере размером 32 элемента (обозначение – tetta).

6 Псевдо-моменты Цернике «значащих» отсчетов дальномерного портрета буфера размером 32 элемента.

***Оценка первых пяти радиолокационных признаков.*** Боеголовки и сферы по своим баллистическим параметрам относятся к стабилизованным БО, ложные цели являются нестабилизированными БО. На рисунке 1 приведен парный график первых пяти признаков распознавания для стабилизированных и нестабилизированных БО.

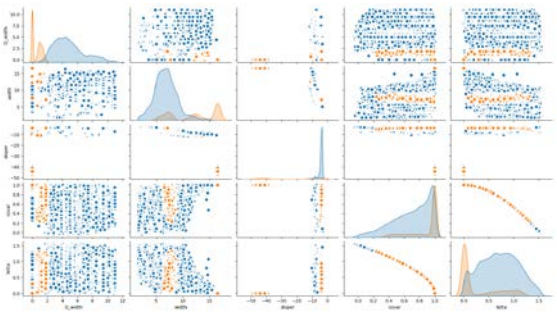


*Рисунок 1*. Парный график признаков распознавания для стабилизированных и нестабилизированных БО

Признак 3 обладает относительно высокой корреляцией со всеми остальными признаками (рисунок 1), между признаки 4 и 5 наблюдается линейная зависимость, признаки 1 и 2 сильно пересекаются между собой. Наиболее контрастными являются признаки 1 и 5 для различения стабилизированных и нестабилизированными БО. Исходя из приведённых рассуждений для селекции стабилизированных и нестабилизированных БО предлагается использовать двумерное пространство с признаками 1 и 5.

При отборе признаков селекции БГ от сфер был построен парный график признаков распознавания (рисунок 2), который показывает взаимосвязь между всеми возможными парами признаков распознавания этих двух типов объектов. Помимо приведенных выше

признаков использовались еще два дополнительных: число «значащих» отсчетов дальномерного портрета (обозначение – nonzero_buffer); среднее значение дисперсии «значащих» отсчетов за все время наблюдения (обозначение – all_disp).

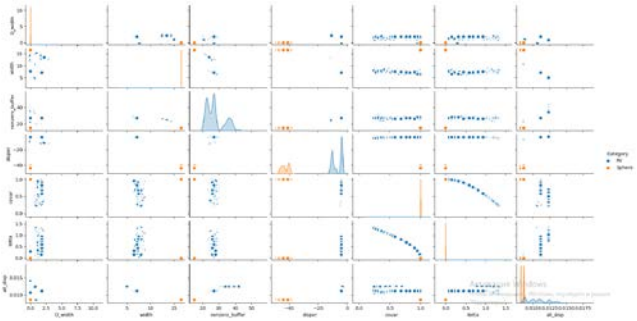


*Рисунок 2.* Парный график признаков разделения БГ и ложных целей сферической формы

Исходя из результатов, отображенных на рисунках 1,2 можно судить, что для селекции БГ на фоне сфер необходимо использовать одномерное пространство признаков – среднее значение дисперсии «значащих» отсчетов дальномерного портрета в окне 32 интервалов наблюдения. Для селекции стабилизированных и нестабилизированных БО использовать двумерное пространство с признаками [8,9] – угол в многомерном пространстве отсчетов дальномерного портрета между точками текущего дальномерного портрета с его усредненным значением в буфере 32; диапазон разброса оценок ширины корреляционных функций «значащих» отсчетов дальномерного портрета буфера 32.

***Псевдо-моменты Цернике.*** В публикациях [2-6] посвященных классификации баллистических объектов, предлагается использовать псевдо-моментов Цернике (pZ-моментов). По утверждениям авторов, данные признаки демонстрируют высокую эффективность за счет устойчивости к геометрическим преобразованиям и шумам наблюдения, к их преимуществам относят:

1 Инвариантность к повороту, масштабу и смещению. pZ-моменты позволяют формировать признаки, которые не зависят от ориентации объекта, его размеров и положения в пространстве. Это критически важно в условиях изменяющихся траекторий и различных углов обзора РЛС.

2 Высокая информативность признаков. Преобразование высокочастотного радиолокационного профиля цели (HRRP) с помощью обратного преобразования Радона (IRT) позволяет получить двумерное представление объекта, содержащее его пространственную структуру и динамические характеристики движения. Из данного представления извлекаются pZ-моменты, обладающие высокой дискриминирующей способностью.

3 Устойчивость к микродвижениям целей. В отличие от традиционных методов, pZ-моменты позволяют минимизировать влияние случайных вариаций и шумов, возникающих из-за прецессии, нутации и вращения объектов.

4 Эффективность в условиях низкого уровня сигнала. Благодаря свойству накопления энергии в обратном преобразовании Радона, метод pZ-моментов позволяет уверенно классифицировать объекты даже при низком сигнал-шумовом отношении (SNR).

*К недостаткам и ограничения признаков псевдо-моментов Цернике относят:*

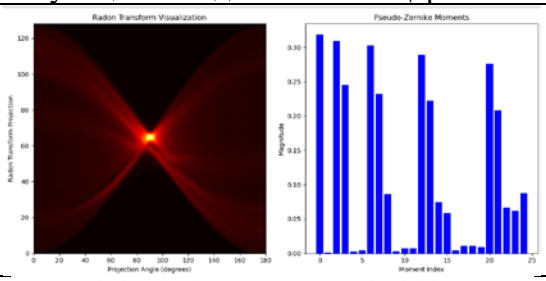
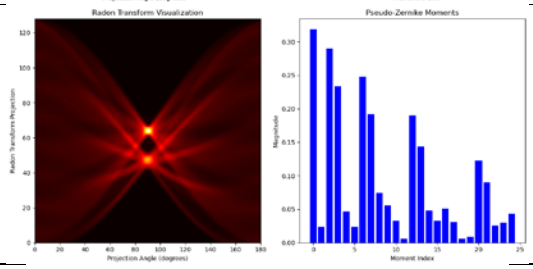
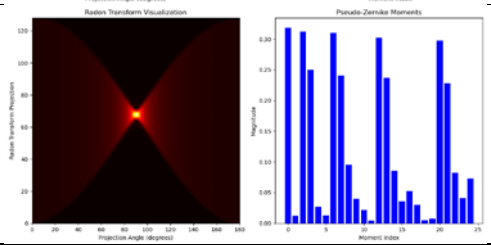
1 Высокая вычислительная сложность. Вычисление pZ-моментов требует значительных ресурсов, особенно при использовании их высоких порядков. Это может быть критичным в системах реального времени с ограниченными вычислительными мощностями.

2 Чувствительность к качеству входных данных. Хотя метод устойчив к микродвижениям, значительные искажения HRRP (например, вызванные помехами, джаммингом или ошибками измерений) могут повлиять на точность распознавания.

3 Зависимость от параметров РЛС. Оптимальность работы алгоритма pZ-моментов может зависеть от характеристик радиолокационной системы, таких как частотный диапазон, разрешающая способность и поляризация сигнала.

При проведении исследований методом математического моделирования были получены псевдо-моментов Цернике для набора из 64 БО. Для каждого из объектов были рассчитаны прямые и обратные преобразования Радона и 25 моментов Цернике. Пример полученных результатов для трех типовых БО – боеголовки, сферы и ложной цели представлен в таблице 1.

Таблица 1. Преобразование Радона и соответствующие псевдо-моменты Цернике

| | |
|---|---|
| БО – DECOY, форма – «Конус».<br>Вес – 16 кг.<br>Длина – 3 м., радиус – 0.15-0.5 м.<br>Скорость вращения – 24.8 град/сек. |  |
| БО – RW, форма – «Конус».<br>Вес – 700 кг.<br>Длина – 3.44 м., радиус – 0.06-0.6 м.<br>Скорость вращения – 1003 град/сек. |  |
| БО – SPHERE, форма – «Шар».<br>Вес – 0.1 кг.<br>Радиус – 0.5 м. |  |

С целью выявления из основных моментов Цернике признаков обладающих минимальной корреляцией между собой и обеспечивающих максимальную контрастность была рассчитана ниже таблица корреляции, см. рисунок 3.

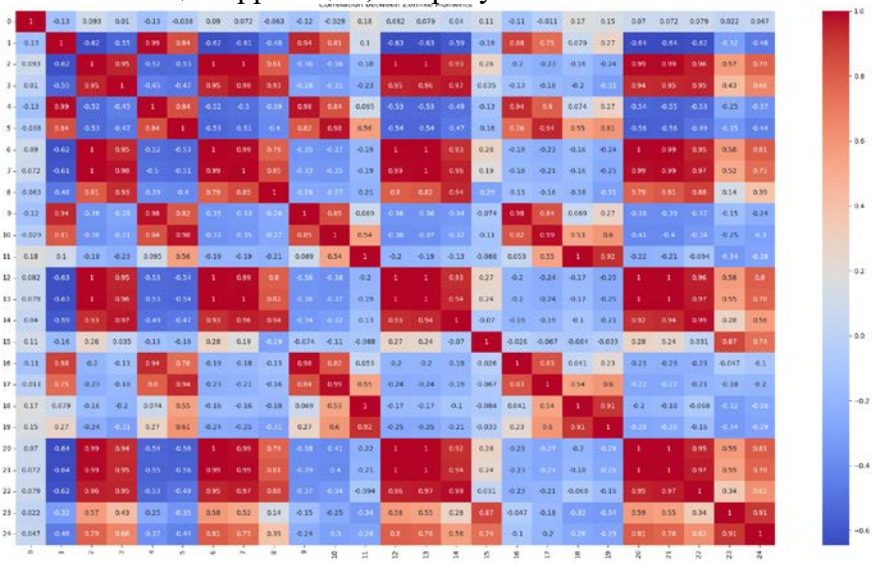


*Рисунок 3.* Таблица корреляции между псевдо-моментами Цернике

Для выбора наиболее эффективных моментов Цернике использовался коэффициент разделимости:

$$F_j = \frac{\left(\mu_{RW,j} - \mu_{Decoy,j}\right)^2}{\sigma^2_{RW,j} + \sigma^2_{Decoy,j}},$$

*где* $\mu_{RW,j}$ *– среднее значение признака j для класса RW,*

$\mu_{Decoy,j}$ *– среднее значение признака j для класса Decoy,*

$\sigma^2_{RW,j}$ *– дисперсия признака j в классе RW,*

$\sigma^2_{Decoy,j}$ *– Дисперсия признака j в классе Decoy.*

Результаты анализа коэффициента разделимости (*Fisher Score*) моментов Цернике представлены в таблице 2.

Таблица 2. Коэффициенты разделимости для моментов Цернике.

| M Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fisher Scores | 0.11411751 | 0.035960 66 | 0.193930 33 | 0.175896 92 | 0.113530 89 | 0.180 97 15 2 | 0.187 67 77 7 | 0.195 51 33 2 | 0.173 74 01 | 0.357338 83 | 0.476428 1 | 0.201 51 25 1 | |
| M Number | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| Fisher Scores | 0.19948478 | 0.197207 95 | 0.207590 93 | 0.054602 82 | 0.684742 11 | 0.800 046 19 1 | 0.217 14 50 5 | 0.386 76 89 9 | 0.197 32 62 6 | 0.207383 33 | 0.236157 86 | 0.125 41 22 2 | 0.143 66 69 |

Коэффициент *Fisher Score* показывает, насколько различаются RW и Decoy по каждому моменту. Чем выше значение, тем лучше этот момент различает классы. Моменты с индексами 17, 18 обладают наибольшей контрастностью. Они могут быть основными признаками для классификации.

**Заключение.** В данной работе рассмотрены признаки радиолокационного распознавания баллистических объектов (БО) на внеатмосферном участке траектории их полета. Проведенный анализ показал, что выбор информативных признаков является ключевым фактором для эффективного разделения объектов на классы «Боеголовка (RW)», «Ложная цель (Decoy)» и «Сфера (Sphere)».

Разработанный набор признаков на основе дальномерных портретов БО, включающий параметры корреляции, дисперсии и углов в многомерном пространстве, позволил отобрать двумерное пространство признаков для селекции стабилизированных и нестабилизированных БО. Полученный набор признаков подходит для применения алгоритмов машинного обучения решая при этом основную задачу – классификации баллистических объектов.

Были рассмотрены псевдо-моменты Цернике (pZ-моментов), полученные после обратного преобразования Радона. Данные признаки обеспечивают высокую устойчивость к геометрическим преобразованиям и шумам, что делает его перспективным для применения в системах радиолокационного наблюдения. Расчет коэффициента разделимости (Fisher Score) позволил выявить наиболее значимые моменты, обеспечивающие наибольшую дискриминационную способность. Полученный набор псевдо-моментов Цернике при решении задачи классификации БО можно легко

использовать в методах машинного обучения, а полученные портеры на основе преобразования Радона можно использовать для обучения нейронных сетей, решающих ту же задачу классификации.

Таким образом, предложенные признаки позволяют обеспечить высокую контрастность результатов наблюдений стабилизированных и нестабилизированных БО по результатам радиолокационного наблюдения.

## Список литературы

[1] A.M. Sessler, J.M. Cornwall, B. Dietz, S. Fetter, S. Frankel, R.L. Garwin, K. Gottfried, L. Gronlund, G.N. Lewis, T.A. Postol, and D.C. Wright, “Countermeasure: A technical evaluation of the operational effectiveness of the planned us national missile defense system,” Tech. Rep., Union of Concerned Scientists MIT Security Studies Program, 2000. – 420p.

[2] Y. Hua, J. Guo, and H. Zhao, “The usage of inverse-radon transformation in isar imaging,” in 2014 IEEE International Conference on Control Science and Systems Engineering, Dec 2014 – pp.167-170.

[3] C. Clemente, L. Pallotta, I. Proudler, A. DeMaio, J.J. Soraghan, and A. Farina, “Pseudo-zernike-based multi-pass automatic target recognition from multi-channel synthetic aperture radar,” Radar, Sonar Navigation, IET,vol.9,no.4, 2015. – pp. 457-466.

[4] H. Gao, L. Xie, S.Wen, andY. Kuang, “Micro-doppler signature extraction from ballistic target with micro-motions,” IEEE Transactions on Aerospace and Electronic Systems, vol. 46,no.4, Oct2010 – pp. 969–982.

[5] M.E. Clark, “High range resolution techniques for ballistic missile targets,” Tech.Rep., British Aero space Land and Sea Systems, Newport Road, Cowes, Isle of Wight, PO318PF, United Kingdom, British Aerospac epic. 1999 – 860p.

[6] S.R.Deans, The Radon Transform and Some of Its Applications, Dover Books on Mathematics Series. Dover Publications, 2007 – 167p.

[7] Paulraj, A.; Roy, R.; Kailath, T., Estimation Of Signal Parameters Via Rotational Invariance Techniques - Esprit, *Nineteenth Asilomar Conference on Circuits, Systems and Computers,* 1985. – pp. 83–89.

[8] L. Li-hua, W. Zhuang, and H. Wei-dong, “Precession period extraction of ballistic missile based on radar measurement,” in 2006 CIE International Conference on Radar, Oct 2006 – 227p.

[9] Zhuo, Z.Y.; Zhou, Y.; Du, L.; Ren, K.; Li, Y. A Noise Robust Micro-Range Estimation Method for Precession Cone-Shaped Targets. Remote Sens., 2021 – 1820p.

[10] Тагаев Т. И., Музафаров Д. А., Мохорт Д. Ю., Космачев В. М., Определение оптимальных тактико-технических характеристик антенной решетки для различения алгоритмом сверхразрешения ESPRIT целей, разнесенных на малое угловое расстояние // Антенны и распространение радиоволн. — СПб.: Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина) (Санкт-Петербург), 2021. — С. 96-98.

## Авторский вклад

**Кожуховский Владислав Леонидович** – реализация и тестирование алгоритмов выделения признаков распознавания баллистических объектов. Исследование эффективности применения псевдо-моментов Цернике на радиолокационных портретах баллистических объектов различного класса.

**Солонар Андрей Сергеевич** – руководство исследованием, постановка задачи для исследования, проведение совместных исследований.

**Брынчикова Анастасия Юрьевна** – разработка метода выделения радиолокационных признаков для распознавания баллистических объектов, тестирование группы отобранных признаков на баллистических объектах различного класса.

# FEATURES OF RADAR RECOGNITION OF BALLISTIC OBJECTS

***V. L. Kozhuhovsky***
*Software engineer, 1st category, ISC "KB Radar", master’s degree*

***A. S. Solonar***
*Senior researcher, ISC "KB Radar", PhD in technical sciences, Associate professor*

***A. Yu. Brynchikova***
*Software engineer, 2nd category, ISC "KB Radar"*

**Abstract**. This article analyzes existing methods and features of radar recognition of ballistic objects. A software system has been developed that allows the simulation of recognition algorithms based on radar features of ballistic objects.

**Keywords**: Zernike pseudo-moments, "significant" samples, correlation coefficient, variance.

УДК 004.021:004.75

# ОЦЕНКА ЗНАЧИМОСТИ ПАРАМЕТРОВ ОБУЧЕНИЯ МОДЕЛИ НЕЙРОННОЙ СЕТИ НА ПРИМЕРЕ КЛАССИФИКАТОРА ОБЪЕКТОВ НА СНИМКАХ ЗЕМНОЙ ПОВЕРХНОСТИ

***Е.И. Ковшер***
*ассистент кафедры электронных вычислительных машин БГУИР, магистрант*
elizavetakovsher@mail.ru

***Н.В. Лапицкая***
*кандидат технических наук, зав. кафедрой программного обеспечения информационных технологий, БГУИР*
lapan@bsuir.by

***Е.И. Ковшер***

*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием искусственного интеллекта и методов машинного обучения.*

***Н.В. Лапицкая***

*В 1994 году окончила факультет прикладной математики Белорусского государственного университета по специальности «прикладная математика». В 2006 году присуждена ученая степень кандидата технических наук по специальности 05.13.17 – Теоретические основы информатики. Область научных интересов связана со статистическим анализом нерегулярных данных, методами многокритериальной оптимизации, логико-вероятностными моделями, управлением рисками.*

**Аннотация.** В исследовании проведена оценка значимости параметров обучения модели нейронной сети на примере классификатора объектов на снимках земной поверхности. Представлен комплексный анализ влияния различных факторов («объем обучающей выборки», «количество батчей» и «количество эпох обучения») на точность и устойчивость модели, что позволило определить оптимальные настройки для её обучения.

Особое внимание уделяется оценке эффективности предложенного решения. Результаты проведённых экспериментов подтверждают превосходство разработанной нейросетевой архитектуры по таким критериям, как точность классификации, скорость обработки изображений и устойчивость к изменению условий съёмки, таких как освещение, угол обзора и разрешение снимков.

Представленные выводы могут быть полезны для дальнейшего совершенствования моделей машинного обучения, используемых в задачах классификации объектов на снимках земной поверхности.

**Ключевые слова:** Машинное обучение, нейронные сети, DARKNET, YOLOv3, обработка геоинформационных данных.

**Введение.** Классификация объектов на снимках земной поверхности – важная задача, которая используется в различных областях науки, техники и экономики. Классификация позволяет извлекать ценные данные из снимков, превращая их в инструмент для анализа и принятия решений. Благодаря этому возможно более эффективное управление ресурсами, планирование и защита окружающей среды.

Объектом исследования являются снимки земной поверхности, а также объекты на снимках земной поверхности, полученных средствами дистанционного зондирования.

В качестве предмета исследования выступают методы и алгоритмы автоматического распознавания, сегментации, классификации и анализа объектов на снимках земной поверхности, основанные на современных технологиях обработки изображений и машинного обучения.

Основная цель данной работы – оценка значимости параметров обучения модели нейронной сети на примере классификатора объектов на снимках земной поверхности.

**Описание модели нейронной сети для автоматического распознавания и анализа объектов на снимках земной поверхности.** Тактическое планирование представляет собой определение способа проведения каждой серии испытаний машинной модели, предусмотренных планом эксперимента, и оно связано с вопросами эффективного использования выделенных для эксперимента машинных ресурсов [2].

Чем сложнее машинная модель *M*, тем важнее этап тактического планирования машинного эксперимента. Процесс планирования машинных экспериментов с моделью *M* носит итерационный характер, т. е. при уточнении некоторых свойств моделируемой системы *S* этапы тактического планирования экспериментов могут чередоваться. Далее будет представлена итоговая модель, в которой будут выделены наиболее значимые факторы, влияющие на точность распознавания объектов [3].

Модель системы базируется на фреймворке *DARKNET* [4] – это высокопроизводительный, гибкий и легковесный фреймворк для разработки и обучения нейронных сетей, написанный на языке программирования *C* с использованием *CUDA* и *OpenCV* для ускорения вычислений на *GPU*.

Основой фреймворка *DARKNET* является модель *YOLOv3*. Это одна из наиболее популярных моделей для обнаружения объектов на изображениях. *YOLOv3* принадлежит к семейству одноэтапных детекторов, которые обрабатывают изображение целиком за один прогон через сеть, что делает их значительно быстрее, чем двухэтапные подходы (например, *Faster R-CNN*).

Модель представляется в виде черного ящика (1):

$$\eta = \eta(x, b), \tag{1}$$

где $x$ – факторы (входные величины); $b$ – неизвестные параметры (коэффициенты); $\eta$ – реакция системы (выходная величина).

Целью анализа экспериментальных данных является определение оценок неизвестных параметров $b$ в некоторой заданной области факторного пространства $X$. Графическое представление статистической модели системы демонстрирует рисунок 1. В реальных условиях, из-за наличия помехи $\varepsilon$, вместо истинного значения выходной величины $\eta$ необходимо измерять величину $Y$. Следовательно, опираясь на результаты измерения, нельзя получить абсолютно точные значения $b$. Вместо истинных параметров $b$ необходимо получить оценки параметров $\beta$. Тогда, оцениваемое уравнение для модели будет иметь вид, представленный в формуле (2).

*Рисунок 1.* Статическая модель системы

$$Y = Y(x, \beta), \tag{2}$$

где $x$ – факторы (входные величины); $\beta$ – оценка неизвестных параметров (коэффициентов); $Y$ – реакция системы с учетом помехи $\varepsilon$ (выходная величина).

**Общая структура планирования эксперимента.** Основные этапы планирования эксперимента изложены ниже.

Вначале происходит определение цели эксперимента. В этот этап входит формулировка проблемы, которую необходимо проверить, и определение того, что нужно исследовать. Цель эксперимента, проводимого в рамках работы, заключается в исследовании влияния различных факторов – «объем обучающей выборки», «количество батчей» и «количество эпох обучения» – на точность распознавания объектов на снимках земной поверхности с использованием фреймворка *DARKNET* и модели *YOLOv3*.

Следом за определением цели эксперимента следует выбор факторов и уровней их варьирования. Данная работа предполагает полный факторный эксперимент: исследуются все возможные комбинации уровней факторов. Этот этап включает в себя определение факторов, которые могут влиять на результат эксперимента, и выбор значений (уровней) для каждого фактора. Для того чтобы понять, как факторы влияют на эффективность модели *YOLOv3* при распознавании объектов на снимках земной поверхности, эти факторы («объем обучающей выборки», «количество батчей» и «количество эпох обучения») будут варьироваться в ходе эксперимента. Каждый из этих факторов может существенно повлиять на результаты, и, играя с их уровнями, представляется возможным найти оптимальные настройки для поставленной задачи.

Чем больше изображений используется для обучения, тем лучше модель «учится» обобщать данные и распознавать различные объекты, но больше данных требуют больше времени и вычислительных мощностей. Данный фактор разделен на два уровня. Уровень 1 (низкий) – 100 изображений. Меньший объем данных может быть полезен для первоначальных экспериментов, чтобы быстро проверить, как модель обучается. Подходит для небольших проектов или быстрых прототипов. Уровень 2 (высокий) – 800 изображений. Больший объем данных позволяет модели «видеть» больше примеров, что делает её более устойчивой и точной при распознавании объектов, особенно в сложных или изменчивых условиях.

Количество батчей влияет на то, сколько примеров будет обрабатываться одновременно перед обновлением модели. Меньшие батчи дают больше вариативности, но и могут быть менее стабильными. Уровень 1 (низкий) – 8 батчей. Меньшие батчи могут быть полезны для более частых обновлений весов, что иногда ускоряет процесс обучения, но на меньших батчах точность обновлений может быть не такой высокой. Уровень 2 (высокий) – 16 батчей. Большие батчи дают более точные обновления, так как модель видит больше примеров перед изменением своих весов. Это помогает сделать обучение более стабильным, но требует больше памяти.

Эпохи – это количество полных проходов через все данные. Большее количество эпох позволяет модели дольше «учиться», но может привести к переобучению, если количество эпох слишком велико. Заключительный фактор так же включает в себя два уровня, представленных ниже. Уровень 1 (низкий) – 10 эпох. Это количество эпох подходит для начальных тестов, когда нужно проверить основные параметры модели. Меньше эпох также помогает избежать чрезмерного переобучения. Уровень 2 (высокий) – 50 эпох. С большим количеством эпох модель будет иметь больше времени для обучения, что помогает достигать лучших результатов на более сложных задачах. Однако важно следить за переобучением, если модель начинает «запоминать» тренировочные данные.

Каждый из 3 предложенных факторов будет варьироваться между двумя уровнями, что даст 8 возможных комбинаций для экспериментов ($2^3 = 8$). Это даст возможность исследовать, как каждый параметр влияет на производительность модели *YOLOv3* при распознавании объектов на изображениях земной поверхности.

Следующим этапом планирования эксперимента является определение выходных переменных (откликов). Это показатели, которые будут измеряться в ходе эксперимента.

Средняя точность ($mAP$) – это один из наиболее важных откликов для оценки производительности модели при решении задач распознавания объектов, таких как распознавание объектов на снимках земной поверхности. Средняя точность (*mAP)* учитывает как точность классификации, так и правильность локализации, формула (3):

$$mAP = \frac{1}{C}\sum_{c=1}^{C}\frac{1}{N_c}\sum_{k=1}^{N_c}\frac{TP_c(k)}{TP_c(k)+FP_c(k)}, \quad (3)$$

где $C$ – количество классов (в рассматриваемой модели 60); $N_c$ – количество точек отсечения для класса $c$; $TP_c(K)$ и $FP_c(k)$ – количество истинных и ложных положительных предсказаний для класса $c$ на уровне отсечения $k$.

Далее происходит обработка данных. Сбор и анализ результатов с использованием статистических методов – является неотъемлемой частью планирования эксперимента.

На этапе обработки данных собираются результаты всех проведенных экспериментов, выполняется их систематизация и проводится статистический анализ для выявления значимости влияния факторов на выходную переменную ($mAP$).

Поскольку эксперимент является полнофакторным, каждый фактор кодируется с помощью бинарных значений (0 и 1) для удобства последующего анализа. Низкий уровень – значение 0, высокий уровень – значение 1.

В ходе эксперимента на каждом этапе фиксировался показатель *mAP* (основной отклик, который используется для оценки точности модели). Все собранные данные сведены в таблицу (расширенную матрицу полнофакторного эксперимента), где каждая строка соответствует одной комбинации факторов. Таблица 1, содержащая эти данные, представлена ниже. Значения *mAP* рассчитывались для двух изображений, местность которых содержат разные классы объектов, пригодных к распознаванию.

Таблица 1. Расширенная матрица полнофакторного эксперимента

| *Номер эксперимента* | *Объем обучающей выборки* ($X_1$) | *Количество батчей* ($X_2$) | *Количество эпох* ($X_3$) | *mAP (1, 2)* |
|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0,124400, 0,107424 |
| 2 | 0 | 0 | 1 | 0,210523, 0,171322 |
| 3 | 0 | 1 | 0 | 0,122273, 0,107425 |
| 4 | 0 | 1 | 1 | 0,217966, 0,191696 |
| 5 | 1 | 0 | 0 | 0,198827, 0,177805 |
| 6 | 1 | 0 | 1 | 0,316848, 0,278746 |
| 7 | 1 | 1 | 0 | 0,286014, 0,244481 |
| 8 | 1 | 1 | 1 | 0,311532, 0,272263 |

**Дисперсионный и регрессионный анализ полученных в ходе планирования эксперимента данных.** Дисперсионный анализ (*ANOVA*) используется для оценки значимости влияния факторов на выходную переменную – в данном случае на *mAP*.

Общая цель дисперсионного анализа – проверить гипотезу $H_0$, а именно: выявить наиболее значимые факторы, которые существенно влияют на точность модели. Это помогает не только лучше понять влияние каждого параметра, но и сократить время на оптимизацию, сосредоточившись на наиболее важных параметрах. Для этого используется расчет отношения межгрупповой и внутригрупповой дисперсий (4):

$$F = \frac{MS_{\text{фактор}}}{MS_{\text{ошибка}}}, \quad (4)$$

где $MS_{фактор} = SS_{фактор} / df_{фактор}$ – средний квадрат отклонений, обусловленных изменением факторов; $MS_{ошибка} = SS_{ошибка} / df_{ошибка}$ – средний квадрат отклонений, обусловленных случайными ошибками; $SS_{фактор}$ – сумма квадратов отклонений из-за влияния фактора; $SS_{ошибка}$ – сумма квадратов отклонений из-за влияния фактора; $df_{фактор}$ – сумма квадратов отклонений из-за влияния фактора; $df_{ошибка}$ – сумма квадратов отклонений из-за влияния фактора.

Общая сумма квадратов отклонений – это сумма квадратов отклонений всех наблюдений от общего среднего значения отклика (5):

$$SS_{\text{общее}} = \sum_{i=1}^{N} (Y_i - \bar{Y})^2, \tag{5}$$

где $Y_i$ – значение отклика (*mAP*) для *i*-й комбинации факторов; $\bar{Y}$ – общее среднее значение *mAP* по всем экспериментам; *N* – общее количество экспериментов.

Далее для каждого фактора рассчитывается его вклад в изменчивость отклика (6):

$$SS_{\text{фактор}} = n \sum_{j=1}^{2} (\bar{Y}_{\text{фактор}_j} - \bar{Y})^2, \tag{6}$$

где $\bar{Y}_{\text{фактор}_j}$ – среднее значение отклика для *j*-го уровня фактора; *n* – число повторений для каждого уровня.

В заключение рассчитывается сумма квадратов ошибок (7):

$$SS_{\text{ошибка}} = SS_{\text{общее}} - \sum SS_{\text{фактор}}, \tag{7}$$

После расчета дисперсионного анализа можно сделать следующие выводы.

Если *F*-критерий для фактора окажется больше табличного значения при уровне значимости $\alpha = 0{,}05$, значит, данный фактор оказывает статистически значимое влияние на отклик (*mAP*). Если *F*-критерий окажется меньше табличного значения, влияние фактора можно считать незначимым.

Для удобства значения *F*-критериев, полученные для каждого фактора, сведены в таблицу 2.

Таблица 2. Значения *F*-критерия для каждого фактора

| *Фактор* | *Значение F-критерия* |
|---|---|
| Объем обучающей выборки ($X_1$) | 10,997371 |
| Количество батчей ($X_2$) | 0,445487 |
| Количество эпох ($X_3$) | 5,741693 |

В данном случае табличное значение *F*-критерия при уровне значимости $\alpha = 0{,}05$ – 5,41. На основании этого и сформулированной ранее гипотезы $H_0$ сделано заключение о том, что влияние таких факторов, как «объем обучающей выборки» и «количество эпох обучения», является значимыми, в то время как влияние фактора «количество батчей» оказывает меньшее влияния, однако его также нельзя исключать в дальнейший

исследованиях.

Регрессионный анализ используется для построения математической модели, описывающей зависимость выходной переменной (*mAP*) от значений факторов. Цель – определить, какие факторы оказывают значимое влияние на результат, и получить уравнение регрессии, с помощью которого можно предсказать значение *mAP* для новых комбинаций факторов. Для оценки коэффициентов модели используется метод наименьших квадратов (МНК).

Множественная линейная регрессия предполагает, что отклик ($Y$, в рассматриваемом случае – *mAP*) является линейной функцией факторов (8):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon, \quad (8)$$

где $Y$ – выходная переменная ($mAP$); $X_1$, $X_2$, …, $X_n$ – факторы; $\beta_0$ – свободный член; $\beta_1$, $\beta_2$, …, $\beta_n$ – коэффициенты регрессии; $\varepsilon$ – случайная ошибка.

Метод наименьших квадратов предполагает, что минимизируется сумма квадратов отклонений фактических значений отклика от предсказанных моделью (9):

$$S = \sum_{i=1}^{N} (Y_i - \widehat{Y}_i)^2 = \sum_{i=1}^{N} (Y_i - (\beta_0 + \sum_{j=1}^{n} \beta_j X_{ij}))^2, \quad (9)$$

Для нахождения коэффициентов регрессии необходимо решить систему нормальных уравнений (10):

$$X^T X\beta = X^T Y, \quad (10)$$

где $X$ – матрица факторов; $\beta$ – вектор коэффициентов регрессии; $Y$ – вектор откликов.

Регрессионный анализ позволяет получить уравнение, описывающее зависимость *mAP* от параметров модели.

В ходе планирования данного эксперимента было получено следующее уравнение, описывающее зависимость *mAP* от параметров модели (11):

$$Y = 0{,}0904 + 0{,}0628X_1 + 0{,}0713X_2 + 0{,}0655X_3, \quad (11)$$

Анализ полученного уравнения так же подтверждает сформулированную гипотезу $H_0$ и говорит о значительном влиянии выдвинутых в ходе работы факторов.

Результат работы модели представлен на рисунке 2.

*Рисунок 2.* Результат работы модели

**Заключение.** На основе проведенного плана эксперимента был сделан вывод, что выдвинутая гипотеза $H_0$ была подтверждена на основании дисперсионного и регрессионного анализа. Анализ данных, полученных в ходе эксперимента, показал, что увеличение объема обучающей выборки, оптимизация числа батчей и увеличение числа эпох обучения приводят к значительному улучшению точности модели, а также повышают ее устойчивость и производительность. Эти факторы оказали существенное влияние на качество работы модели, что подтверждает первоначальное предположение.

## Список литературы

[1] Ковшер, Е. И. Классификация и анализ снимков земной поверхности = Classification and analysis of earth surface / Е. И. Ковшер, Д. Ю. Перцев // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей X Международной научно-практической конференции, Минск, 13 марта 2024 г. : в 2 ч. Ч. 2 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2024. – С. 152–158.

[2] Адлер Ю.П., Макарова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий, М – 1967.

[3] Шовенгердт Р.А. Дистанционное зондирование. Модели и методы обработки изображений. М.: Техносфера, 2010. – 560 с.

[4] Darknet [Электронный ресурс] – Режим доступа: https://github.com/pjreddie/darknet – Дата доступа: 05.03.2025.

## Авторский вклад

**Ковшер Елизавета Игоревна** – оценка значимости параметров обучения модели нейронной сети на примере классификатора объектов на снимках земной поверхности, планирование и проведение эксперимента, интерпретация полученных результатов.

**Лапицкая Наталья Владимировна** – постановка задачи исследования, анализ полученных результатов.

# ASSESSMENT OF THE SIGNIFICANCE OF THE TRAINING PARAMETERS OF A NEURAL NETWORK MODEL USING THE EXAMPLE OF A CLASSIFIER OF OBJECTS IN IMAGES OF THE EARTH'S SURFACE

***E.I. Kovsher***
*Assistant at the Department of Electronic Computing Machines, BSUIR, PhD student*

***N.V. Lapitskaya***
*PhD of Technical Sciences, Head of the Department of Information Technology Software, BSUIR*

**Abstract.** The research assesses the significance of the training parameters of the neural network model using the example of a classifier of objects in images of the earth's surface. A comprehensive analysis of the influence of various factors ("training data size", "number of batches" and "number of training epochs") on the accuracy and stability of the model is presented, which made it possible to determine the optimal settings for its training.

Particular attention is paid to assessing the effectiveness of the proposed solution. The results of the experiments confirm the superiority of the developed neural network architecture in such criteria as classification accuracy, image processing speed and resistance to changes in shooting conditions, such as lighting, viewing angle and image resolution.

The presented results can be useful for further improvement of machine learning models used in the tasks of classifying objects in images of the earth's surface.

**Keywords**: Machine learning, neural networks, DARKNET, YOLOv3, geoinformation data processing.

УДК 004.942

# МОДЕЛЬ СЛОЖНЫХ БИОЛОГИЧЕСКИХ ТКАНЕЙ ДЛЯ ИССЛЕДОВАНИЯ ВОЗДЕЙСТВИЯ ЭЛЕКТРИЧЕСКОГО ТОКА

***А.Н. Голубович***
*Студентка БГУИР*
anastasiagolubovich5@gmail.com

***П.В. Камлач***
*Заместитель декана факультета компьютерного проектирования БГУИР, кандидат технических наук, доцент*

***Чураков А.В.***
*врач анестезиолог-реаниматолог УЗ «5 ГКБ», кан.мед.наук, доцент*

***В.М. Бондарик***
*доцент кафедры электронной техники и технологии*
*доцент; канд. техн. наук*
*декан факультета доуниверситетской подготовки и профессиональной ориентации*
bondarik@bsuir.by

***А.Н. Голубович***

*Окончила белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методов построения информационно-компьютерных систем.*

***П.В. Камлач***

*Окончил белорусский государственный университет информатики и радиоэлектроники. Область научного интереса – проектирование медицинских электронных систем.*

***А.В. Чураков***

*Врач анестезиолог-реаниматолог УЗ «5 ГКБ», кан.мед.наук, доцент. Область профессиональных интересов/исследований - эфферентные методы терапии, квантовая медицина, программное моделирование медико-биологических технологий в инновационных платформах*

***В.М. Бондарик***

*Окончил Минский радиотехнический институт. Область научного интереса – медицинская электронная техника.*

**Аннотация.** Разработана геометрическая модель мышечной ткани для исследования воздействия электрического тока. Модель включает кожу, электроды, мышцы и промежуточный слой соединительной ткани. Модель создана в программном обеспечении SolidWorks.

**Ключевые слова:** моделирование, воздействие электрического тока, модель мышечной ткани.

**Введение.** Современные методы численного моделирования позволяют детально исследовать процессы распространения электрического тока в биологических тканях, что актуально для медицинской диагностики, физиотерапии и реабилитации. В данной работе разработана геометрическая модель мышечной ткани для анализа электрического воздействия. Модель включает кожу, электроды, мышцы и соединительную ткань, созданную в SolidWorks с учетом анатомических и биофизических особенностей тканей, включая их неоднородность и анизотропию.

Целью работы является исследование влияния электрического тока на мышечные ткани путем построения точной цифровой модели, которая в дальнейшем будет использована для численного анализа в SolidWorks.

Разработка и последующий численный анализ модели позволят оптимизировать параметры электростимуляции и определить наиболее эффективные режимы воздействия. Полученные результаты могут быть применены в физиотерапии, реабилитации, спортивной медицине и разработке персонализированных методов лечения.

**Создание модели в SolidWorks.** Для создания модели в SolidWorks, которая будет использована для дальнейшего численного анализа электрического воздействия в COMSOL Multiphysics, было выполнено несколько этапов. Была построена базовая геометрия, включающая ключевые биологические слои, такие как кожа, фасция и мышцы. Эти слои были созданы с учетом их анатомической структуры и физических характеристик, таких как толщина, форма и взаимное расположение. Кожа была смоделирована как наружный слой, состоящий из двух элементов, который будет контактировать с электродами, фасция — как промежуточный слой между мышечными тканями, а мышцы — как основной объект исследования.

Для увеличения точности модели добавлена мелкомасштабная структура мышечных волокон. Это детализированные элементы, представляющие собой отдельные мышечные волокна, что позволит более точно отразить особенности взаимодействия электрического тока с тканями. Созданы электроды на поверхности кожи. Эти элементы смоделированы в виде тонких пластин, которые контактируют с кожей.

**Описание модели.** Модель состоит из кожи, которая является внешним слоем и служит для контакта с электродами. Далее, добавлена фасция, которая является промежуточным слоем между мышцами и оказывает влияние на распространение электрического тока. Добавлена структура мышечных волокон. Созданы электроды на поверхности кожи. Модель представлена на рисунке 1.

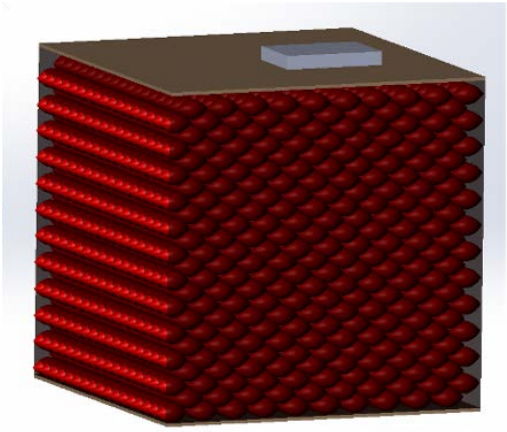


*Рисунок 1.* Модель мышечной ткани

Кожа, которая служит внешним слоем и контактирует с электродами. В реальной анатомической структуре кожи можно выделить несколько слоев, таких как эпидермис и дерма. В модели кожа была представлена как однородный слой толщиной 2мм.

Электроды были размещены на поверхности кожи. Эти элементы были смоделированы как тонкие пластины, представляющие собой контактные поверхности для подачи электрического тока. Размеры и расположение электродов были подобраны, чтобы обеспечить правильное распределение электрического тока.

В модели так же представлена фасция (соединительная ткань), которая служит промежуточным слоем между мышцами. Фасция не только структурно поддерживает мышцы, но и оказывает влияние на распространение электрического тока. Хотя фасция имеет более низкую проводимость по сравнению с мышечной тканью, она все же играет роль в распределении тока, особенно в местах, где она окружает группы мышечных волокон. Модель фасции представлена на рисунке 2.

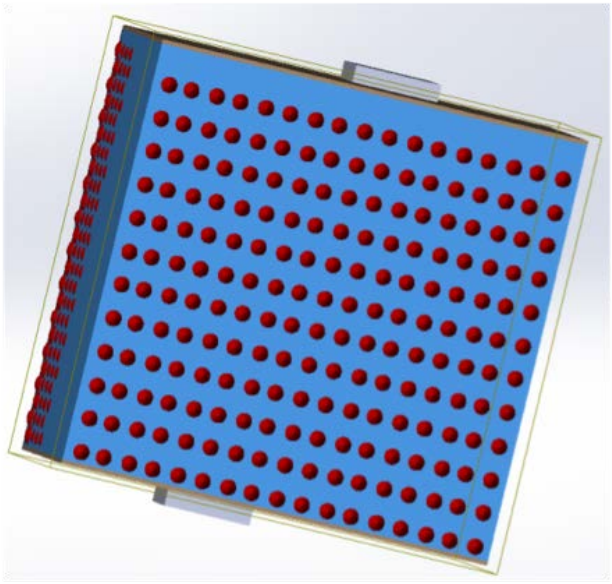


*Рисунок 2.* Модель фасции

Самым важным объектом в модели является мышечная ткань, которая является основным объектом исследования. Мышечные волокна были смоделированы в виде сетчатой структуры, что позволяет точнее отобразить реальные характеристики ткани и ее взаимодействие с электрическими импульсами. Мышечная ткань имеет высокую проводимость, особенно в областях, где расположены отдельные волокна, и она является тем слоем, в котором происходит основная электрическая стимуляция. В модели была использована детализированная структура, которая позволяет более точно смоделировать распределение электрического тока и его воздействие на каждое мышечное волокно.

Каждая часть модели, включая кожу, фасцию и мышцы, была создана с учетом анатомических и физических характеристик реальных тканей. В будующем планируется исследование модели в COMSOL Multiphysics, где будет проведен численный анализ воздействия электрического тока, распределения электрического поля и его влияния на ткани.

**Заключение.** Была разработана геометрическая модель мышечной ткани для исследования воздействия электрического тока. Модель, включающая кожу, фасцию, мышцы и электроды, была создана в SolidWorks с учетом анатомических и биофизических особенностей тканей. После проведения численного анализа в COMSOL Multiphysics, результаты работы будут служить основой для более точных рекомендаций по выбору параметров электростимуляции и расположению электродов, что поможет улучшить эффективность лечения и восстановительных процедур.

**Список литературы**

[1] Гуляев, В. Ю. Электродиагностика, электростимуляция и импульсная низкочастотная электротерапия / В. Ю. Гуляев, В. А. Матвеев, И. Е. Оранский. Екатеринбург, 2004. 104 с.

[2] . Резникова С.В. Терапевтическое применение постоянного электрического тока. Учебное пособие. Астрахань : ГОУ ВПО «АГМА» Росздрава, 2007. — 28 с.

# MODEL OF COMPLEX BIOLOGICAL TISSUES FOR INVESTIGATING THE IMPACT OF ELECTRICAL CURRENT

***A. N. Golubovich***
*Student of BSUIR*

***P.V. Kamlach***
*Deputy dean of the Faculty of Computer Design of BSUIR, PhD of Technical Sciences, Associate Professor*

***A.V. Churakov***
*PhD, anesthesiologist-resuscitator of the State Health Institution "5th City Clinical Hospital"*

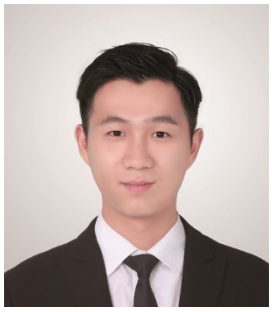
***V.M. Bondarik***
*Associate Professor of the Department of Electronic Engineering and Technology Associate Professor; Ph.D. in Engineering*
*Dean of the Faculty of Pre-University Training and Career Guidance*

**Abstract.** A geometric model of muscle tissue has been developed to study the effects of electric current. The model includes the skin, electrodes, muscles, and an intermediate layer of connective tissue. The model was created using SolidWorks software.

**Keywords**: modeling, electrical current impact, muscle tissue model.

УДК 004.021

# VEHICLE AND LABEL DETECTION METHOD BASED ON *YOLOV11*



***Guo Qicheng***
*PhD student at the department of Infocommunication Technologies BSUIR*
2733064287@qq.com

***Guo Qicheng***
*PhD student at the Department of Infocommunication Technologies, BSUIR. The field of scientific interest is related to the study of computer vision, image processing, and SLAM (Simultaneous Localization and Mapping).*

**Abstract.** This paper proposes a vehicle and label detection method based on the *YOLOv11* algorithm. By constructing a custom dataset of vehicles and labels, the *YOLOv11* model is trained to achieve precise detection of vehicles and their rear-mounted labels. This paper details the dataset creation process, label design, model training procedure, and the selection of optimal training parameters. Finally, the algorithm's performance is evaluated using a validation set. Experimental results demonstrate that the *YOLOv11*-based vehicle and label detection method exhibits strong performance in terms of accuracy and real-time capability, meeting the requirements of practical applications.

**Keywords:** Object Detection, *YOLOv11*, Vehicle Detection, Label Recognition.

**Introduction.** With the rapid advancement of computer vision technology, object detection has been widely applied in fields such as autonomous driving, robotics, and surveillance. In these applications, accurately identifying moving objects and their associated labels is crucial. To achieve precise detection of vehicles and their rear labels, this paper proposes a vehicle and label detection method based on the *YOLOv11* (You Only Look Once Version 11) algorithm.

The *YOLO* series of algorithms [1] has achieved remarkable success in object detection, particularly in real-time detection and efficiency, making it widely adopted. Compared to earlier versions, *YOLOv11* [2] has demonstrated significant improvements in detection speed, accuracy, and scalability. Therefore, this algorithm is selected as the core detection framework for this study.

**Dataset Construction**. To train the *YOLOv11* model, this study first requires the creation of a custom dataset containing images of small vehicles and their corresponding labels. The dataset construction process involves the following steps:

1)Image Acquisition of Vehicles and Labels

The dataset is built using a four-wheeled experimental vehicle model as the primary subject. Photographs of the vehicle and its rear-mounted labels were captured. During the image acquisition process, care was taken to ensure diversity in the dataset by varying perspectives and distances. The label design incorporates easily recognizable visual features, with red rectangular frames selected as the label markers. This study constructed a dataset comprising 660 images. Figure 1 shows a subset of the dataset images featuring the vehicle and its labels.

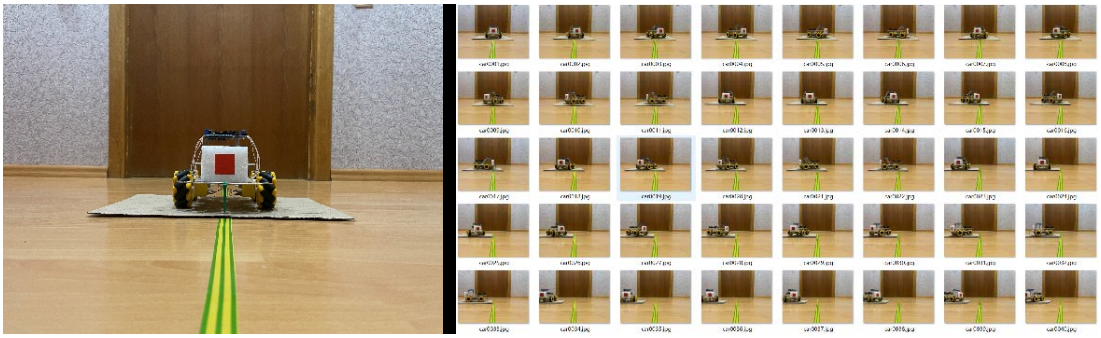


*Figure 1.* Partial dataset images of the vehicle and its corresponding labels.

2)Label Annotation

The captured images were manually annotated using the labeling tool (LabelImg). In each image, the bounding boxes of the vehicles and the positions of the rear labels were annotated. Each label category was associated with its corresponding vehicle to ensure the accuracy of the data annotation. After the annotation process, *YOLO*-format label files were generated, which contain the category information of each object and the coordinates of the bounding boxes.

3) Dataset Partition

The annotated dataset needs to be divided into training, validation, and test sets. Typically, the training set accounts for 70%-80% of the dataset, the validation set for 10%-15%, and the test set for 10%-15% [3]. The partitioning of the dataset should ensure a uniform distribution of samples across all categories to avoid overfitting or underfitting of the model. In this study, the dataset was partitioned in an 8:1:1 ratio.

***YOLOv11* Algorithm Overview.** *YOLOv11* (You Only Look Once Version 11) is the latest iteration in the *YOLO* series of algorithms. It achieves object detection through a single forward pass, simultaneously performing category prediction and location regression, demonstrating highly efficient detection capabilities. *YOLOv11* incorporates state-of-the-art deep learning techniques, such as efficient convolutional networks and more advanced loss functions, significantly improving speed while maintaining detection accuracy.

The *YOLOv11* network architecture primarily consists of three components: the backbone network, the detection head, and the output layer. The backbone network is responsible for feature extraction, while the detection head generates category predictions and bounding box regression values through convolutional operations. By performing a single forward pass on the input image, *YOLOv11* can simultaneously output object categories and locations, greatly enhancing detection efficiency. *YOLOv11* has the following advantages:

– efficiency – *YOLOv11* achieves object detection in a single forward pass, making it significantly faster compared to traditional two-stage detection methods;

– accuracy – By integrating various advanced techniques, *YOLOv11* improves the model's accuracy and robustness;

– scalability – *YOLOv11* is adaptable to a wide range of hardware environments, enabling efficient operation across different devices.

**Experiments and Results Analysis.** The *YOLOv11* framework was employed to train the model for vehicle and label detection. During the training process, hyperparameters such as learning rate, batch size, and weight decay were adjusted, and data augmentation techniques (e.g., rotation, flipping, brightness adjustment) were applied to further enhance the model's generalization capability. Through iterative experiments, appropriate hyperparameters were selected to optimize the training performance. The trained model was evaluated using validation and test sets. Evaluation metrics included mean Average Precision (mAP), recall, and F1-score. By comparing the performance of models under different training parameters, the optimal training strategy was determined.

After multiple rounds of training, the final model achieved a mAP of over 87% on the validation set, with excellent performance in both recall and precision. Figure 2 illustrates the training results. The results on the test set also demonstrated that the *YOLOv11* model can accurately detect vehicles and their rear labels, with detection speed meeting real-time requirements. Figure 3 presents the prediction results.

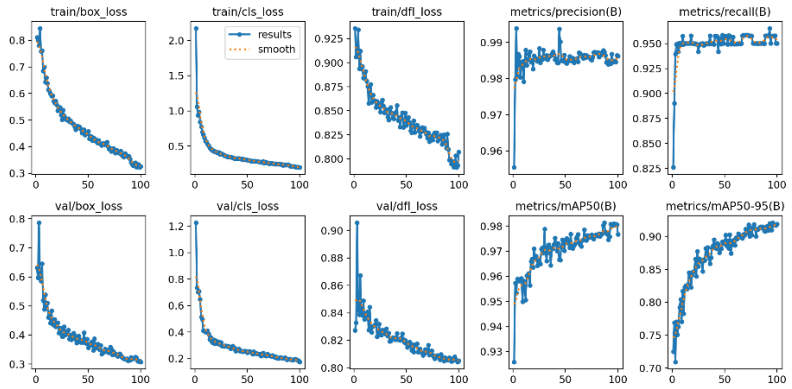



*Figure 2.* Training results

Due to the diversity of the dataset and the data augmentation techniques employed during the training process, the *YOLOv11* model has demonstrated strong robustness, enabling it to effectively handle vehicle detection tasks under varying lighting conditions and perspectives.

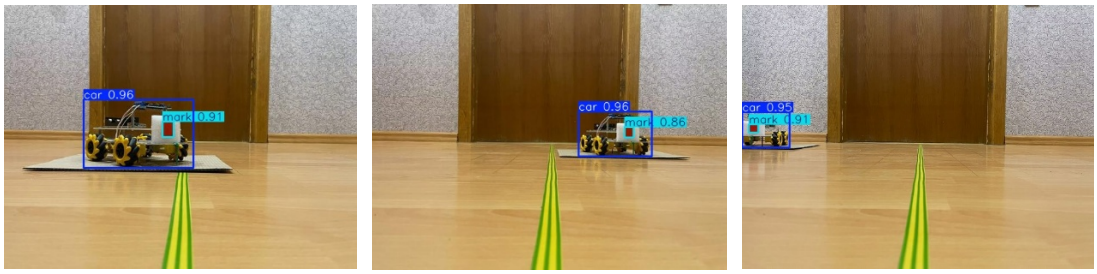



*Figure 3.* Prediction results

**Conclusion.** The proposed method for vehicle and label detection based on *YOLOv11*, through training and optimization on a custom dataset, has achieved promising experimental results. The efficiency and accuracy of *YOLOv11* in object detection enable this method to be effectively applied to real-world vehicle detection tasks. Future work will focus on further optimizing model performance and expanding the dataset to enhance adaptability in more complex environments.

## References

[1] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779-788.

[2] Ultralytics. (2025). YOLOv11 Documentation. Retrieved from https://docs.ultralytics.com/zh/models/yolo11/

[3] Chalapathy, Raghavendra, and Sanjay Chawla. Deep learning for anomaly detection: A survey. IEEE Transactions on Neural Networks and Learning Systems 30.9 (2019): 1-18.

УДК 004.934.2+534.784

# АНАЛИЗ ПОДХОДОВ К ПОСТРОЕНИЮ СИСТЕМ РАСПОЗНАВАНИЯ ЭМОЦИЙ ПО РЕЧИ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

***Д.В. Краснопрошин***
*аспирант каф. электронных вычислительных средств, БГУИР*
daniil.krasnoproshin@gmail.com

***М.И. Вашкевич***
*д-р техн. наук, доцент, проф. каф. электронных вычислительных средств, БГУИР*
vashkevich@bsuir.by

***Д.В. Краснопрошин***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем.*

***М.И. Вашкевич***
*Окончил Белорусский государственный университет информатики и радиоэлектроники (БГУИР) в 2008 г., в 2013 г. защитил кандидатскую диссертацию, а в 2022 г. докторскую диссертацию. С 2023 г. работает профессором кафедры электронных вычислительных средств БГУИР. Проводит научные исследования, связанные с применением методов машинного обучения для анализа и классификации речевых сигналов.*

**Аннотация.** В статье представлен обзор методов и алгоритмов распознавания эмоций в речи, охватывающий этапы предобработки данных, извлечения признаков и выбора моделей классификации. Рассмотрены преимущества и недостатки применения статистических моделей. Особое внимание уделено методам с использованием нейронных сетей, анализу их преимуществ и недостатков в контексте распознавания эмоций. Оценены перспективы дальнейших исследований, направленных на улучшение эффективности и интерпретируемости моделей, включая использование мультимодальных данных и сокращение числа обучаемых параметров моделей для повышения производительности.

**Ключевые слова:** глубокое обучение, распознавание эмоций в речи, сверточные нейронные сети, рекуррентные нейронные сети.

**Введение.** Распознавание эмоционального состояния по речевому сигналу является важной проблемой, вызывающей все больший интерес у научного сообщества. Это связано с многочисленными прикладными аспектами данного направления, такими как диалоговые системы и помощники, электронное обучение, клинические исследования, распознавание лжи, сфера развлечений, компьютерные игры и колл-центры [1, 2]. Создание эффективных моделей распознавания эмоций в речи может значительно улучшить пользовательский опыт в системах, включающих взаимодействие человека и машины, например, в областях искусственного интеллекта (ИИ) или мобильного здравоохранения (mHealth) [3].

В настоящее время уже предложены различные системы для распознавания в этих областях. Исследователи используют различные методы и классификаторы для определения эмоций, например, метод на основе линейного дискриминантного анализа (ЛДА), метод опорных векторов (МОВ), а также различные архитектуры искусственных нейронных сетей [3].

Тем не менее, данная задача по-прежнему остается весьма сложной, что связано с необходимостью выбора подходящего признакового пространства, наличием фонового шума в аудиозаписях, наличием большого количества естественных языков, а также индивидуальными особенностями говорящих [2].

В данной работе произведен обзор методов обработки речевых сигналов и моделей машинного обучения, которые используются в настоящее время для построения систем распознавания эмоций по речи. Также выполнен обзор наиболее распространенных речевых баз, использующихся для оценки производительности подобных систем. В конце работы сделаны выводы относительно текущего состояния проблемы распознавания эмоций, а также обозначены возможные направления для дальнейших исследований.

**Речевые базы данных для задач распознавания эмоций.** Исходные наборы данных играют ключевую роль в построении любой системы машинного обучения, в т.ч. систем распознавания эмоций в речи. В настоящее время речевые базы, используемые для обучения, могут содержать как реальные примеры проявления эмоций, так и специально подготовленные записи. Очевидно, что чем более исходные данные ближе к "естественным", тем они сложнее для анализа. Среди наиболее популярных речевых баз данных для распознавания эмоций являются следующие:

1 **Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS):** Данный набор включает записи от 24 актеров (12 мужчин и 12 женщин), представленных по 104 высказывания на каждого актера (60 речевых и 44 песенных). Набор состоит из 1440 аудиофайлов в wav-формате (16 бит, 48 кГц). RAVDESS содержит в себе различные эмоциональные состояния, такие как нейтральность, спокойствие, счастье, грусть, гнев, страх, удивление и отвращение [4].

2 **Toronto emotional speech set (TESS):** состоит из 2800 аудиозаписей, созданных двумя актрисами в возрасте 26 и 64 лет, которые выражают шесть основных эмоций — гнев, отвращение, страх, счастье, печаль и удивление — вместе с нейтральным эмоциональным состоянием. Каждая запись содержит одно из 200 целевых слов, встроенных в «несущую» фразу «Произнесите слово [цель]», что обеспечивает единообразие в структуре предложения. Аудиофайлы хранятся в wav-формате (16 бит, 24 кГц) [5].

3 **Berlin Emotional Speech Database (Emo-DB):** данный набор включает записи от 10 актеров (5 мужчин и 5 женщин), представленных по 535 высказывания на каждого актера. В базе данных представлены различные эмоции, такие как радость, гнев, грусть, удивление, отвращение, страх, нейтральность и другие [6].

4 **Danish Emotional Speech Database (DES)**: датасет включает записи от 10 актеров (5 мужчин и 5 женщин), представленных по 35 высказываний на каждого актера. В базе данных представлены различные эмоции, такие как радость, гнев, грусть, удивление, страх и нейтральное состояние [7].

5 **Russian Emotional Speech Dialogs with annotated text (RESD):** русскоязычный набор данных эмоциональных речевых диалогов. Этот набор данных был собран из ~3,5 часов живой речи актеров, которые озвучивали заранее распределенные эмоции в диалоге в течение ~3 минут каждый. Эмоции представлены в 7 состояниях: гнев, отвращение, страх, энтузиазм, счастье, нейтральность и грусть [8].

6 **Interactive Emotional Dyadic Motion Capture (IEMOCAP):** мультимодальный набор данных, который состоит из аудио, видео и образцов захвата движения лица, собранных у пяти пар актеров-мужчин и актеров-женщин. Образцы распределены по пяти сессиям, каждая из которых содержит данные от определенной пары. Актеры играли, используя театральные сценарии или импровизировали. Аудиофайлы из данного набора данных разделены на десять классов эмоций: злость, счастье, грусть, нейтральность, разочарование, возбужденность, страх, удивление, отвращение и «другое» [9].

Стоит отметить, что это лишь небольшая часть существующих наборов данных, используемых в данной области. Однако они пользуются наибольшей популярностью и

являются своего рода эталонными при оценке производительности систем распознавания эмоциональных состояний в речи [10].

**Распознавание эмоций с использованием классических методов машинного обучения.** Для систем распознавания эмоций в речи исходными данными являются аудиозаписи речи. Когда эти записи получены в реальных условиях, необходимо выполнить этап удаления шумов, чтобы улучшить качество данных. В случае с аудиозаписями, полученными в студийных условиях, данный этап можно пропустить. Далее, из обработанных аудиофайлов извлекаются признаки, которые могут быть интерпретированы моделями машинного обучения. Извлечённый набор признаков подается на вход классификатора, в результате чего на выходе модели генерируется метка, соответствующая наиболее вероятной эмоции, определённой на основе анализируемых данных. На рисунке 1 представлен пример описанной выше системы.

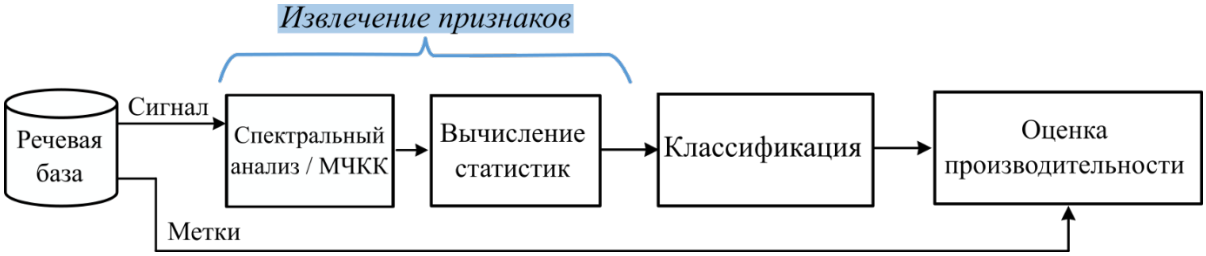



*Рисунок 1.* Типовая схема системы распознавания эмоций в речи

Большинство современных систем используют мел-частотные кепстральные коэффициенты (МЧКК) в качестве базовых речевых признаков [11]. На основе полученных МЧКК могут также вычисляться первые и вторые производные, а также различные статистики, такие как межквантильный размах, коэффициент асимметрии и эксцесс [Краснопрошин-12]. Вычисление данных характеристик позволяет более точно описать быстроту изменения спектральных характеристик звука во времени. Подобный подход позволяет создать вектор признаков фиксированной длины вне зависимости от длины входного сигнала, что является критическим как для классических моделей машинного обучения (МОВ, ЛДА), так и для различных типов нейронных сетей.

В рамках данного подхода также могут применяться алгоритмы отбора наиболее значимых признаков с целью повысить эффективность классификатора и снизить размерность признакового пространства. Важно подчеркнуть, что, как правило, большинство исследователей используют одни и те же характеристики, но в разном сочетании.

После извлечения и отбора признаков, следующим шагом является выбор подходящего классификатора, который вносит значительный вклад в точность распознавания эмоций. Классические модели такие как МОВ и ЛДА все еще остаются популярными в силу их простоты, неприхотливости к вычислительным ресурсам и хорошей интерпретируемости.

Классические методы классификации речевых эмоций требуют сложной предварительной обработки данных, включая извлечение и отбор признаков. Малые изменения в этих признаках могут значительно повлиять на результаты. Кроме того, увеличение объема данных снижает производительность этих методов.

**Применение сверточных нейронных сетей для распознавания эмоций.** С развитием методов глубокого обучения, на первый план стали выходить модели на основе различных архитектур нейронных сетей.

Одним из первых типов нейронных сетей, завоевавших популярность среди ученых в области распознавания эмоций в речи, являются различные вариации сверточных нейронных сетей (СНС). Так, например, в работе [13] была предпринята попытка исследовать эффективность СНС в сравнении с иными методами обработки речи, такими

как рекуррентные нейронные сети (РНС) и классические машинные алгоритмы. Из речевого сигнала извлекались пять различных типов признаков такие как МЧКК, мел-спектрограммы, хромаграммы и проч. Затем полученные признаки усреднялись по оси времени и конкатенировались в массив фиксированной длины (*193 признака*), который затем подавался на вход одномерной СНС в качестве входных данных. В работе подчеркивалось, что порядок добавления признаков в итоговый входной вектор имеет критические значение, и что при малейших изменениях результат работы классификатора может поменяться. Выдвигалась гипотеза о том, что смешивание этих признаков во входных данных обеспечивает более разнообразное представление звукового файла, что может привести к лучшему обобщению и лучшим результатам классификации. В качестве финального классификатора был предложен ансамбль моделей на основе СНС. Схематически его работу можно представить следующим образом:

**Алгоритм.** Ансамбль СНС для классификации эмоции в речевом сигнале.

**Вход:**

$X$ – исходный вектор признаков;

**Результат: метка класса**– эмоция, предсказанная классификатором;

**Begin:**

**1:** $P(\text{отвращение}) = classifier_disgust(X)$

2: **if** $(P(\text{отвращение}) \geq threshold)$:

3: return «отвращение»

4: **end if**

**5:** $P(\text{скука}) = classifier_boredom(X)$

6: **if** $(P(\text{скука}) \geq threshold)$:

7: return «Скука»

8: **end if**

3: $predicted_emotion$ = argmax(softmax($multi_class_emotion_classifier(X)$))

4: **return** $predicted_emotion$

**End**

Примером еще одной работы по распознанию эмоций с помощью СНС может послужить [14]. В ней предлагается выполнять распознавание на основе кватернионных СНС (*QCNN – Quaternion Convolutional Neural Networks*). В отличие от стандартных вещественных СНС, кватернионные СНС (КСНС) используют многомерные представления данных, что позволяет эффективно кодировать зависимые компоненты акустических признаков.

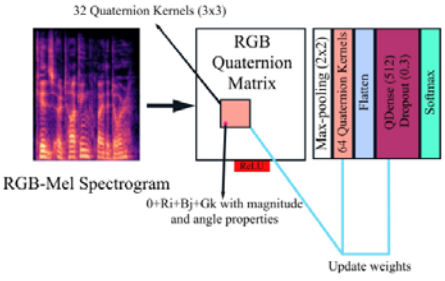



*Рисунок 2.* Архитектура кватернионной СНС (изображения взято из [14])

В [14] входные данные представлялись в виде многоканальных акустических признаков (например, МЧКК, мел-энергетические спектрограммы), которые обрабатывались в кватернионном пространстве. КСНС позволяют учитывать корреляции между различными акустическими параметрами, что приводит к более информативным представлениям признаков. Архитектура сети включает кватернионные свертки, которые снижают избыточность и улучшают способность модели к обобщению на новых данных (рисунок 2). В [14] экспериментально доказано, что КСНС демонстрирует улучшенную точность распознавания эмоций по сравнению с обычными СНС благодаря более

эффективному представлению взаимосвязанных акустических признаков. При этом удается снизить количество параметров модели, что уменьшает требования к вычислительным ресурсам и снижает вероятность переобучения (оттуда и способность хорошо обобщаться на новые данные). Более того, в ходе экспериментов было установлено, что КСНС могут быть эффективно использованы в условиях ограниченного объема данных, поскольку лучше моделируют зависимости между признаками.

**Применение рекуррентных нейронных сетей для распознавания эмоций.** Так как речевой сигнал представляет собой последовательность, не менее широкое распространение в задачах распознавания эмоций получили различные модификации РНС. Формально взаимосвязь речевого сигнала и РНС можно описать следующим образом. Речевой сигнал $s(t)$ представляет собой непрерывную функцию времени. В процессе обработки он дискретизируется с частотой $f_s$, что даёт последовательность дискретных отсчётов:

$$s[n] = s(nT_s),\ T_s = \frac{1}{f_s}.$$

В контексте задачи распознавания эмоций на вход РНС обычно подаются временные последовательности акустических признаков, характеризующих речевой сигнал. К таким признакам относятся МЧКК, хроматические признаки, формантные частоты, а также просодические параметры, включая частоту основного тона, интенсивность и прочее. Эти признаки, извлекаемые из фрагментов речевого сигнала, позволяют модели учитывать временные зависимости и динамические изменения в речи, что критически важно для корректной классификации эмоционального состояния говорящего.

Так, с помощью оконного преобразования Фурье и вычисления акустических признаков (например, МЧКК), речевой сигнал преобразуется в последовательность векторов:

$$X = [x_0, x_1, \ldots, x_T], \qquad x_t \in \mathbb{R}^d, \tag{1}$$

где сигнал $T$ – длина последовательности, а $d$ – размерность признакового пространства.

РНС моделируют временные зависимости в последовательности $X$ с помощью скрытых состояний $h_t$, которые передаются между слоями нейронной сети:

$$h_t = f(W_h h_{t-1} + W_x x_t + b_h) \tag{2}$$

где $h_t \in \mathbb{R}^m$ – скрытое состояние на шаге $t$, $W_h \in \mathbb{R}^{m \times m}$ и $W_x \in \mathbb{R}^{m \times d}$ – матрицы весов, $b_h \in \mathbb{R}^m$ – вектор смещения, $f(\cdot)$ – функция активации.

Сеть обучается так, чтобы скрытые состояния $h_t$ содержали информацию о предыдущих входных значениях, позволяя моделировать контекст в речевом сигнале.

На основе скрытых состояний РНС формируются выходные предсказания, например, вероятность принадлежности речевого фрагмента к определённому эмоциональному классу:

$$y_t = g(V_h h_t + b_y) \tag{3}$$

где $V_h$ – матрица весов, $g(\cdot)$ – функция активации, $b_y$ – вектор смещения.

Схематическое представление РНС показано на рисунке 3.

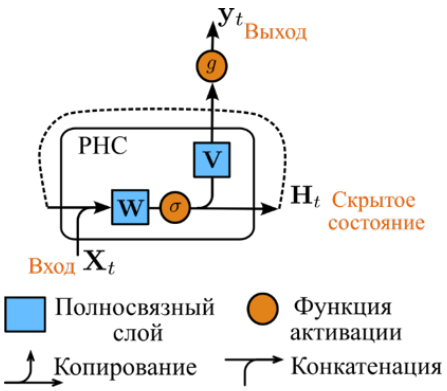



*Рисунок 3.* Архитектура простой РНС

Базовой моделью для решения задачи классификации эмоций в речи может послужить простая однонаправленная РНС. Выходное состояние такой модели на последнем шаге $h_T$ могут использоваться в качестве входа для функции активации $\mathrm{softmax}(\cdot)$, то есть классификация выполняется по вектору $h_T$, полученному после обработки всей последовательности входных признаков $X = \{x_0, x_1, \dots, x_T\}$. Формально, вероятность принадлежности к классу $c$ вычисляется как:

$$P(y = c \mid X) = \mathrm{softmax}(V_h h_t + b_y) \quad (5)$$

Таким образом, вся временная информация сжимается в единый вектор $h_T$, позволяя сети учитывать контекстную информацию. Однако данный подход обладает и рядом ограничений, связанных с неравномерным распределением эмоциональной информации в речевом сигнале. Например, если в начале последовательности присутствует ярко выраженная эмоция (например, гнев или радость), а к концу высказывания речь становится нейтральной, то скрытое состояние $h_T$ будет преимущественно отражать последние, менее эмоциональные фреймы. Это приводит к снижению качества классификации. В дополнение к этому, фреймы речевого сигнала, содержащие тишину или нейтральные интонации, могут затруднять обучение модели, так как финальное скрытое состояние будет неинформативным.

РНС имеет недостаток в виде проблем «исчезающего градиента». Это происходит, когда сеть прекращает обучение в результате изменения выходных данных по отношению к изменению входных данных. В связи с этим начали появляться усовершенствованные модели, такие как сети долгой краткосрочной памяти (*LSTM – Long short-term memory*), а также сети, именуемые управляемым рекуррентным блоком (*GRU – Gated recurrent unit*). Именно эти две усовершенствованные модели стали новым витком в развитии систем распознавания эмоций.

Существует два базовых подхода для построения классификаторов эмоций в речи на основе двунаправленных РНС [15]. Первый и наиболее «наивный» подход заключается в том, чтобы использовать общую эмоцию для каждого фрейма (рисунок 4) и, таким образом, обучать РНС на каждом фрейме и затем с помощью метода обратного распространения ошибки, оптимизируя веса после каждой итерации (в обе стороны). Проблема заключается в том, что не все фреймы в рамках одного высказывания выражают одну и ту же эмоцию. Связано это с тем, что некоторые фреймы могут являться тишиной (молчанием) или просто не содержать никакой эмоции [15]. Более того, поскольку мы предполагаем, что выходные данные РНС являются долгосрочными агрегациями входного сигнала, мы не должны ожидать, что выходные данные будут иметь желаемое долгосрочное представление, начиная с первого фрейма. Наоборот, РНС необходимо получить доступ к истории входного контекста, прежде чем она сможет правильно классифицировать последовательность целиком.

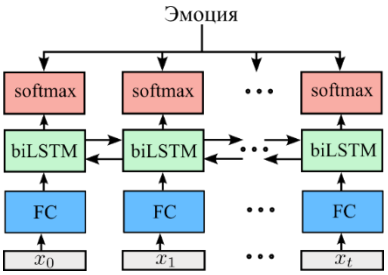



*Рисунок 4.* Схема с использованием одной метки (эмоции) для каждого фрейма

Альтернативой описанному выше методу является идея использовать только последние скрытые представления (каждого из двух направлений РНС) в качестве входа для функции активации softmax [15] (рисунок 5). Данный способ гарантирует, что модель сможет получить достаточно контекстуальной информации, прежде чем выполнить классификацию. Однако проблема, связанная с тем, что некоторые фреймы могут являться тишиной (молчанием) или просто не содержать никакой эмоции, все еще остается нерешенной. Например, если предложение начинается с ярко выраженной эмоции счастья, но эмоция угасает к концу, выход РНС начнет отклоняться от желаемого представления счастья, поскольку он столкнется с «неэмоциональными» фреймами к концу высказывания. Следовательно, решение полагаться лишь на последнее скрытое состояние РНС не является оптимальным [15].

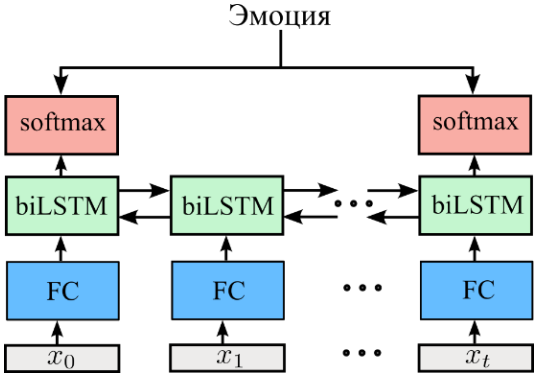



*Рисунок 5.* Схема с использованием только последних скрытых представлений (каждого из двух направлений РНС) в качестве входа для функции активации

Хорошим примером применения LSTM для задачи распознавания эмоций в речи служит [15], где предлагается новый метод автоматического распознавания эмоций в речи на основе двунаправленной LSTM (*biLSTM – bidirectional LSTM*) с механизмом локального внимания. Механизм внимания позволяет улучшить точность классификации за счет избирательного фокусирования на наиболее информативных участках речевого сигнала, что помогает модели игнорировать менее значимые фрагменты. В качестве входных признаков используются высота тона, вероятность звучания голоса, энергия, скорость пересечения нуля, МЧКК, среднее и дисперсия от МЧКК и др. [15]. Локальный механизм внимания, который позволяет модели фокусироваться только на важных сегментах речевого сигнала, позволяет избежать чрезмерного сглаживания информации, характерного для глобального внимания. Более того, была произведена конкатенация извлеченных признаков и скрытых состояний РНС для улучшенной финальной классификации эмоций. Предложенный подход позволил лучше выделить релевантные сегменты сигнала, а также снизить влияние фонового шума. Тем самым, экспериментально доказано, что локальное (селективное) внимание может применяться для подобных задач. Также в работе [15] отмечено, что локальное внимание способно снижать избыточность обработки, позволяя модели работать более эффективно и интерпретируемо.

Интересный подход был предложен в работе [16], который представляет собой комбинацию CHC и РНС (рисунок 6).

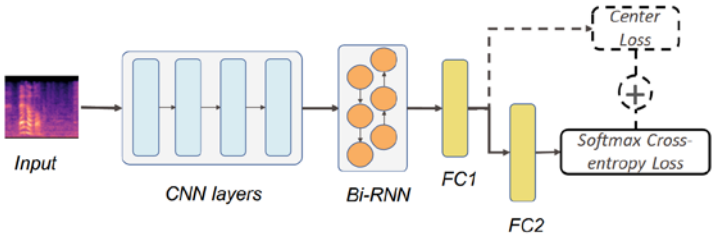



*Рисунок 6.* Система на основе комбинированного подхода (CHC + РНС) с двумя функциями потерь (изображения взято из [*16*])

В работе решается проблема внутриклассовой изменчивости, возникающей в связи с использованием МЧКК и коэффициентов линейного предсказания и приводящей к снижению эффективности моделей. В качестве альтернативы в статье предлагается использовать мел-спектрограммы, которые позволяют сохранять временные и частотные характеристики речи, в качестве входных данных, которые затем подаются на вход CHC. Наряду с этим вводится дополнительная функция потерь именуемая CenterLoss. Цель данной функции – снизить внутриклассовую изменчивость признаков и тем самым повысить разделимость признакового пространства. В [16] утверждается, что стандартная кросс-энтропийная функция потерь фокусируется на правильной классификации образцов, но не всегда оптимизирует представление данных в признаковом пространстве. Это приводит к высокой внутриклассовой вариативности. Функция CenterLoss решает эту проблему, «притягивая» представления одного класса ближе к его центру в признаковом пространстве. Таким образом, признаки описывающий один и тот же класс становятся более компактными. В итоге, предлагается комбинированный подход, основанный на комбинировании стандартной кросс-энтропийной функция потерь с CenterLoss. Было экспериментально доказано, что данная комбинация усиливает дискриминативную способность модели, что в свою очередь позволяет ей выучить более эффективное признаковое представление входных сигналов.

Итоговая система представляет собой следующее: слои CHC извлекают пространственную информацию из спектрограммы переменной длины, чтобы получить последовательность переменной длины. Двунаправленная РНС (в основе лежит GRU) сжимает последовательность переменной длины до вектора фиксированной длины. Затем первый полносвязный нейронный слой проецирует выходные данные РНС на желаемую размерность. После этого второй полносвязный нейронный слой, выход которого обозначает апостериорные вероятности классов, используется для вычисления softmax энтропийной функции потерь. Эта функция позволяет сети обучаться и выделять разделяющие признаки, а CenterLoss одновременно снижает внутриклассовую вариацию признаков.

В результате экспериментов в [16] делается вывод, что добавление CenterLoss значительно повышает точность классификации за счет уменьшения внутриклассовой изменчивости. Более того, предлагаемый показывает хорошие результаты в условиях ограниченного количества данных, поскольку он помогает извлекать более стабильные признаки. При этом отмечается, что спектрограммы оказываются более информативными, чем МЧКК, так как они сохраняют больше временных и частотных деталей.

**Оценка производительности.** Оценка производительности моделей классификации эмоций в речи является ключевым этапом в разработке систем автоматического анализа эмоционального состояния и во многом определяет практическую применимость предлагаемых алгоритмов. Для объективного анализа качества работы модели используются различные количественные метрики. Существуют также различные методы разделения данных на тренировочные и тестовые выборки.

Для количественной оценки качества классификации эмоций в речи наиболее часто применяются следующие метрики:

1 **Правильность** (**Accuracy**) – является базовой метрикой, определяющей долю правильно классифицированных примеров относительно общего числа примеров. Однако в задачах с несбалансированными классами данная метрика может быть недостаточно информативной.

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (6)$$

где $TP$ – доля истинных положительных результатов (*TP – true positive*), *TN* – доля истинных отрицательных результатов (*TN – true negative*), *FP* – доля ложных положительных результатов (*FP – false positive*), *FN* – доля ложных отрицательных результатов (*FN – false negative*).

Данный параметр применялся в [12, 14] при оценке классификаторов на таких наборах данных как RAVDESS, EMO-DB и IEMOCAP. Значение правильности находится в диапазоне от 0 до 1.

2 **Среднее арифметическое (невзвешенное) полноты (unweighted average recall, UAR).** UAR – это показатель, используемый для измерения общей производительности модели многоклассовой классификации, вычисляет средний уровень запоминания по всем классам, придавая каждому классу одинаковую важность без учета классового дисбаланса:

$$UAR = \frac{1}{N_c}\sum_{i=1}^{N_c}\frac{A_{ii}}{\sum_{j=1}^{N_c}A_{ij}} \quad (7)$$

где $A$ – матрица ошибок (confusion matrix); $N_c$ – количество классов.

Значение UAR находится в диапазоне от 0 до 1. Данная метрика использовалась, например, в [12, 17] при оценке классификаторов на основе МОВ и СНС.

Метод разбиения исходного набора данных на тренировочную и тестовую выборки, наряду с выбором метрики качества классификации, может оказать существенное влияние на оценку результатов работы системы распознавания. В связи с этим возникает задача тщательно продумать процесс разделения данных на тренировочные и тестовые выборки. Наиболее распространённым подходом является случайное разбиение данных, например, в пропорции 80% на обучение и 20% на тест, как это было сделано в [12, 15].

Тем не менее, в задачах классификации эмоций в речи необходимо учитывать дикторо-зависимые или дикторо-независимые сценарии. Так в дикторо-независимых сценариях данные одного говорящего не должны присутствовать одновременно в тренировочной и тестовой выборках, чтобы исключить эффект переобучения на индивидуальные особенности голоса. В [12] был предложен подход, призванный устранить данную проблему. В качестве исходного набора данных использовался RAVDESS. Суть подхода заключается в следующем:

1 Набор разбивается на *k* блоков.

2 В цикле для *i* = 1, 2, …, *k* выполняются следующие операции:

– блок *i* устанавливается в качестве тестового набора данных (test data);

– оставшиеся блоки принимаются как тренировочные данные (train data);

– выполняется обучение модели классификатора на тренировочных и оценивается ее эффективность на тестовых данных;

– сохранение результатов классификации для данных из тестового набора;

– сброс параметров модели до исходного состояния для следующей итерации.

3 Расчет оценки эффективности модели на основе сохраненных результатов классификации тестовых данных.

Выбранная стратегия заключается в том, что каждый блок должен содержать одинаковое количество случайно выбранных образцов для каждого класса. При этом должно выполняться условие, что каждый актер представлен либо в обучающей, либо в тестовой выборке, но не в обеих.

Еще одной стандартной техникой для получения более стабильной оценки производительности модели является кросс-валидация (cross-validation) или перекрестная проверка. В задачах классификации эмоций в речи чаще всего применяется k-fold кросс-валидация, при которой данные делятся $k$ равных частей (блоков). На каждом шаге одна из частей используется для тестирования, а оставшиеся — для обучения. Спикер-независимая кросс-валидация требует, чтобы данные одного говорящего полностью присутствовали только в одной из частей, что позволяет оценить обобщающую способность модели на новых говорящих. Например, исходные данные (RAVDESS), согласно схеме, предложенной в [17], разбивали на блоки следующим образом (в скобках указаны номера актеров):

– блок 0: (2, 5, 14, 15, 16);
– блок 1: (3, 6, 7, 13, 18);
– блок 2: (10, 11, 12, 19, 20);
– блок 3: (8, 17, 21, 23, 24);
– блок 4: (1, 4, 9, 22).

Схематическое представление кросс-валидации показано на рисунке 7.

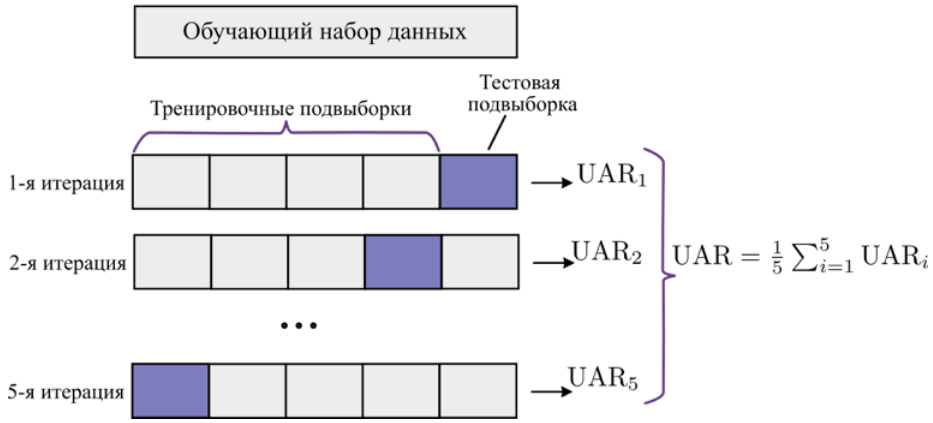



*Рисунок 7.* Пример кросс-валидации

**Заключение.** В работе представлен обзор современных систем распознавания эмоций в речи, включая ключевые этапы, такие как предобработка аудиоданных, извлечение признаков и выбор модели классификации. Особое внимание уделено различным архитектурам нейронных сетей, активно применяемым для решения данной задачи. Несмотря на достигнутые успехи, существующие методы, включая традиционные алгоритмы, имеют значительные ограничения, связанные с необходимостью сложной предварительной обработки данных, высокой чувствительностью к изменениям в признаках и снижением производительности при увеличении объема данных. Рассмотренные подходы подчеркивают важность дальнейших исследований для создания более эффективных, интерпретируемых и универсальных моделей. Перспективными направлениями являются разработка устойчивых и масштабируемых методов, таких как нейронные сети, а также интеграция мультимодальных данных для повышения точности распознавания эмоций. Кроме того, важным направлением для будущих исследований является сокращение числа параметров моделей, что позволит улучшить их производительность и снизить вычислительные затраты.

## Список литературы

[1] ДИТ Аналитика [Электронный ресурс] / Применение передовых технологий в работе контакт-центра. – URL: https://moscowanalytics.ru/in- dex/research/call_center (дата обращения 12.11.2024).

[2] Han K., Yu D., Tashev I. Speech emotion recognition using deep neural network and extreme learning machine //Proceedings of Interspeech-2014. – 2014. – P. 223-227.

[3] Badshah A. M. et al. Speech emotion recognition from spectrograms with deep convolutional neural network //2017 International conference on platform technology and service (PlatCon). – 2017. – P. 1-5.

[4] Livingstone S. R., Russo F. A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English //PloS one. – 2018. – Т. 13. – №. 5. – P. 1-35.

[5] Guilford, J. P., & Schultheiss, H. H. (2008). The Toronto Emotional Speech Set (TESS) [Database]. University of Toronto. – URL: https://utoronto.scholaris.ca/collections/036db644-9790-4ed0-90cc-be1dfb8a4b66. (дата обращения 08.12.2025).

[6] Burkhardt F. et al. A database of German emotional speech // Proceedings of Interspeech-2005. – 2005. – Т. 5. – С. 1517-1520.

[7] Jensen, R., and A. Nielsen. The Danish Emotional Speech Database (DES). Aalborg University, 2010, – URL: https://vbn.aau.dk/en/publications/design-recording-and-verification-of-a-danish-emotional-speech-da. (дата обращения 08.12.2025).

[8] Amentes, Artem, Nikita Davidchuk, and Ilya Lubenets. Russian Emotional Speech Dialogs with Annotated Text (RESD), 2022. – URL https://paperswithcode.com/dataset/resd. (дата обращения 08.12.2025).

[9] Busso C. et al. IEMOCAP: Interactive emotional dyadic motion capture database // Language resources and evaluation. – 2008. – Vol. 42. – P. 335-359.

[10] Видман В. В. и др. Распознавание эмоций из речевого сигнала // Молодежь и современные информационные технологии: сборник трудов XVI Международной научно-практической конференции студентов, аспирантов и молодых учёных, 3-7 декабря 2018 г., г. Томск.—Томск, 2019. – 2019.

[11] Huang X. et al. Spoken Language Processing: A guide to theory, algorithm, and system development. – Prentice hall PTR, 2001.

[12] Краснопрошин Д. В., Вашкевич М. И. Метод распознавания эмоций в речевом сигнале с использованием машины опорных векторов и надсегментных акустических признаков //Доклады БГУИР. – 2024. – Т. 22. – №. 3. – С. 93-100.

[13] Issa D., Demirci M. F., Yazici A. Speech emotion recognition with deep convolutional neural networks //Biomedical Signal Processing and Control. – 2020. – Т. 59. – С. 101894.

[14] Muppidi A., Radfar M. Speech emotion recognition using quaternion convolutional neural networks // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – 2021. – P. 6309-6313.

[15] Mirsamadi S., Barsoum E., Zhang C. Automatic speech emotion recognition using recurrent neural networks with local attention // Proceedings of IEEE International conference on acoustics, speech and signal processing (ICASSP). – 2017. – P. 2227-2231.

[16] Dai D. et al. Learning discriminative features from spectrograms using center loss for speech emotion recognition // Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). –2019. – P. 7405-7409.

[17] Luna-Jiménez C. et al. Multimodal emotion recognition on RAVDESS dataset using transfer learning // Sensors. – 2021. – Т. 21. – №. 22. – P. 1-29.

### Авторский вклад

**Краснопрошин Даниил Вадимович** – решение задач исследования, анализ полученных результатов, формирование структуры статьи.

**Вашкевич Максим Иосифович** – постановка цели и задач исследования, руководство исследованием, анализ полученных результатов, редактирование статьи.

## ANALYSIS OF APPROACHES TO BUILDING SPEECH EMOTION RECOGNITION SYSTEMS USING DEEP LEARNING METHODS

***D.V. Krasnoproshin***
*PhD Student at the Department of Electronic Computing Facilities, Belarusian State University of Informatics and Radioelectronics*

***M.I. Vashkevich***
*Dr.Sc., Professor at the Department of Electronic Computing Facilities, Belarusian State University of Informatics and Radioelectronics*

**Abstract.** The article provides an overview of methods and algorithms for speech emotion recognition problems, covering the stages of data preprocessing, feature extraction, and classification model selection. The application of statistical models and their limitations are discussed. Special attention is given to neural network technologies, analyzing their advantages and disadvantages in the context of solving the problem. The prospects for further research aimed at improving the efficiency and interpretability of models are evaluated, including the use of multimodal data and dimensionality reduction to enhance performance.

**Keywords**: deep learning, speech emotion recognition, convolutional neural networks, recurrent neural networks.

УДК 378.1: 614.84

# АНАЛИЗ ПОДХОДОВ К ПОСТРОЕНИЮ АВТОМАТИЗИРОВАННЫХ СИСТЕМ ЧТЕНИЯ ПО ГУБАМ

***Д.А. Макар***
*Аспирант кафедры электронных вычислительных средств БГУИР*
my-makar@mail.com

***М.И. Вашкевич***
*д-р техн. наук, доцент, проф. каф. электронных вычислительных средств, БГУИР*
vashkevich@bsuir.by

***Д.А. Макар***

*Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методики распознания речи для людей с нарушением речевого аппарата.*

***М.И. Вашкевич***

*Окончил Белорусский государственный университет информатики и радиоэлектроники (БГУИР) в 2008 г., в 2013 г. защитил кандидатскую диссертацию, а в 2022 г. докторскую диссертацию. С 2023 г. работает профессором кафедры электронных вычислительных средств БГУИР. Проводит научные исследования, связанные с применением методов машинного обучения для анализа и классификации речевых сигналов.*

**Аннотация.** Автоматическое чтение речи по губам представляет собой перспективную технологию, позволяющую распознавать устную речь на основе анализа движения губ и мимики лица. С развитием методов глубокого обучения произошел значительный прогресс в этой области, что открывает новые возможности для помощи людям с нарушением слуха, приобретенной потерей речи, а также в других задачах, где необходимо распознавать речь в условиях агрессивного шумового воздействия. В работе рассматриваются структура автоматизированной системы чтения по губам, методы предобработки и извлечения признаков из видеоданных, а также перспективы развития данной технологии.

**Ключевые слова:** автоматическая система чтения по губам, глубокое обучение, сверточные нейронные сети, рекуррентные нейронные сети, распознавание речи

**Введение.** Автоматическое чтение речи по губам – технология, позволяющая распознать звучащую речь на основе визуальных данных, таких как движение губ и мимика лица. Чтение по губам трудноприобретаемый навык для человека, так люди с потерей слуха обычно способны не более 20 % произносимых слов. Поэтому важной задачей является автоматизации процесса чтения по губам. Машинное чтение по губам имеет огромный практический потенциал, поскольку может применяться в устройствах, помогающих людям с нарушением слуха [1], в системах распознавания речи в условиях агрессивного акустического загрязнения [2], для биометрической идентификации, а также для построения человеко-компьютерных интерфейсов для людей, потерявших голос.

**Статистические модели.** До революционных изменений, произошедших в связи с разработкой методов глубокого обучения, автоматические системы чтения по губам (АСЧГ) строились с использованием сложных методов предобработки видеокадров (фреймов), включающих извлечение признаков, нахождение характерных точек, контуров

губ [3], с последующей обработкой динамики, извлеченных характеристик. Таким образом, процедура чтения по губам производилась с помощью анализа изменения носогубных мышц лица. В качестве основных признаков обычно использовались внешний и внутренний контур губ (рисунок 1).

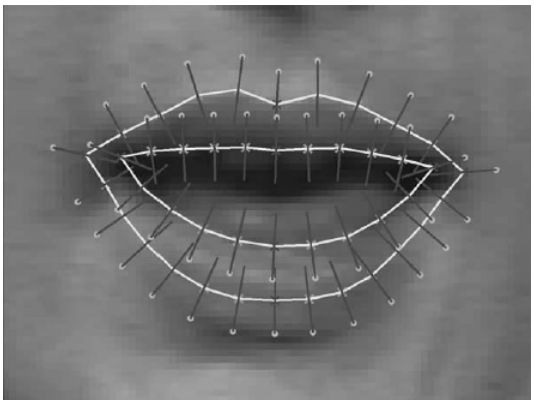


*Рисунок 1. Внутренний и внешний контуры модели губ. Белые точки представляют собой первичные и вторичные характерные точки. Линии обозначают нормали, проведенные через каждую точку. (Изображение взято из [4])*

Для отслеживания внутреннего и внешнего контура губ чаще всего осуществляется с использованием активных контурных моделей [4, 5]. Для распознавания визуальных фонем, которые также называют виземами, в ранних работах использовался математический аппарат скрытых марковских моделей (СММ). Следует заметить, что АСЧГ могут производить распознавание на уровне визем, на уровне слов, а также на уровне предложений [6]. В ранних работах [3, 4] описываются системы, позволяющие распознавать на видеозаписях отдельные виземы. Общий вид стуктуры АСЧГ, использующей временную последовательность признаков, полученных в результате анализа кадров видеоизображения показан на рисунке 2. Для непосредственного распознования фонем система использует скрытые марковские модели.

*Рисунок 2. Структура автоматической системы чтения по губам на основе СММ*

В частности, структура АСЧГ, показанная на рисунке 2, использовалась в работе [4], где рассматривалась задача распознавания изолированных фонем. В наилучшей конфигурации системы точность распознавания составляла 41,9%, что оставляло большой простор для возможных улучшений. Аналогичная система в работе [7] тестировалась на задаче распознавания цифр показала точность на уровне 37%. Важно отметить, что системы, основанные на извлечении построенных вручную признаков (англ. «*handcrafted features*») с последующей классификацией на основе статистических моделей, часто были дикторо-зависимыми, что существенно ограничивало их применение. Одним из наилучших результатов, достигнутых в рамках рассматриваемого статистического подхода, можно считать работу [8], где авторы использовали преобразованные при помощи дискретного косинусного преобразования области рта, а также комбинированную систему классификации на основе СММ и моделей гауссовых смесей (МГС). В результате была получена дикторозависимая точность распознавания 86,4%.

**Использование сверточных нейронных сетей в АСЧГ.** Подходы к построению АСЧГ значительно изменились с момента широкого распространения глубоких нейросетевых моделей. В первую очередь изменения коснулись процесса извлечения

визуальных признаков из области рта на видеоизображении. Вместо построенных вручную признаков стали использоваться сверточные нейронные сети (СНС). Впервые СНС для извлечения визуальных признаков в АСЧГ была использована в работе [9]. На рисунке 3 схематично представлена система из работы [9].

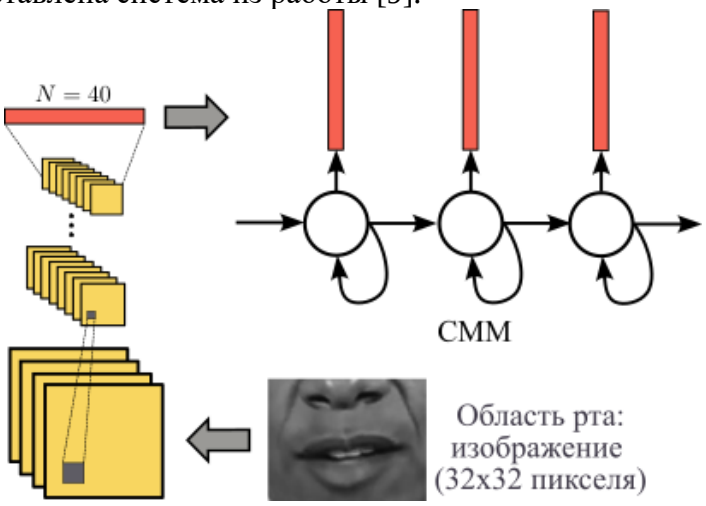



*Рисунок 3. Использование СНС автоматизированной системе чтения по губам*

Система [9] включала в себя СНС и СММ-МГС для выделения визуальных признаков и распознавания отдельных слов/фонем соответственно. Система выделения визуальных признаков использовала семислойную СНС, которая позволяла распознавать фонемы из последовательности изображений области рта. СНС моделировала нелинейное отображение исходных изображений в оттенках серого в соответствующее апостериорное распределение вероятности меток фонем. Временные последовательности, полученные из этих выходных данных, рассматриваются как визуальные характеристики для чтения по губам. Далее, полученные последовательности визуальных признаков, обрабатывались при помощи СММ-МГС для распознавания отдельных слов. В результате система показала 58% точность распознавания отдельных фонем, распознавание слов при этом находилась на уровне 37%.

**3D-сверточные нейронные сети для аудиовизуального распознания.** 3D-сверточные нейронные сети (3D-СНС) представляют собой инструмент для обработки видеоданных, где требуется учет как пространственной, так и временной информации. В контексте задачи чтения по губам, 3D-СНС используются для сопоставления аудио и визуальных потоков, что позволяет улучшить точность распознавания речи, особенно в условиях шума или отсутствия аудиоданных [10]. Особенностью 3D-СНС является их возможность эффективно захватывать динамику движений губ и мимики лица.

Основная задача аудиовизуального распознавания с акцентом на чтение губ заключается в соответствии между аудио и визуальным потоком. Архитектура 3D-СНС, которая обрабатывает обе модальности (аудио и видео) и объединяет их в общее пространство признаков для оценки их соответствия описана в [11]. Система, описанная в [11], состоит из двух нейронных сетей и позволяет эффективно использовать временную корреляцию между аудио и визуальными данными, что особенно важно для задач, таких как чтение по губам. Обе сети обучаются совместно, но имеют разные наборы весов. В качестве аудиоданных на вход сети подаются мел-частотные кепстральные коэффициенты (МЧКК). Входной тензор данных для аудио сети имеет размерность 15×40×3, где 15 – количество временных фреймов, 40 – количество МЧКК, а 3 – каналы (МЧКК, первая и вторая производные МЧКК). В качестве видеоданных на вход поступает последовательность кадров, на которых выделена область рта. Входной тензор данных для видео сети имеет размерность 9×60×100, где 9 – число кадров, 60×100 – размер изображения области рта в оттенках серого.

Аудио-часть сети [11] состоит из нескольких сверточных слоев, которые обрабатывают временные и спектральные характеристики. Каждый слой применяет свертки по временной и частотной осям, что позволяет сети извлекать как локальные, так и глобальные признаки. Видео-часть сети также состоит из нескольких сверточных слоев, которые обрабатывают пространственную и временную информацию. Свертки применяются по всем трем измерениям (время, высота, ширина), что позволяет учитывать динамику движения губ.

После обработки каждой модальности по отдельности, признаки объединяются на более поздних слоях сети. Это позволяет сети учитывать взаимосвязь между аудио и визуальными данными. Общая структура системы показана на рисунок 4.

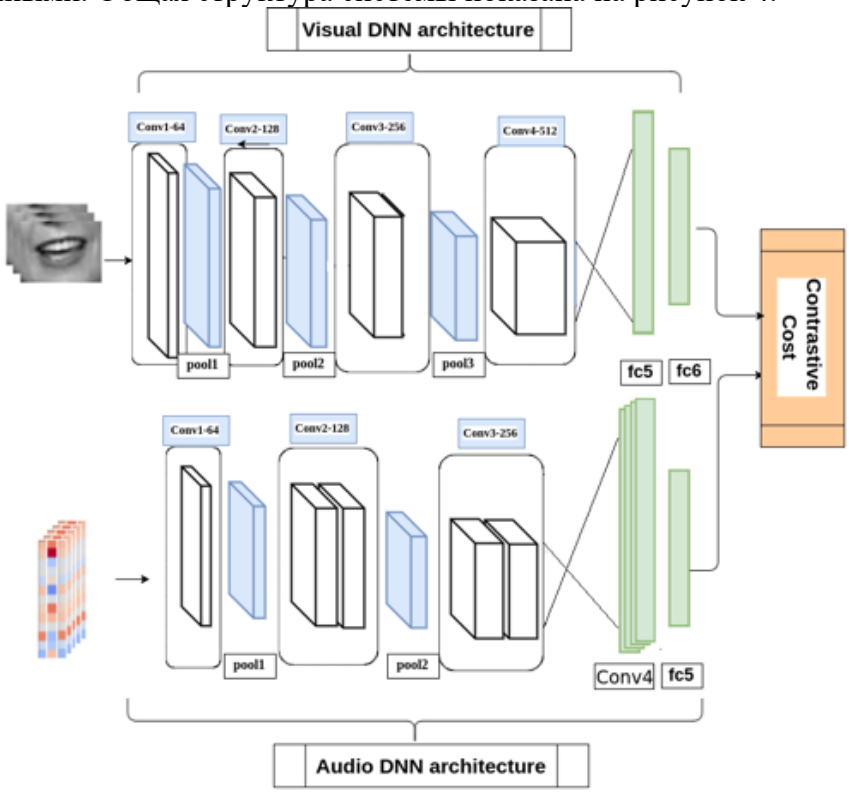



*Рисунок 4. Архитектура связанная с 3D сверточной нейронной сетью (изображение взято из [11])*

**Использование рекуррентных нейронных сетей и механизма внимания в АСЧГ**. Существенный прогресс в задаче визуального распознавания речи был достигнут в работе [6], в которой была предложена модель WLAS («Watch, Listen, Attend and Spell»). Модель WLAS предназначена для распознавания символов в произносимых предложениях по видеозаписи говорящего лица, как со звуком, так и без него. Модель состоит из трех ключевых компонент: 1) кодировщика изображения Watch; 2) кодировщика аудио Listen; 3) декодировщика символов Spell. Однако наиболее важным элементом модели являются блоки внимания, которые применяются отдельно для признаков, извлеченных из аудио- и из видеоданных. Общая структура модели WLAS приведена на рисунке 5. Отметим, что как кодировщики Watch и Listen, так и декодировщик Spell построены на базе рекуррентной нейронной сети (РНС) с долговременной краткосрочной памятью (англ. *LSTM – long short-term memory*). В качестве аудиовхода модель получает вектора МЧКК, а для выделения визуальных признаков, подаваемых на вход обработки видеоданных, используется шестислойная СНС.

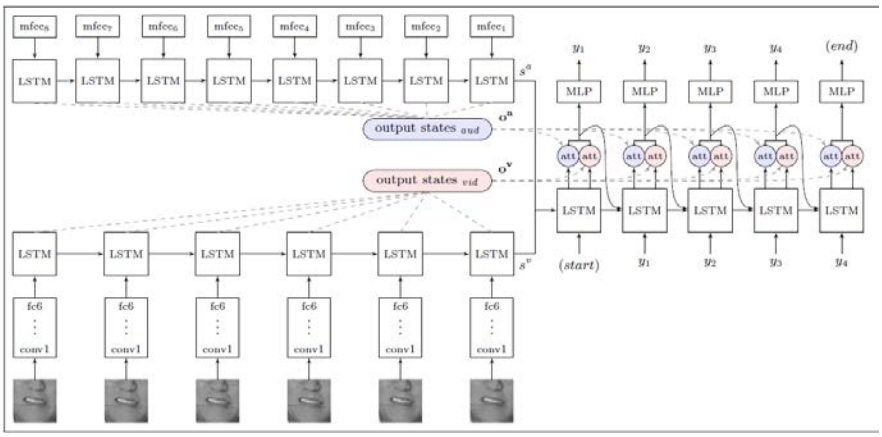



*Рисунок 5. Модель аудио-визуального распознавания речи (изображение взято из [6])*

По наблюдению авторов модели WLAS критически важной её частью модели является двойной механизм внимания [12], который может работать как с визуальными, так и с аудио данными. Без механизма внимания модель «забывает» входной сигнал и имеет очень низкую производительность.

В работе [6] также описан набор данных «Lip Reading Sentences» (LRS) для визуального распознания речи, содержащий более 100 тыс. естественных предложений из британских телепередач BBC, включающий разнообразные условия: различные позы, выражения, освещение, фоны и этническое происхождение говорящих (рисунок 6).

*Рисунок 6. Вверху: Оригинальные статические изображения из набора данных для чтения по губам. Внизу: Движения рта для слова «afternoon» от двух разных спикеров. Модель WLAS «видит» области внутри красных квадратов. (изображение взято из [6])*

*В работе [6] используется стратегия обучения «curriculum learning» (обучение начинается с коротких последовательностей, отдельных слов, затем постепенно переходит к полным предложениям), а также мультимодальное обучение (модель обучается на трех типах данных: только аудио, только видео, и аудио+видео, чтобы избежать доминирования одного из модальностей [23].*

Модель WLAS превосходит все предыдущие работы на стандартных бенчмарках для чтения по губам (LWR и GRID). На тестовом наборе LRS модель достигает: CER (Character Error Rate) – 7,9% (аудио+видео), 39,5% (только видео); WER (Word Error Rate) – 13,9% (аудио+видео), 50,2% (только видео); модель превосходит профессионального чтеца по губам на видео с BBC.

**Общая структура автоматизированных систем чтения по губам.** Обобщая рассмотренные работы, можно сделать вывод, что обобщенная структура АСЧГ состоит из четырех блоков:

- предобработка видео;
- извлечение визуальных признаков;
- моделирование временных зависимостей;
- классификация и генерация текста.

На этапе предобработки видео происходит обнаружение лица и губ, а также нормализация видео. Нормализация включает изменение размера, яркости и контрастности кадров для минимизации влияния изменений в освещении и ракурсе [2].

Для извлечения признаков из видео применяются СНС. Они позволяют выделить важные визуальные характеристики, такие как форма губ и движение рта. Для учета временной информации используются 3D-СНС, которые обрабатывают последовательность кадров.

Моделирование временных зависимостей движения губ применяются РНС. Эти модели способны запоминать информацию из предыдущих кадров и использовать ее для предсказания текущего состояния.

Классификация или генерация текста происходит на заключительном этапе, когда система либо классифицирует последовательности в определенные слова или фразы, либо генерирует текст, соответствующий произнесенным словам. Классификация и генерация текста – это два различных подхода к обработке и пониманию языка, которые могут использоваться в контексте распознавания речи.

Выбор между классификацией и генерацией текста зависит от конкретной задачи и требований систем. Классификация подходит для задач, где важно точно определить, к какому классу принадлежит входная информация, в то время как генерация используется в более сложных сценариях, где необходимо создавать осмысленный и контекстно правильный текст.

**Обзор баз данных для обучения моделей АСЧГ.** Для обучения и тестирования моделей АСЧГ используются базы данных, которые содержат видео с движением губ и соответствующие текстовые транскрипции, некоторые из них:

GRID Corpus. Содержит видео с произнесением 1000 предложений, каждое из которых состоит из 6 слоев. Используется для задач, связанных с распознаванием предложений. Разрешение видео: 320x288 пикселей. Количество данных: 34 часа видео [13].

LWR (Lip Reading in the Wild). Содержит более 500000 видеоклипов из телевизионных передач. Каждый клип длится около 1 секунды и соответствует одному слову. Разрешение видео: 224x224 пикселей [6].

TCD-TIMIT. База данных, созданная в Trinity College Dublin, включает видео с произнесением предложений из корпуса TIMIT. Разрешение видео: 1920x1080 пикселей. Количество данных 60 часов видео [14].

**Перспективы развития АСЧГ.**

В целом, внедрение нейронных сетей в АСЧГ позволило значительно повысить точность и надежность систем распознавания, а также расширить их применение в различных областях, таких как помощь людям с нарушением слуха или в условиях, где звуковая информация недоступна.

Однако задача автоматизированного чтения по губам остается сложной из-за высокой вариативности данных (разные люди, освещение, ракурс) и недостатки размеченных данных для обучения моделей.

Для дальнейшего развития АСЧГ необходимо решить несколько ключевых задач, включая создание более крупных и разнообразных баз данных, разработку более эффективных архитектур и интеграцию мультимодальных подходов. Одним из перспективных направления является использование предобученных моделей на больших наборах данных для улучшения точности распознавания на меньших наборах данных.

**Заключение.** АСЧГ остается сложной, особенно из-за высокой вариативности данных и ограниченного количества размеченных обучающих наборов. Современные модели на основе СНС и РНС позволяют достичь высокой точности распознавания. Таким образом, технологии АСЧГ обладают значительным потенциалом для практического применения. Перспективы их развития открывают новые горизонты в области человеко-

компьютерного взаимодействия и могут существенно изменить подход к распознаванию и пониманию речи в будущем.

## Список литературы

[1] Zhang, K., Zhang, Z. Joint Face Detection and Alignment Using Multi-Task Cascaded Convolutional Networks // IEEE Signal Processing Letters, 2016 – 5 с.

[2] O'Reilly, M., McMahon, P.A Survey of Recent Advancements in Lip Reading // International Journal of Computer Vision. – 2020 – 17 с.

[3] Zhao G., Barnard M., Pietikainen M. Lipreading with local spatiotemporal descriptors // IEEE Transactions on Multimedia. – 2009. – Т. 11. – №. 7. – 1254-1265 с.

[4] Matthews I. et al. Extraction of visual features for lipreading IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2002. – Т. 24. – №. 2. – 198-213 с.

[5] Ковшов Е. Е., Завистовская Т. А. Система обработки движения губ человека для речевого ввода информации // Cloud of science. – 2014. – Т. 1. – №. 2. – 279-291 с.

[6] Chung, J. S., Senior, A., Vinyals, O., Zisserman, A. Lip Reading Sentences in the Wild // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2017. – P. 6447-6456.

[7] Fu Y., Yan S., Huang T. S. Classification and feature extraction by simplexization // IEEE Transactions on Information Forensics and Security. – 2008. – Т. 3. – №. 1. – 91-100 с.

[8] Gergen S. et al. Dynamic Stream Weighting for Turbo-Decoding-Based Audiovisual ASR // Proceedings of Interspeech. – 2016. – P. 2135-2139.

[9] Noda K. et al. Lipreading using convolutional neural network // Proceedings of Interspeech. – 2014. — P. 1149-1153.

[10] Tran, D., Bourdev, L., & Fergus, R. Learning Spatiotemporal Features with 3D Convolutional Networks // Proceedings of IEEE International Conference on Computer Vision. – 2015. – P. 4489-4497.

[11] Torfi A. et al. 3D convolutional neural networks for cross audio-visual matching recognition // IEEE Access. – 2017. – Т. 5. – P. 22081-22091.

[12] Bahdanau, D., Cho, K., Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate // Processing of ICLR, 2015 – P. 1-15.

[13] Cooke, M. P., Lu, Y. The GRID Corpus for Speech Recognition // Proceedings of the International Conference on Spoken Language Processing, 2006 – P. 1-15 с.

[14] Harte N., Gillen E. TCD-TIMIT: An audio-visual corpus of continuous speech //IEEE Transactions on Multimedia. – 2015. – Т. 17. – №. 5. – P. 603-615.

## Авторский вклад

**Вашкевич Максим Иосифович** – постановка задачи исследования в области АСЧГ.

**Макар Дарья Александровна** – проведение анализа современных подходов к построению АСЧГ.

# ANALYSIS OF APPROACHES TO DESIGN AUTOMATED LIPE READING SYSTEMS

***D.A. Makar***
*Postgraduate student of the Department of Electronic Computing Facilities of BSUIR*

***M.I. Vashkevich***
*Professor of the Department of Electronic Computing Facilities of BSUIR, Doctor of Technical Sciences*

**Abstract.** Automatic lip reading (ALR) is a promising technology that enables the recognition of spoken language based on the analysis of lip movements and facial expressions. With the advancement of deep learning methods, significant progress has been made in this field, opening new opportunities to assist people with hearing impairments and in noisy environments. This paper discusses the structure of ALR, methods for preprocessing and feature extraction, as well as the future prospects of the technology.

**Keywords**: automatic lip reding, deep learning, convolutional neural networks, recurrent neural networks, speech recognition.

УДК 004.855.5

# ПРОБЛЕМАТИКА МАШИННОГО ОБУЧЕНИЯ: ВЛИЯНИЕ ГИПЕРПАРАМЕТРОВ, ПЕРЕОБУЧЕНИЕ И НЕДООБУЧЕНИЕ НЕЙРОННЫХ СЕТЕЙ

***А.А. Навроцкий***
*Заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент*
navrotsky@bsuir.by

***А.О. Шведко***
*Аспирант кафедры информационных технологий автоматизированных систем БГУИР, магистр*
kv.space@yandex.ru

***В.И. Сакович***
*Научный сотрудник, магистр*
kv.space@yandex.ru

***А.А. Навроцкий***
*Заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент.*

***А.О. Шведко***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Магистр.*

***В.И. Сакович***
*Научный сотрудник, магистр.*

**Аннотация.** В рамках настоящего исследования выполнен анализ влияния входных данных, параметров первоначальной настройки на процесс переобучения и недообучения нейронной сети. Это исследование имеет важное значение, поскольку первоначальная настройка модели обучения является ключевыми аспектом в процессе обучения моделей глубокого обучения.

В процессе анализа модели были обучены с использованием различных комбинаций количества нейронов с одинаковыми входными данными и алгоритмом оптимизации, что позволило получить информацию о том, как обучение влияет на тренировочную и валидационную потери.

**Ключевые слова:** TensorFlow, гиперпараметры, переобучение и недообучение, нейронные сети, машинное обучение.

**Введение**. Наиболее распространенными программными библиотеками для машинного обучения являются *TensorFlow* и *PyTorch*. *TensorFlow*, разработанный компанией *Google*, известен своей мощной экосистемой, включающей инструменты для развертывания моделей и их оптимизации. *PyTorch*, созданный *Facebook*, выделяется интуитивно понятным интерфейсом и гибкостью, что делает его предпочтительным выбором для исследователей. Обе библиотеки активно поддерживаются и развиваются, предоставляя пользователям доступ к последним достижениям в области машинного обучения, а их популярность в академическом и промышленном секторах подтверждает эффективность и надежность.

Однако, процесс обучения нейронных сетей сопряжен с рядом значительных трудностей, которые могут существенно повлиять на качество и надежность получаемых моделей. Эти трудности включают в себя переобучение, недообучение и грамотное использование гиперпараметров, а их преодоление является важной задачей для

исследователей, стремящихся улучшить производительность и обобщающую способность моделей [1].

**Влияние гиперпараметров на обучение модели**. Гиперпараметры играют ключевую роль в процессе обучения нейронных сетей, поскольку они определяют стратегию оптимизации и влияют на динамику обучения, производительность и способность модели к обобщению.

Конкретные гиперпараметры, такие как скорость обучения, оказывают значительное влияние на производительность модели. Например, эксперименты с *ImageNet* показали, что изменение скорости обучения на порядок может существенно повлиять на скорость сходимости модели. Слишком высокая скорость обучения может привести к нестабильности
и невозможности достижения оптимума, в то время как слишком низкая – к медленной сходимости.

Однако, такие параметры, как скорость оптимизации и тип используемого оптимизатора (например, *SGD, Adam, RMSprop*), также влияют на обучение. Каждый оптимизатор имеет свои особенности и может быть более или менее эффективным для разных задач и архитектур.

Правильная настройка скорости обучения и оптимизации является критически важным этапом в обучении нейронных сетей.

Для наглядности приведенных сведений необходимо построить несколько простых полносвязных сетей (т.е. многослойных перцептронов) со следующими параметрами:

Таблица 1. Параметры конфигурации простых полносвязных сетей

| Наименование модели | Модель-основание | Малая модель (недообучение) | Большая модель (переобучение) |
|---|---|---|---|
| Алгоритм оптимизатора | *Adam* | *Adam* | *Adam* |
| Размер гиперпараметра (*batch size*) | 512 | 512 | 512 |
| Количество эпох | 20 | 20 | 20 |
| Ректифицированный линейный юнит первого слоя (количество нейронов) | 16 | 4 | 300 |
| Ректифицированный линейный юнит первого слоя (количество нейронов) | 16 | 4 | 300 |
| Сигмоидная функция (количество нейронов) | 1 | 1 | 1 |

**Причины переобучения**. Переобучение является важной проблемой, с которой сталкиваются при машинном обучении. Оно возникает, когда модель демонстрирует отличные результаты на обучающем наборе данных, но её производительность на тестовых данных значительно ниже. Это указывает на неспособность модели к адаптации, т. е. к применению выявленных закономерностей на новых, ранее не обработанных данных. Таким образом, переобучение свидетельствует о чрезмерной адаптации модели к особенностям обучающего набора, включая информационный шум и случайные вариации.

Основными причинами переобучения являются чрезмерная сложность модели
и недостаточное количество обучающих данных. Когда модель имеет слишком большое количество параметров, она может легко запомнить обучающие данные, включая их шум, вместо того, чтобы выучить общие закономерности.

Для наглядного представления создадим график тренировочной и валидационной потери большой модели относительно модели-основания (рисунок 1), где синим цветом представлена модель-основание, а зеленым большая модель. Непрерывные линии показывают тренировочную потерю, а пунктирные линии показывают валидационную потерю(следует помнить, более низкое значение валидационной потери означает лучшее качество модели). Здесь большая сеть начинает переобучаться позже, чем модель-основание и ее производительность уменьшается намного медленней после начала переобучения.

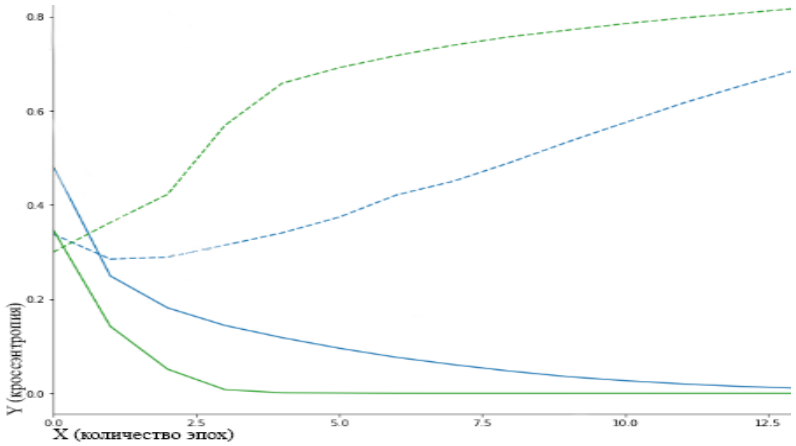



*Рисунок 1.* Пример тренировочных и валидационных потерь

В этом контексте переобучение можно определить, как ситуацию, «когда ошибка на данных обучения мала, а на данных уточнения значительно больше».

**Недообучение и причины возникновения**. Недообучение возникает в тех случаях, когда модель нейронной сети не может эффективно обучиться на предоставленных данных, что приводит к низкой производительности как на обучающем, так и на тестовом наборах. Основными причинами этого являются недостаточная сложность архитектуры модели, недостаточное количество эпох обучения и использование неподходящих методов оптимизации, которые не позволяет ей выявить все необходимые закономерности во входных данных.

Как и в случае переобучения модели, создадим график, показывающий недообучение меньшей модели относительно модели-основания (рисунок 2), где синим цветом представлена модель-основание, а оранжевым меньшая модель. Здесь меньшая сеть начинает обучение гораздо раньше, чем модель-основание, однако ее производительность оставляет желать лучшего из-за нехватки входных данных.

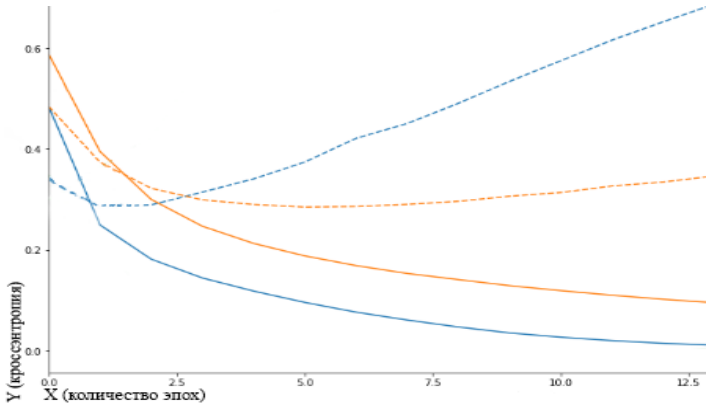



*Рисунок 2.* Пример тренировочных и валидационных потерь

Недообучение может привести к значительному снижению производительности модели, что выражается в высокой вероятности возникновения погрешности на всех этапах проверки и ограничивает возможность использования модели в реальных приложениях, так как она не способна адекватно анализировать новые входные данные.

**Заключение**. Широкий спектр задач, решаемых нейронными сетями, делает невозможным создание универсальных и мощных моделей, что, в свою очередь, требует разработки специализированных решений, функционирующих по различным алгоритмам.

Понимание и решение трудностей, связанных с обучением нейронных сетей, является важным этапом в развитии искусственного интеллекта. Перед началом обучения крайне важно провести тщательный анализ данных, который включает проверку их полноты, наличие пропусков и аномалий, а также оценку распределения. Правильная предобработка, включая нормализацию и масштабирование, может значительно улучшить качество модели. Например, пропуски в данных могут быть заполнены средними значениями или предсказаны
с использованием других признаков, что минимизирует их влияние на обучение.

## Список литературы

[1] Шакла Н. Машинное обучение и TensorFlow. СПб.: Питер, 2019. – 336 с.

## Авторский вклад

**Навроцкий Анатолий Александрович** – постановка задачи на исследование и руководство по оценке влияния гиперпараметров на процесс машинного обучения, а также влияние переобучения и недообучения на выходную модель данных.

**Шведко Александр Олегович** – исследование и оценка доступных материалов по машинному обучению, краткое описание проблематики по теме статьи, построение модели-основания и малой модели.

**Сакович Владимир Игоревич** – построение большой модели нейронной сети, оформление графического материала, написание заключения.

# PROBLEMS OF MACHINE LEARNING: IMPACT OF HYPERPARAMETERS, OVERFITTING AND UNDERFITTING OF NEURAL NETWORKS

***A.A. Navrotsky***
*Head of the Department of Information Technologies of Automated Systems BSUIR, PhD in Physics and Mathematics, Associate Professor*

***A.O. Shvedko***
*Postgraduate student of the Department of Information Technologies of Automated Systems of BSUIR, Master of Science*

***V.I. Sakovich***
*Researcher, Master of Science*

**Abstract:** This study analyzes the influence of input data and initial tuning parameters on the process of overfitting and underfitting of a neural network. This study is important because the initial tuning of the training model is a key aspect in the training process of deep learning models.

In the analysis, the models were trained using different combinations of the number of neurons with the same input data and optimization algorithm, which provided information on how training affects the training and validation losses.

**Keywords**: TensorFlow, hyperparameters, overfitting and underfitting, neural networks, machine learning.

UDC 330.33:004

# DIGITAL MATURITY CATEGORIES IN BUSINESS: INTEGRATING AND ASSESSMENT

***L.I. Arkhipova***
*Associate professor, PhD, BSUIR*
*Belarusian State University of Informatics and Radioelectronics,*
*Republic of Belarus,*
l.arkhipova@gmail.com

***L.I. Arkhipova***

*Graduated from the Belarusian State University and Academy of public administration under the aegis of the President of the Republic of Belarus. Has years of experience in Semiconductor industry (Integral, Minsk). Conduct scientific researches in the field of digital business transformation and marketing technologies.*

**Abstract.** In the context of rapid digital transformation, businesses need to align their organizational capabilities and expectations with technological advances in order to succeed. This study explores the interplay between organizational, customer, and technological digital maturity. The study emphasizes the challenges of integrating these different maturity levels, highlighting the importance of aligning technological investments with organizational readiness and customer expectations. By exploring this dynamic, this article offers insights on how to optimize digital transformation initiatives for long-term business growth.

**Key words:** organizational capabilities, digital organizational maturity (ODM), customer digital maturity (CDM), technology maturity (TM), data-driven decision-making.

**Introduction**. Successful digital transformation requires a strong approach that integrates organizational digital maturity, customer digital maturity, and the maturity of digital technologies. By understanding the interplay between these three categories, businesses can create strategies that enhance operational efficiency and meet the evolving expectations of customers. As organizations embark on their digital transformations, they need to continually assess and adapt strategies to align with technological advancements and market dynamics. Ultimately, the triad of digital maturity serves as the framework for businesses to thrive in the digital world.

This consideration is not just about having technology tools, but also about having the right attitude, processes, and structures to use technology effectively. Each company shows its level of digital usage and business success. Specific terminology exists to measure this: digital maturity and a digital maturity score.

*Digital maturity* refers to a company's ability to use digital technologies to achieve business objectives. Digital transformation entails identifying and solving challenges, creating opportunities, and decision-making in a digitally connected environment. The best tools for measuring digital maturity include those created by Deloitte, Gartner, Forrester, IBM, and Capgemini [1-3].

*Digital maturity score* is calculated by considering factors such as online presence, web analytics, marketing automation and technology in customer service. Businesses should conduct such an evaluation in order to:

- understand how the company uses digital technologies.
- identify areas for improvement;

– monitor progress over time;
– assess the impact of digital technology on business performance;
– identify risk factors and opportunities for growth.

Such an assessment helps companies understand their current digital state – to determine the organization's digital maturity level (Figure 1).

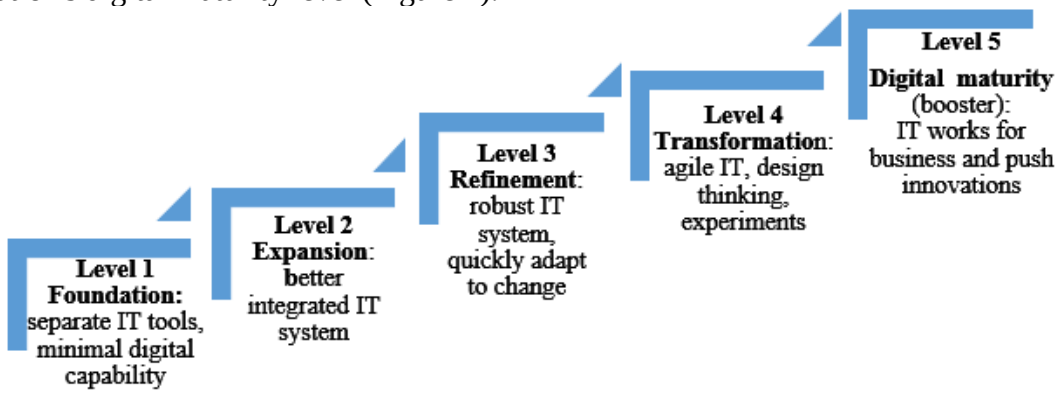



*Figure 1*. Five levels of digital maturity score (adapted) [3]

The described approach is not sufficient to judge the level of digital maturity because digital maturity is not a simple metric – it is a complex OKR for businesses. Therefore, the existing model has several disadvantages due to the difficulty in splitting and recognizing the sources and factors influencing business outcomes with greater power. To solve this problem, this article proposes an updated concept – to determine three main categories of digital maturity and calculate their weight in final business outcomes.

**Analysis of the principles and basis for integrating three components of business digital maturity**. The main task of a business under this approach is to combine three categories: organizational digital maturity, customers' digital maturity, and maturity of digital technology.

***Organizational Digital Maturity (ODM)***. Organizational digital maturity refers to the degree of development of a company's capabilities, processes, and culture in order to effectively use digital technologies.

***Vision and strategy*** – the presence of a clear digital strategy driven by leadership commitment.

*Processes and operations* – organizations must assess and optimize internal processes in order to become more efficient and responsive, through operational agility, which is the ability to adapt workflow and processes to include digital tools.

***Talent and skills*** – human resources need to have the necessary digital skills and training to effectively use digital tools and adjust to changing business needs.

***Data-driven decision-making*** – using analytics and insights to support strategic and operational decision-making.

Digital maturity is not just about adapting to technology, but about embedding digital thinking into the organization's DNA. For businesses undergoing digital transformation, reaching high organizational digital maturity is crucial to ensure that technology investments yield measurable results.

***Customer Digital Maturity (CDM)***. Customer digital maturity reflects the evolving expectations and behavior of consumers in a digital world. Understanding customer digital maturity involves:

***Digital literacy*** – the ability of customers to effectively use digital platforms and tools.

***Digital engagement*** – customers are increasingly interacting with brands through digital channels. Businesses need to provide seamless experiences across these channels to ensure customers can easily interact with the brand.

***Personalization*** – customers become more informed about digital products and services; they expect customized experiences tailored to their preferences and behaviors through

omnichannel experiences. Organizations need to leverage data analytics and customer insight to deliver relevant content and offerings to their customers.

***Feedback and interaction*** – customers are not just passive recipients of marketing; they actively participate in conversations with brands. Businesses should embrace social listening mechanisms and feedback to evaluate customer experience and improve them.

***Trust and security***– organizations need to prioritize transparency and ethical data practices in order to build trust among their customers.

***Adoption of digital services*** – the willingness of customers to switch from traditional to digital interactions (online shopping, mobile banking).

Businesses need to assess and address their customers' digital maturity in order to design solutions that meet their needs. This may involve educating customers, improving user experiences, or introducing advanced digital features.

***Technology Maturity(TM)***. The final stage is when the technology has reached mainstream adoption with proven value and is ready for practical use. Businesses must evaluate the technology cycle stage to make informed decisions. For example:

– *emerging technologies*: technologies in early stages of hype cycle (generative AI, quantum computing) offer high potential, but come with significant uncertainty and risk (Figure 2).

– *stabilizing technologies*: technologies approaching plateau of productivity (cloud AI services, intelligent applications, computer vision) are more reliable and proven value (Figure 3).

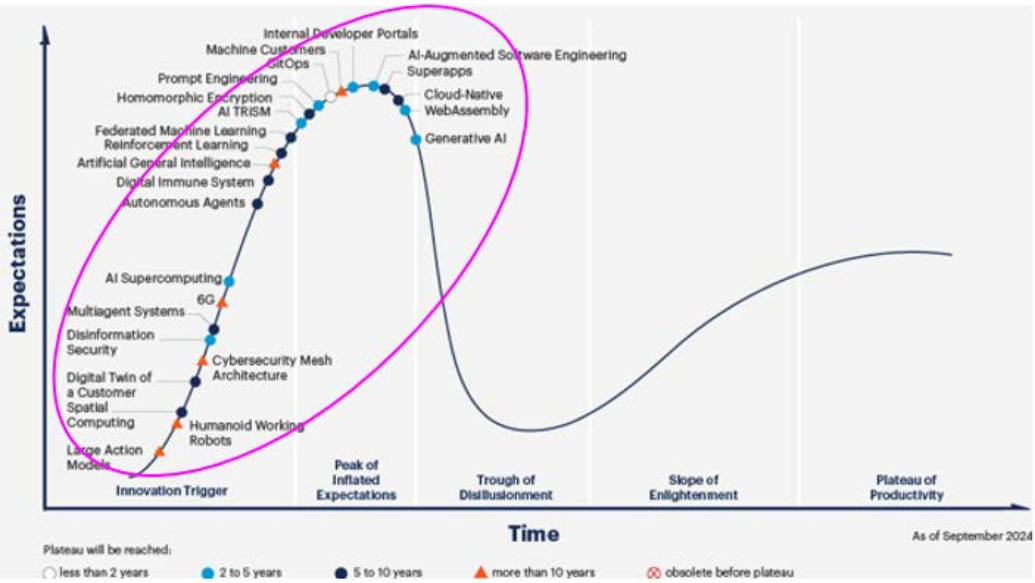



*Figure 2.* Gartner Hype Cycle for Emerging Technologies, 2024 [4]

Gartner's Hype Cycle for Emerging Technologies report (September 2024) outlines the technological innovations through the four core themes: autonomous AI; developer productivity; total experience; human-centric security and privacy [5]. These emerging technologies present big opportunities and challenges for businesses – affecting competitive differentiation and efficiency, transformational potential for business capabilities in various use cases, and conformity with organizational ability and strategy.

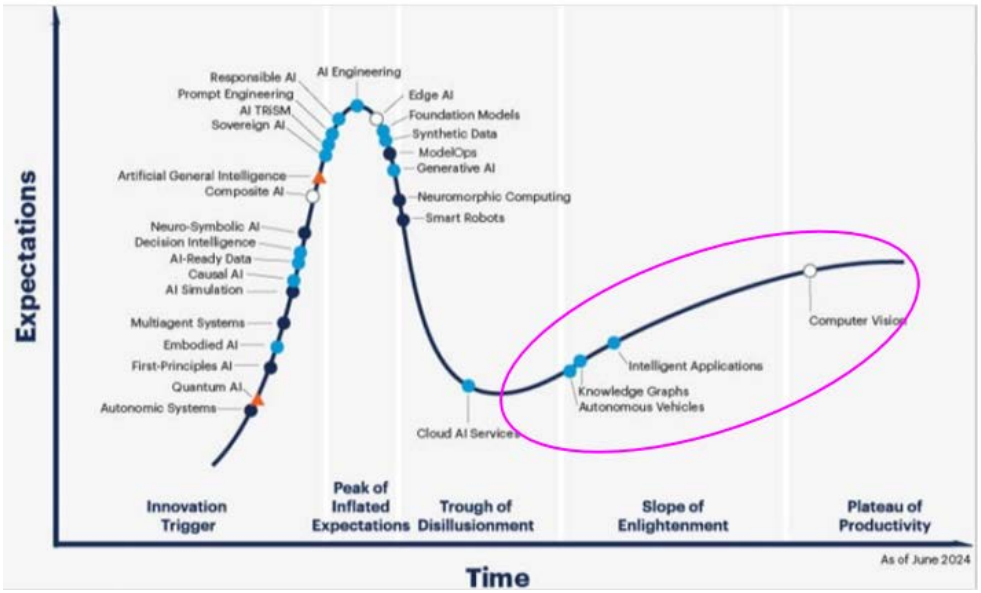



*Figure 3.* Gartner Hype Cycle for Artificial Intelligence, 2024 [5]

Difficult task for business is to determine company's strategy, design customer-driven marketing strategies and choose the right technology to find the best way of integrating business tasks with technology. To solve this issue, the concept of integrating three dimensions (ODM, CDM and TM) into one framework was proposed.

***Concept of integrating three dimensions***. The challenge of digital transformation lies in balancing these three aspects:

1 A highly mature organization with advanced technology may struggle to achieve ROI if their customer base lacks digital literacy to adopt new solutions.

2 Businesses with digitally mature customers might fail to take advantage of opportunities if their internal processes and culture are not aligned with digital objectives.

3 Introducing cutting-edge technology without considering its maturity level can lead to wasted resources and unrealized expectations.

4 Considering these facts, businesses should:

5 Regularly evaluate organizational and customer's digital maturity as well as relevant technologies' maturity.

6 Ensure that digital initiatives align both with internal capabilities and customers' readiness.

Aligning technology adoption with organizational and customer maturity ensures businesses invest in solutions that are not only innovative, but also practical and effective. The best way to achieve this is to assign specific criteria for each dimension of ODM (Organizational Digital Maturity), CDM (Customer Digital Maturity) and TM (Technology Maturity), with specific weights. This is achieved through weight coefficients based on these concepts:

ODM (40%). This relates to the core organizational processes and aligns with organization maturity models (CMMI Capability Maturity Model or ITIL – technology service model) that assess processes, governance and alignment with strategic goals.

CDM (30%). Focusing on customer-centric outcomes. It reflects principles from customer-centric frameworks (Lean Startup, Design Thinking) and outcome-based metrics (NPS, CSAT).

TM (30%). Draws on technology maturity assessments (ITIL, DevOps) and digital transformation metrics.

**CDM: 40%, ODM: 30%, TM: 30%**

The 40 -30 -30 weighting formula and composite score are not universally standardized, but are a customizable approach to prioritizing organizational needs. The formula is a simple linear

aggregation commonly used in multi-criteria decision analysis (MCDA). If an organization prioritizes customer outcomes over technology, weights can be shifted.

In addition, to calculate the score inside each model, it is recommended to use a checklist with a scoring system from 1-5 (Table 1).

Table 1. Checklist for digital maturity assessment. Scoring: 1 (Low) to 5 (High).

| Criteria | Questions / checklist | Score |
|---|---|---|
| **Organizational Digital Maturity (ODM)**<br>Objective: Assess the organization's readiness to adopt digital processes | | |
| Digital Process Automation | 1. Do they use ERP/CRM systems for core processes?<br>2. Is IoT integrated into operations? | |
| Data Utilization | 1. Do companies leverage data analytics for decision-making?<br>2. Do they employ predictive analytics? | |
| Digital Culture | 1. Do employees use digital tools confidently?<br>2. Is training on digital skills provided? | |
| Ecosystem Integration | 1. Have suppliers / partners been connected via digital platforms?<br>2. Are collaborative tools used? | |
| | Average: | |
| **Customer Digital Maturity (CDM)**<br>Objective: Assess customers' readiness to engage with digital services. | | |
| Digital Channel Adoption | 1. Do customers use online portals / mobile apps?<br>2. Is preference for chat bots / self-service used? | |
| Personalization Readiness | 1. Do customers expect tailored experiences?<br>2. Do they use of preference settings in apps? | |
| Expectations from Digital Services | 1. Do customers prioritize speed / convenience?<br>2. Do they concerns about data privacy/security? | |
| Trust in Digital Platforms | 1. Willingness to share data for personalized services.<br>2. Satisfaction with digital support | |
| | Average: | |
| **Technology Maturity (TM)**<br>Objective: Evaluate technology adoption using Gartner's Hype Cycle | | |
| Gartner Hype Cycle Stage | 1. Is technology in the "peak of inflated expectations" or "plateau of productivity"? | |
| Technology Accessibility & Cost | 1. Are technologies scalable and cost-effective?<br>2. Availability of skilled resources | |
| Adaptability to Needs | 1. Can the technology align with customer/organizational goals?<br>2. What about customization options? | |
| | Average: | |
| | **Total:** | |

Finally, using formulas, sub-criteria, and scoring, the concept was created for practical application in businesses. The concept of integrating three digital models into one framework is described in Table2.

Table 2. Framework: digital maturity evaluation and maturity score calculation

| **Category** | | |
|---|---|---|
| **Category 1: Organizational Digital Maturity (ODM)** | **Category 2: Customer Digital Maturity (CDM)** | **Category 3: Technology Maturity (TM)** |
| **Measures the organization's readiness to adopt and integrate digital processes** | **Assesses customers' readiness to engage with digital channels and services** | **Evaluates the stage of technology adoption using Gartner's Hype Cycle** |
| **Sub-criteria:**<br>- **Digital Process Automation (ERP, CRM, IoT adoption).** | **Sub-criteria:**<br>- **Digital Channel Adoption (use of online ordering, mobile** | **Sub-criteria:**<br>- **Gartner Hype Cycle Stage (Peak of Inflated Expectations,** |

| | | |
|---|---|---|
| - **Data Utilization (data analytics, Big Data, predictive insights).**<br>- **Digital Culture (employee skills and adoption of digital tools).**<br>- **Ecosystem Integration (collaboration with suppliers/partners via digital platforms)** | **apps, chatbots).**<br>- **Personalization Readiness (preference for tailored digital experiences).**<br>- **Expectations from Digital Services (speed, convenience, security).**<br>- **Trust in Digital Platforms (willingness to share data for personalized services)** | **Plateau of Productivity).**<br>- **Technology Accessibility and Cost (availability, scalability, cost-effectiveness)** |
| Scoring: 1 (Low) to 5 (High). | Scoring: 1 (Low) to 5 (High). | Scoring: 1 (Low) to 5 (High) |
| **Total Score Calculation:** Sum scores for **ODM** , **CDM** , and **TM**<br>Total possible score: **15** (5 points per category) | | |
| **Weight (depend on company priority)** | | |
| **40%** | **30%** | **30%** |
| **Maturity Clusters:**<br>Leader (12–15): high maturity in all areas.<br>Moderately Mature (7–11): some strengths but gaps exist.<br>Low Maturity (0–6): critical improvements needed | | |
| Formula: Composite Score=(0.4×ODM)+(0.3×CDM)+(0.3×TM) | | |

Integrating three dimensions (ODM, CDM, and TM) provides a clear description of the most important components of a framework in terms of a model. In this context, each component can be considered as part of the framework, and is used for evaluating maturity – the ability to achieve business goals and calculate a digital maturity score for improvements.

The concept of integrating the ODM, CDM and TM components of digital maturity is based on a structured formulation of four core principles that support this concept from the perspective of maturity (Table 3). These principles ensure cohesion, flexibility, and actionable insights.

Principle 1: Systematic integration of maturity dimensions. Maturity cannot exist in isolation in any one dimension; progress in one aspect amplifies or limits progress in others.

Principle 2. Dynamic balance of priorities. The weights of maturity (40% ODM, 30% CDM, and 30% TM) must adapt to the organization's strategy, industry environment, and market demands.

Principle 3. Maturity as an iterative journey, not a final destination. Maturity evolves through feedback loops and small improvements in each dimension. It is very important to follow the CMMI regulation (capability maturity model integration) – which is an advanced framework designed to improve and integrate processes across various disciplines such as software engineering, systems engineering, and HR management.

Principle of Data-driven alignment of outcomes. Maturity indicators must be measurable and aligned with business outcomes (e.g., revenue, customer retention, innovation).

Table 3. Summary: concept' principles and their linkages

| Principle | ODM focus | CDM focus | TM focus |
|---|---|---|---|
| Systemic integration of maturity dimensions | Processes enable CX/tech | CX drives org priorities | Tech scales processes/CX |
| Dynamic balance of priorities | Adjust weights for strategy | Align with market demands | Adapt to innovation cycles |
| Maturity as a journey, not a destination | CMMI levels, process rigor | CX evolution, personalization | Tech adoption, scalability |
| Data-Driven Alignment of outcomes | Audit scores, efficiency | NPS, CLV | Uptime, AI accuracy |

The proposed concept could be adapted to different industries. In this case, it is necessary to know industry standards and specific requirements. To adjust this concept to the semiconductor industry (as an example), it is necessary to understand that there are unique challenges.

Semiconductors are capital-intensive with long development cycles, high-precision equipment, and strong requirements for technology. Key aspects include research and development, supply chain complexity, advanced manufacturing processes, and stringent quality control. Customer demands for smaller, faster and more energy-efficient chips are constantly increasing. Furthermore, industry-specific frameworks like SEMI standards should also be considered. An example of the application of integrated concepts in the semiconductor industry can be seen in Table 4.

Table 4. Integrated principles and their linkages for semiconductor

| Principle | ODM focus | CDM focus | TM focus |
|---|---|---|---|
| Systemic integration of maturity dimensions | Yield optimization. | Customer co-design (foundry) | AI/ML for process control. |
| Dynamic balance of priorities | Precise product design. Process rigor. Lean. | Time-to-market. Cycle time. | Advanced node investment. |
| Maturity as a journey, not a destination | SEMI compliance. Design thinking. Agile. | Six sigma. Defect prevention | Smart factory adoption. |
| Data-Driven Alignment of outcomes | Cycle time. SEMI metrics VSM (value stream map). | NPS. CSAT Defect density. | IT platforms. IaaS Business Ecosystem |

**Conclusion.** Success of digital transformation hinges on effective integration of organizational digital maturity, customer digital maturity and maturity of digital technologies as outlined by the Gartner Hype Cycle. Understanding and harmonizing these dimensions is essential for business success in a competitive market.

## Reference list

[1] Цифровая зрелость как фактор конкурентного преимущества в бизнесе / Л. Ф. Медведева, Л. И. Архипова // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня: VII Международная научно-практическая конференция [Электронный ресурс]: сборник материалов VII Международной научно-практической конференции, Минск, 19-20 мая 2021 года / Белорусский государственный университет информатики и радиоэлектроники; редкол.: В. А. Богуш [и др.]. – Минск, 2021. – С. 86–98.

[2] Digital Business Maturity Model: 9 Essential Competencies to Assess Digital Business Maturity [Электронный ресурс]. – Режим доступа: https://www.gartner.com/en/documents/3983264/digital-business-maturity-model-9-essential-competencies. – Дата доступа: 10.02.2025

[3] Digital Maturity Score (Examples, Levels, and Key elements) [Электронный ресурс]. – Режим доступа: https://www.erp-information.com/digital-maturity-scorehttps://www.linkedin.com/pulse/evolution-digital-business-ecosystems-ashraf-alsadiq/. – Дата доступа: 10.03.2025

[4] The 2024 Hype Cycle for Emerging Technologies rests on 4 core themes **[**Электронный ресурс**]. –** Режим доступа: https://www.gartner.com/en/articles/hype-cycle-for-emerging-technologies. – Дата доступа: 11.02.2025

[5] The Hype Cycle for Emerging Technologies points to disruptive forces [Электронный ресурс]. – Режим доступа: https://www.gartner.com/en/articles/hype-cycle-for-emerging-technologies**.** – Дата доступа: 04.03.2025

УДК 004.272.2

# ОБРАБОТКА БОЛЬШИХ МАССИВОВ ИЗМЕРИТЕЛЬНОЙ ИНФОРМАЦИИ С ПРИМЕНЕНИЕМ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ

***Т.А. Ализаде***
*Заведующий лабораторией методов распознавания и интеллектуального анализа сигналов Института систем управления (Азербайджан), кандидат технических наук, доцент*
tahiralizada1973@gmail.com

***А.А. Мустафаев***
*Аспирант Института систем управления (Азербайджан)*
akshinmustafayev@outlook.com

***Т.А. Ализаде***

*Окончил Бакинский государственный университет. Область научных интересов связана с обработкой сигналов и разработкой методов и алгоритмов спектрального анализа, распознаванием технического состояния объектов нефтедобычи.*

***А.А. Мустафаев***

*Окончил Азербайджанский государственный университет нефти и промышленности. Область научных интересов связана с разработкой алгоритмов и программ распараллеливания вычисления спектральных характеристик сигналов.*

**Аннотация.** В настоящее время использование параллельных вычислений в информационных системах становится необходимостью, особенно когда требуется обработка больших объёмов информации в режиме реального времени. Этому способствует, прежде всего, появление и доступность многоядерных процессоров. В последнее время были разработаны новые методы оптимизации, появился фреймворк параллельного программирования, который открыл новые перспективы для решения численно сложных задач.

Исследована обработка измерительной информации с применением параллельных вычислений на примере спектрального анализа сигналов методом Фурье. Предлагается метод вычисления коэффициентов ряда Фурье, разработанный с учетом использования многоядерных процессоров с целью существенного сокращения времени обработки. Результаты вычислительных экспериментов показывают, что применение параллельного программирования позволяет в разы увеличить производительность вычислений при анализе Фурье.

**Ключевые слова:** параллельные вычисления, коэффициенты ряда Фурье, многоядерные процессоры, анализ сигналов.

**Введение.** Исторически сложилось так, что классическими источниками больших данных являются интернет вещей и социальные медиа [1]. Методы обработки больших данных изначально ориентировались на данные, соответствующие этим источникам. Однако в последние годы стали интенсивно развиваться методы обработки больших данных в промышленности и научных исследованиях, когда данными является измерительная информация, поступающая с измерительных устройств [2]. В данной работе рассмотрено применение параллельных вычислений для обработки измерительной информации на примере спектрального анализа сигналов классическим методом Фурье. В

дальнейшем планируется расширить данное исследование путём применения параллельных вычислений для неклассических методов Фурье, таких как БПФ (быстрое преобразование Фурье), оконное преобразование Фурье, полиспектральный анализ [3].

Периодический сигнал может быть представлен в течение определенного интервала времени в виде линейной комбинации ортогональных функций. Если эти ортогональные функции являются тригонометрическими, то представление в виде ряда Фурье будет выглядеть как тригонометрический ряд Фурье. Для вычисления коэффициентов тригонометрического ряда Фурье функции $f(x)$ которая определена на интервале $[0,\ T]$, используется следующая формула [4-6]:

$$f(x) = a_0 + \sum_{n=1}^{\infty} (a_n \cos\left(\frac{2\pi nx}{T}\right) + b_n \sin\left(\frac{2\pi nx}{T}\right))$$

где $a_0$ – среднее значение функции (нулевой коэффициент), $a_n$ и $b_n$ – коэффициенты ряда Фурье.

Ряд Фурье всегда является периодическим по своей природе, даже если исходная функция $f(x)$ не является периодической. Периоды $\sin\left(\frac{2\pi nx}{T}\right)$ и $\cos\left(\frac{2\pi nx}{T}\right)$ равны $\frac{2\pi}{n}$ поэтому наибольший период, $T = 2\pi$, получается из $n = 1$ членов, и ряд Фурье имеет период $2\pi$. Это означает, что ряд должен представлять периодические функции с периодом $2\pi$.

Для расчета коэффициентов мы будем использовать формулы, приведенные ниже [7-13]:

1 Для среднего значения $a_0$:

$$a_0 = \frac{1}{N} \sum_{k=0}^{N-1} f(x_k)$$

2 Для коэффициентов $a_n$:

$$a_n = \frac{2}{N} \sum_{k=0}^{N-1} f(x_k) \cos\left(\frac{2\pi kn}{N}\right), n = 1,2, \ldots$$

3 Для коэффициентов $b_n$:

$$b_n = \frac{2}{N} \sum_{k=0}^{N-1} f(x_k) \sin\left(\frac{2\pi kn}{N}\right), n = 1,2, \ldots$$

где $x_k$ представляют собой точки выборки сигнала, $N$ – количество точек выборки, $k$ – индекс точки.

**Понимание процессов обработки сигналов и оптимизации производительности.** Производители оборудования предоставляют инструменты, библиотеки и среды для разработки, тестирования, проверки и масштабирования параллельных алгоритмов и рабочих нагрузок как для реальных приложений, так и для исследовательских целей. Например, Hewlett Packard Enterprise (HPE) предоставляет среду *Cray Compiler Environment*, совместимую со многими стандартами, включая стандарт *OpenMP*, которая выполняет анализ кода во время компиляции и автоматически генерирует высоко оптимизированный код [14].

Технические характеристики аппаратных вычислительных средств, используемых для проведения расчетов и испытаний, приведены в таблице 1 [15].

Таблица 1. Технические характеристики аппаратных вычислительных средств

| ***Технические характеристики*** | ***Сервер HP ProLiant DL 360 Gen9*** |
|---|---|
| Процессор | Intel Xeon e5-2660 v3 2.6GHZ |

| | |
|---|---|
| Сокет | Intel Socket 2011 |
| Операционная система | Windows Server 2016 |
| Микроархитектура | Sandy Bridge-EP |
| Память | 256 GB |
| Количество ядер | 8 |
| Количество потоков | 16 |
| Кэш L1 | 64 KB (на ядро) |
| Кэш L2 | 256 KB (на ядро) |
| Кэш L3 | 20 MB (общее) |
| Процессор | Intel Xeon e5-2660 v3 2.6GHZ |

Обработка сигналов оказывает большое влияние на многие области, начиная от связи и аудиоанализа и заканчивая биомедицинской инженерией и радарными системами. Анализ частотных составляющих сигнала, который часто выполняется с помощью преобразования Фурье, является одним из основных методов обработки сигналов. С увеличением размера сигнала эта операция становится все более требовательной к вычислительным ресурсам. Обработка больших сигналов на современном оборудовании без оптимизации проблематична из-за сложности этих вычислений [16].

Для решения этой проблемы было разработано множество алгоритмов, однако, поскольку наборы данных и требования к обработке в режиме реального времени продолжают расти, даже использование этих алгоритмов может оказаться не столь эффективным. Именно здесь на помощь приходит параллелизм. Используя возможности современных многоядерных процессоров, параллелизм позволяет разбивать вычисления на небольшие блоки и распределять нагрузку между многими ядрами, значительно сокращая время выполнения.

В этой части мы рассмотрим реализацию и оптимизацию алгоритмов обработки сигналов на C++ [17], уделяя особое внимание следующему:

1 Генерация тестовых сигналов, имитирующих реальные сигналы.

Функция *generateSignal* возвращает результат дискретной версии непрерывной функции $f(x)$ путем ее дискретизации через регулярные интервалы в определенном диапазоне. Это действие начинается с разбиения диапазона на равноотстоящие друг от друга интервалы numPoints. Каждый из этих интервалов соотносится с размером шага (*step*). Для каждой точки выборки функция вычисляет координату *x* и вычисляет значение $f(x)$, а затем сохраняет результат в векторе с именем *signal*. Выбор данной реализации обусловлен тем, что в отличие от генерации случайных чисел, она предоставляет более адекватные результаты для анализа. Ниже приведен фрагмент кода генерации тестового сигнала:

```
std::vector<double> generateSignal(int numPoints)
{
    std::vector<double> signal(numPoints);
    double step = 2 * PI / numPoints;
    for (int i = 0; i < numPoints; ++i)
    {
        double x = -PI + i * step;
        signal[i] = f(x);
```

```
    }
    return signal;
}
```

2 Реализация непараллельных параллельных вычислений коэффициентов ряда Фурье.

Простое представление вычисления коэффициентов ряда Фурье из сгенерированного сигнала находится в функции *computeFourierCoefficients*. Стоит обратить внимание, что в этой функции не используются методы распараллеливания.

```
void computeFourierCoefficients(const std::vector<double> &signal, int maxN, std::vector<double> &an, std::vector<double> &bn)
{
    int numPoints = signal.size();
    double step = 2 * PI / numPoints;
    // Compute a0
    an[0] = 0.0;
    for (int i = 0; i < numPoints; ++i)
    {
        an[0] += signal[i];
    }
    an[0] *= step / (2 * PI);
    // Compute an and bn
    for (int n = 1; n <= maxN; ++n)
    {
        double an_local = 0.0;
        double bn_local = 0.0;
        for (int i = 0; i < numPoints; ++i)
        {
            double x = -PI + i * step;
            an_local += signal[i] * cos(n * x);
            bn_local += signal[i] * sin(n * x);
        }
        an[n] = (2.0 / numPoints) * an_local;
        bn[n] = (2.0 / numPoints) * bn_local;
    }
}
```

Предлагаемая нами функция *computeFourierCoefficientsParallel* повышает эффективность за счет разделения вычислений между несколькими потоками. Внешний цикл, в котором вычисляются коэффициенты, разделяется между потоками. Количество потоков определяется *OpenMP* автоматически, по умолчанию оно обычно равно количеству ядер. Каждый поток вычисляет определенный диапазон частот независимо друг от друга. Этот метод позволяет оптимально использовать многоядерные процессоры, что ускоряет вычисления, особенно для больших сигналов [18-21]:

```
void computeFourierCoefficientsParallel(const std::vector<double> &signal, int maxN, std::vector<double> &an, std::vector<double> &bn)
{
    int numPoints = signal.size();
    double step = 2 * PI / numPoints;
    // Compute a0. Parallelizing a0 computation may introduce overhead
    // a0 is computed differently from an and bn
    an[0] = 0.0;
    for (int i = 0; i < numPoints; ++i)
    {
        an[0] += signal[i];
    }
    an[0] *= step / (2 * PI);
    // Compute an and bn (parallelized outer loop over n)
```

```
#pragma omp parallel for
    for (int n = 1; n <= maxN; ++n)
    {
        double an_local = 0.0;
        double bn_local = 0.0;
        for (int i = 0; i < numPoints; ++i)
        {
            double x = -PI + i * step;
            an_local += signal[i] * cos(n * x);
            bn_local += signal[i] * sin(n * x);
        }
        an[n] = (2.0 / numPoints) * an_local;
        bn[n] = (2.0 / numPoints) * bn_local;
    }
}
```

3 Оценка производительности для демонстрации преимуществ использования параллелизма.

Измерение производительности выполнения различных методов осуществляется с помощью библиотеки *std::chrono* [22]. Расчет времени выполнения определяется путем вычитания времени начала и окончания функции. Для записи точного времени используются часы с высоким разрешением. Этот подход показывает, как каждый метод справляется с увеличением объема данных, и подчеркивает, как параллелизм сокращает время вычислений, особенно для больших сигналов.

**Блок-схема распараллеливания.** Блок-схема распараллеливания (рисунок 1) показывает, как задачи разделяются и выполняются параллельно для оптимизации производительности при выполнении функции *computeFourierCoefficientsParallel* где:

1 Вычисление $a_0$ выполняется последовательно, поскольку оно не зависит от частотных членов и, таким образом, распараллеливание его вычислений может привести к дополнительным расходам вычислительных ресурсов. Этот коэффициент вычисляется путем суммирования значений сигналов.

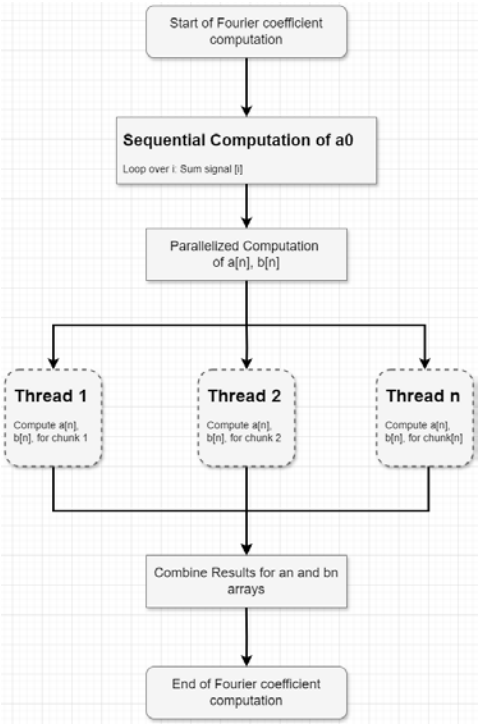



*Рисунок 1.* Блок-схема распараллеливания

2 В коде функции перед внешним циклом *for* указывается строка с *#pragma omp parallel for*. По умолчанию *OpenMP* выбирает количество потоков на основе доступных ядер. После определения количества ядер каждый поток закрепляется за определенным ядром с определенным диапазоном значений $n$, который отражает различные гармонические частоты. В процессе вычислений суммируются произведения сигнала с косинусами и синусами для каждого назначенного $n$. Потоки работают параллельно, поэтому между ними нет пересечений или зависимостей. После того как каждый поток завершает вычисления для назначенного ему диапазона, результаты для $a_n$ и $b_n$ объединяются в массивы.

**Компьютерные эксперименты и анализ их результатов.** Выполнив вышеизложенный код, мы получили результаты компьютерных экспериментов при непараллельной и параллельной реализациях: в таблице 2 показано соотношение между количеством коэффициентов ряда Фурье и временем вычислений (в миллисекундах) при непараллельной и параллельной реализациях.

Таблица 2. Результаты компьютерных экспериментов

| ***Количество коэффициентов ряда Фурье*** | ***Время при непараллельном вычислении, мс*** | ***Время при параллельном вычислении, мс*** |
|---|---|---|
| 10 | 0.8183 | 0.5169 |
| 100 | 6.9809 | 2.293 |
| 200 | 13.9411 | 3.9712 |
| 300 | 21.1175 | 6.644 |
| 400 | 27.8889 | 7.2111 |
| 500 | 34.8119 | 9.2266 |
| 1000 | 71.1401 | 19.4607 |

Соотношение между количеством коэффициентов ряда Фурье при непараллельной и параллельной реализациях показано на рисунке 2.

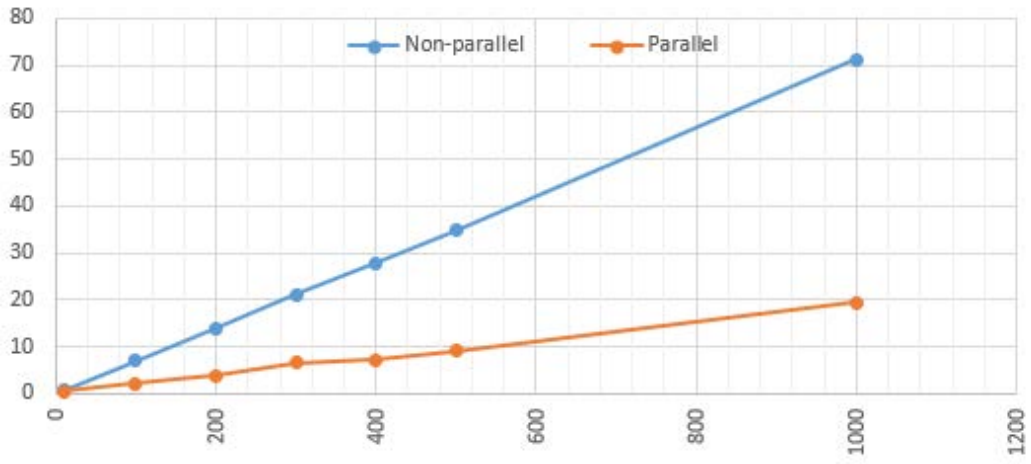



*Рисунок 2.* Зависимость числа коэффициентов ряда Фурье от времени для непараллельной и параллельной реализаций

На основании полученных результатов мы можем определить следующие ключевые

моменты:

– время выполнения непараллельной реализации значительно увеличивается с ростом числа коэффициентов ряда Фурье. Это говорит об ограниченности однопоточного подхода, при котором все вычисления выполняются последовательно. Например, при небольшом количестве термов (10) время выполнения непараллельной реализации равно 0,8183 миллисекунды. Однако при увеличении числа терминов до 1000 время увеличивается до 71,1401 миллисекунды, что указывает на нелинейный рост вычислительных затрат с увеличением числа членов;

– параллельная реализация демонстрирует меньшее время вычислений по сравнению с непараллельным вариантом. Например, при 1000 членах время параллельного выполнения (19,4 миллисекунды) значительно меньше, чем у непараллельного варианта (71,1 миллисекунды). Таким образом, скорость параллельных вычислений приблизительно в 3,7 раза выше непараллельных.

**Заключение.** Результаты данного исследования подтверждают эффективность параллельных вычислений коэффициентов ряда Фурье, особенно в сценариях с большим количеством коэффициентов. Использование параллельного фреймворка *OpenMP* позволило устранить узкое место в вычислениях, связанное с последовательной обработкой. Достигнутый прирост производительности показывает пользу распределения вычислительной нагрузки на несколько ядер процессора. Даже для небольшого количества коэффициентов ряда Фурье параллельная версия работает лучше, чем последовательная. Для дальнейшей оптимизации вычислений можно использовать специализированные стандарты, такие как интерфейс передачи сообщений (*MPI*) для распределенных систем. Данная работа показывает, что параллельные вычисления важны не только с точки зрения эффективности, но и позволяют справиться с требованиями, предъявляемыми современной обработкой сигналов. Результаты, полученные в рамках данной работы, могут быть развиты и в дальнейшем применены для оптимизации вычислений в других методах анализа сигналов, таких как БПФ (быстрое преобразование Фурье), оконное преобразование Фурье, полиспектральный анализ [3] и т.д.

## Список литературы

[1] Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung. Big Data. Related Technologies, Challenges, and Future Prospects. Spinger, 2014. — 100 p. doi.org/10.1007/978-3-319-06245-7

[2] Черняк Л. Большие Данные — новая теория и практика // Открытые системы. СУБД, №10, 2011. www.osp.ru/os/2011/10/13010990

[3] Новиков А.К. Полиспектральный анализ. СПб.: ЦНИИ им. акад. А. Н. Крылова, 2002. - 178 с.

[4] R. E. Edwards, “Fourier Series: A Modern Introduction Volume 1”, Институт перспективных исследований, Австралийский национальный университет, Canberra, Второе издание, Декабрь. 2012, с. 14-43.

[5] Антоний Зигмунд, “Trigonometric Series”, Cambridge University Press, Февраль. 2015, с. 1-4.

[6] Марк А. Пински, "Introduction to Fourier Analysis and Wavelets", Северо-Западный университет, 2008, с. 13-34.

[7] Данхэм Джексон, "Fourier Series and Orthogonal Polynomials", American Mathematical Soc., 1941, с. 1-42.

[8] Тим Олсон, "Applied Fourier Analysis", Birkhäuser New York, NY, 2017, с. 1-3, 19-30, 45-80.

[9] Дэвид В. Каммлер, "A First Course in Fourier Analysis", Cambridge University Press, 2010, с. 1-40.

[10] Libretexts, "Fourier Trigonometric Series", [Онлайн], Доступно по адресу: https://math.libretexts.org/Bookshelves/Differential_Equations/Introduction_to_Partial_Differential_Equations_(Herman)/03%3A_Trigonometric_Fourier_Series/3.02%3A_Fourier_Trigonometric_Series.

[11] Университет Северной Каролины, "Trigonometric Fourier Series", [Онлайн], Доступно по адресу: https://people.uncw.edu/hermanr/pde1/pdebook/Fourier.pdf.

[12] Википедия, “Fourier series”, [Онлайн], Доступно по адресу: https://en.wikipedia.org/wiki/Fourier_series.

[13] Tutorialspoint, "Trigonometric Fourier Series – Definition and Explanation", [Онлайн], Доступно по адресу: https://www.tutorialspoint.com/trigonometric-fourier-series-definition-and-explanation.

[14] Hewlett Packard, "Cray Compiler Environment", [Онлайн], Доступно по адресу: https://cpe.ext.hpe.com/docs/24.03/cce/index.html#overview.

[15] TechPowerUP, “Intel Xeon E5-2660”, [Онлайн], Доступно по адресу: https://www.techpowerup.com/cpu-specs/xeon-e5-2660.c976.

[16] Кули, Дж. У. и Тьюки, Дж. У. (1965), “An Algorithm for the Machine Calculation of Complex Fourier Series”, Математика вычислений, с. 297–301.

[17] Справочник по C++, "Standard Containers”, [Онлайн], Доступно по адресу: https://cplusplus.com/reference/stl.

[18] CsInParallel, "Parallel Computing in the Computer Science Curriculum", [Онлайн], Доступно по адресу: https://csinparallel.org/csinparallel/ppps/openmp.html.

[19] Microsoft, "Enable OpenMP Support", [Онлайн], Доступно по адресу: https://learn.microsoft.com/en-us/cpp/build/reference/openmp-enable-openmp-2-0-support?view=msvc-170.

[20] OpenMP Документация, "Guide to parallelizing computations in C++ for multi-core processors", [Онлайн], Доступно по адресу: https://www.openmp.org.

[21] OpenMP Примеры, "OpenMP Application Programming Interface Examples", [Онлайн], Доступно по адресу: https://www.openmp.org/wp-content/uploads/openmp-examples-5.2.2-final.pdf.

[22] Справочник по C++, "Date and time utilities" 2024, [Онлайн], Доступно по адресу: https://en.cppreference.com/w/cpp/chrono.

### Авторский вклад

**Ализаде Тахир Али** – постановка задачи, консультации при проведении компьютерных экспериментов и анализе их результатов, формирование структуры статьи.

**Мустафаев Акшин Ариф** – разработка программного кода для классических и параллельных вычислений коэффициентов ряда Фурье, проведение компьютерных экспериментов, анализ полученных результатов.

## PROCESSING LARGE ARRAYS OF MEASUREMENT INFORMATION USING PARALLEL COMPUTING

***T.A. Alizada***
*Head of Laboratory of methods of recognition and intelligent analysis of signals of Institute of Control Systems (Azerbaijan), PhD of Technical sciences, Associate Professor*

***A.A. Mustafayev***
*Postgraduate student at the Institute of Control Systems (Azerbaijan)*

**Annotation.** Currently, the use of parallel computing in information systems has become a necessity, especially when it comes to processing large volumes of data in real time. This is primarily driven by the emergence and availability of multicore processors. In recent years, new optimization methods have been developed, and a parallel programming framework has been introduced, opening up new prospects for solving numerically complex tasks.

This study explores the processing of measurement data employing parallel computing, using the example of spectral signal analysis via the Fourier method. A method for computing Fourier series coefficients is proposed, designed with multicore processors in mind, aiming to significantly reduce processing time. The results of computational experiments demonstrate that using parallel programming can greatly increase computation performance during Fourier analysis.

**Keywords:** parallel computing, Fourier series coefficients, multi-core processors, signal analysis.

УДК 004.65+001.895

# ИННОВАЦИОННЫЕ ПОДХОДЫ К АНАЛИЗУ БОЛЬШИХ ДАННЫХ: ОТ ТЕОРИИ К ПРАКТИКЕ

***В.В. Верняховская***
*Заместитель декана инженерно-экономического факультета БГУИР, магистр экономических наук*
verniahovskaya@bsuir.by

***О.М. Раптунович***
*Старший преподаватель кафедры экономической информатики БГУИР*
oraptunovich@gmail.com

***В.В. Верняховская***

*Окончила Белорусский государственный университет информатики и радиоэлектроники. Магистр экономических наук. Работает заместителем декана инженерно-экономического факультета БГУИР. Направления исследований: трансфер технологий, инновационная деятельность, информационные технологии в маркетинге.*

***О.М. Раптунович***

*Окончила Частый Институт Управления и Предпринимательства. Область научных интересов связана с актуальными вопросами экономики, логистики, маркетинга и информационных технологий и способами их решений.*

**Аннотация.** Выполнен обзор современных инновационных подходов к анализу больших данных с использованием глубокого обучения, искусственного интеллекта и квантовых вычислений. Рассмотрены ключевые алгоритмы и технологии, обеспечивающие эффективную обработку и интерпретацию больших объемов информации. Проанализированы примеры практического применения перспективных подходов в различных отраслях деятельности человека.

**Ключевые слова:** большие данные, искусственный интеллект, глубокое обучение, квантовые вычисления.

**Введение.** Как общество, мы находимся в потоке информации обо всех аспектах нашей жизни – от социальных взаимодействий и научных открытий до профессиональной карьеры и личного здоровья. По мере развития технологий мы начали генерировать все больше данных, и их объем достиг уровня, когда традиционные инструменты обработки перестали справляться. В результате появилось понятие «большие данные», описывающее огромные и трудноуправляемые массивы информации.

Развитие информационных технологий приводит к тому, что часть данных изначально возникает в цифровом формате и реализуется через Интернет. Это значительно ускоряет процесс их распространения и увеличивает объем информации, подлежащей анализу. В связи с этим перед исследователями и разработчиками аналитических систем встает задача поиска новых методов и инструментов, позволяющих эффективно обрабатывать и интерпретировать большие массивы данных в условиях их постоянного роста и изменения.

Несмотря на существенный прогресс в области развития компьютерных систем и интернет-технологий, проблемы, связанные с обработкой и управлением данными, сохраняют свою актуальность в эпоху больших данных. Термин «большие данные»

предполагает такие масштабы информации, которые не могут быть эффективно обработаны и проанализированы с использованием большинства существующих информационных систем и традиционных методов. Это обусловлено не только тем, что объем данных превышает возможности одного компьютера, но и тем, что классические методы интеллектуального анализа данных, разработанные для централизованных систем, зачастую оказываются неприменимыми для работы с большими данными.

В данной статье рассмотрены современные инновационные методы анализа больших данных, которые находят применение как в теоретических исследованиях, так и в практической деятельности. Особое внимание уделено переходу от теоретических моделей к их практической реализации.

**Инновационные подходы в анализе больших данных.** С увеличением объемов информации традиционные методы обработки данных теряют свою эффективность, что обусловливает необходимость внедрения передовых технологий, таких как глубокое обучение (Deep Learning, DL) и искусственный интеллект (AI). Эти технологии позволяют автоматизировать процессы анализа данных, выявлять сложные закономерности и ускорять принятие решений. В свою очередь, рисунок 1 показывает прогноз на 3–5 лет, который свидетельствует, что 95% компаний намерены стать «умными предприятиями»» и активно использовать аналитику данных, что свидетельствует о растущей потребности в автоматизированных методах машинного обучения и искусственного интеллекта [1].

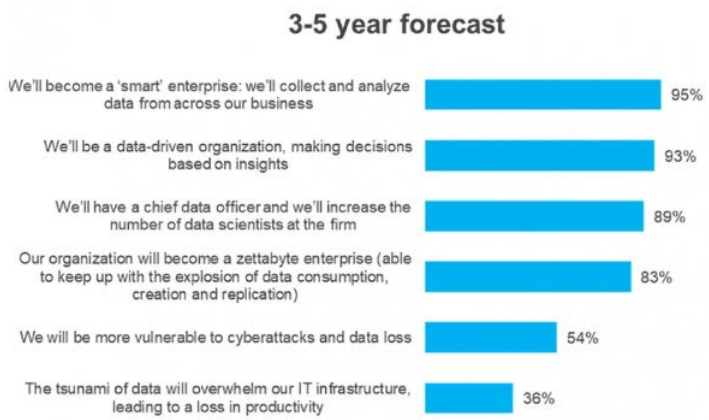



*Рисунок 1.* Прогноз организаций об использовании данных

В то же время чуть менее 40% организаций не уверены в своей способности работы с потоком больших данных, поскольку не смогут обрабатывать зеттабайты данных в будущем. То есть в условиях стремительного роста объемов данных и ограниченности традиционных методов обработки, использование автоматизированных решений становится не только целесообразным, но и необходимым. Одна из новых тенденций заключается в том, что в последние годы эти инструменты стали более востребованными для компаний любого размера и в самых разных отраслях.

Способность глубокого обучения извлекать высокоуровневые и сложные абстракции, а также представления данных из больших объемов информации, особенно неразмеченных данных, делает его ценным инструментом для анализа больших данных. Однако знания, полученные с помощью алгоритмов глубокого обучения и доступные для использования, в значительной степени остаются неиспользованными в контексте анализа больших данных.

В частности, такие задачи, как семантическая индексация, маркировка данных, быстрый поиск информации и дискриминативное моделирование, могут быть более эффективно решены с применением глубокого обучения. Традиционные методы

машинного обучения и инженерии признаков оказываются недостаточно эффективными для выявления сложных и нелинейных закономерностей, характерных для больших данных. Глубокое обучение, извлекая такие признаки, позволяет использовать относительно простые линейные модели, что особенно важно при разработке методов анализа данных в условиях их масштабности.

Глубокое обучение представляет собой ключевое направление в развитии искусственного интеллекта, целью которого является автоматизация процесса извлечения многоуровневых представлений (абстракций) из данных. Данный подход во многом вдохновлён принципами работы человеческого мозга, который анализирует, обучается и принимает решения на основе поступающей информации. Одним из центральных аспектов глубокого обучения является способность нейросетевых моделей самостоятельно выявлять сложные закономерности в данных, снижая зависимость от предварительно заданных человеком признаков.

Основу глубокого обучения составляют многослойные нейронные сети, где каждый последующий уровень абстракции формируется на основе информации, полученной с предыдущих слоёв. Благодаря такой иерархической структуре возможно представление данных в виде высокоуровневых нелинейных функций, что особенно важно для решения сложных задач искусственного интеллекта, включая обработку изображений, естественного языка и временных рядов [2].

Ключевым преимуществом глубокого обучения является возможность обучения на больших объемах неразмеченных данных, что минимизирует необходимость ручного проектирования признаков и экспертного вмешательства. Это делает глубокие нейронные сети мощным инструментом для извлечения информативных представлений, существенно повышая эффективность аналитических и прогнозных систем. В целом, глубокое обучение можно рассматривать как особый класс методов обучения представлений, где знания формируются через многоуровневый процесс преобразования входных данных.

Глубокое обучение и другие методы искусственного интеллекта позволяют автоматизировать значительную часть задач, особенно связанных с рутинными операциями, обработкой данных и сбором информации. В контексте анализа больших данных и разработки интеллектуальных систем понимание степени автоматизируемости различных профессий становится критически важным. Представленная статистика на рисунке 2 позволяет оценить, какие сферы деятельности в наибольшей степени подвержены автоматизации, а также определить потенциал применения методов глубокого обучения для оптимизации рабочих процессов [1]. Эти данные подчеркивают важность разработки стратегий адаптации к изменяющимся требованиям рынка труда и дальнейшего изучения возможностей искусственного интеллекта в сфере автоматизации.

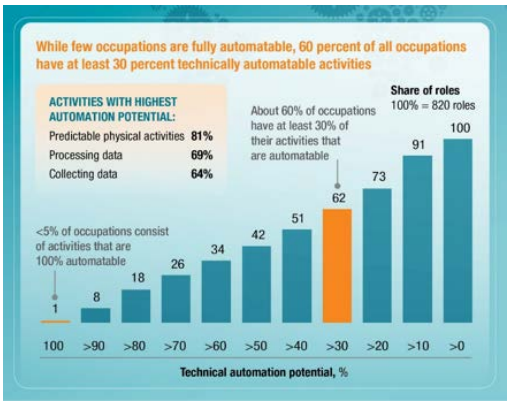



*Рисунок 2.* Степень автоматизируемости различных профессий

Диаграмма иллюстрирует степень автоматизируемости различных профессий, подчеркивая, что хотя полностью автоматизируемых профессий менее 5%, около 60%

профессий содержат не менее 30% технически автоматизируемых задач. Наибольшему риску автоматизации подвержены предсказуемые физические задачи (81%), обработка данных (69%) и сбор данных (64%).

Помимо глубокого обучения за рамки классических подходов выходят квантовые вычисления, которые в свою очередь открывают принципиально новые возможности в анализе больших данных благодаря своей способности выполнять вычисления параллельно и использовать суперпозицию и квантовую запутанность. Квантовый анализ данных (Quantum Data Analytics, QDA) обещает значительно повысить эффективность обработки информации, особенно в задачах оптимизации, моделирования и машинного обучения.

Три квантовых ресурса, ни один из которых не имеет аналогов в традиционных вычислениях, обеспечивают потенциал квантовых вычислений: квантовый параллелизм, квантовая интерференция и квантовая запутанность [3].

Квантовый параллелизм – это способность квантового компьютера одновременно обрабатывать все возможные состояния входных данных. Это принципиально отличается от классических компьютеров, которые обрабатывают данные поочередно, выполняя вычисления над одним набором входных данных за раз, что ускоряет выполнение вычислений. В классических вычислениях данные хранятся в битах, принимающих значения 0 или 1. В квантовом компьютере используются кубиты, которые могут находиться в суперпозиции состояний 0 и 1 одновременно. Хотя квантовый параллелизм обеспечивает одновременную обработку множества состояний, измерение разрушает суперпозицию – после измерения кубиты принимают определенное классическое значение (0 или 1) с определенной вероятностью. Это означает, что нельзя просто измерить все состояния сразу и получить экспоненциальное ускорение. Чтобы использовать квантовый параллелизм на практике, разработаны специальные алгоритмы (например, квантовое преобразование Фурье и амплитудное усиление, которые усиливают вероятность правильного ответа).

Квантовая запутанность – это явление, при котором состояние одного кубита мгновенно связано с состоянием другого, независимо от расстояния между ними. Благодаря способности запутанных кубитов мгновенно передавать информацию и взаимодействовать друг с другом, квантовые системы обладают потенциалом для экспоненциального ускорения вычислений, обмена данными и улучшения алгоритмов обработки данных.

Квантовая интерференция – это фундаментальное явление квантовой механики, позволяющее усиливать правильные вычислительные пути и подавлять неверные за счет когерентного наложения волновых функций. В квантовых вычислениях состояние кубитов описывается волновой функцией, которая может интерферировать конструктивно (усиливая вероятность правильного решения) или деструктивно (ослабляя вероятность неправильного решения). Квантовая интерференция позволяет усилить вероятные сценарии будущих событий, что делает прогнозирование более точным.

**Практическое применение.** Использование искусственного интеллекта в анализе больших данных получило широкое распространение в различных отраслях. Каждая сфера сталкивается с уникальными вызовами и требованиями, и ИИ предлагает адаптивные решения для эффективного их преодоления.

В финансовом секторе ИИ играет ключевую роль в обработке огромных объемов данных для различных целей. Он используется для прогнозирования цен на акции, оценки кредитных рисков и выявления мошеннических операций. Применение прогнозной аналитики является неотъемлемой частью этапов процесса анализа данных, где методы сбора и подготовки информации играют ключевую роль в обеспечении точности и достоверности результатов аналитических выводов [4].

American Express, финансовая корпорация, применяет искусственный интеллект и прогностическую аналитику для выявления мошеннических транзакций. Система анализирует модели поведения и потребительские привычки в расходах, что позволяет

обнаруживать транзакции, подозрительные на мошенничество, и предупреждать как компанию, так и клиента в реальном времени, с целью минимизации потерь.

Здравоохранение – еще одна область, где влияние ИИ на анализ больших данных является революционным. Алгоритмы ИИ способны выявлять закономерности в медицинских данных, что способствует ранней диагностике заболеваний, разработке новых лекарств и определению наиболее эффективных методов лечения.

Аналогично тому, как глубокое обучение используется для анализа изображений, в медицине нейросетевые модели способны анализировать медицинские данные и выявлять заболевания на ранних стадиях. Например, при диагностике онкологических заболеваний сверточная нейронная сеть на первом уровне может выявить базовые текстурные особенности тканей, на втором уровне – обнаружить более сложные паттерны, указывающие на аномальные клетки, а на третьем уровне – определить наличие злокачественных образований, отличая их от доброкачественных. Эти представления используются для автоматического выявления раковых опухолей по рентгеновским, КТ или МРТ-снимкам, позволяя врачам быстрее и точнее диагностировать заболевание. Однако, как и в случае с обработкой изображений, глубокое обучение не следует строго заданной последовательности признаков, а адаптируется к данным, находя наиболее информативные характеристики, влияющие на диагностику.

Этот алгоритм анализирует снимки на различных уровнях абстракции: на первом этапе он обнаруживает текстурные особенности ткани, на втором – выделяет подозрительные области с аномальной плотностью, а на третьем уровне оценивает вероятность наличия опухоли, сравнивая результаты с обширной базой данных. В результате система способна обнаруживать рак молочной железы на ранних стадиях, даже в тех случаях, когда изменения слишком незначительны для визуального восприятия врача.

Образовательные технологии также переживают трансформацию благодаря глубокому обучению. Адаптивные обучающие платформы, использующие нейросетевые алгоритмы, персонализируют учебный процесс, подстраиваясь под уровень знаний и индивидуальные потребности студентов. Автоматизированные системы оценки успеваемости, анализируя ответы учащихся, предоставляют рекомендации по улучшению учебного процесса и помогают преподавателям оптимизировать методы обучения.

Одним из ярких примеров применения глубокого обучения в образовании является платформа Coursera, которая использует нейросетевые алгоритмы для персонализации обучения [5]. Система анализирует успеваемость студентов, их взаимодействие с курсами и временные затраты на выполнение заданий, чтобы предлагать оптимальные учебные траектории. На основе данных об ошибках и успехах студентов алгоритмы могут рекомендовать дополнительные материалы, адаптировать сложность задач и предлагать повторение тем, вызывающих затруднения.

Чтобы использовать преимущества квантовых ресурсов на практике, необходимо применять специализированные алгоритмы, которые способны эффективно преобразовывать квантовые состояния в полезные вычислительные результаты.

В основе квантового преобразования Фурье лежит представление данных в виде комплексных амплитуд, которые могут интерферировать друг с другом. Это позволяет выявлять частотные компоненты в данных и ускорять такие задачи, как спектральный анализ временных рядов, что особенно полезно при обработке сигналов, финансовых прогнозов и анализе тенденций в больших данных, распознавании образов.

Алгоритм Шора использует квантовое преобразование Фурье и квантовую интерференцию для поиска периодичности в числовых последовательностях. В области анализа больших данных он может быть использован для решения сложных задач, связанных с факторизацией чисел, структурированием данных и оптимизацией. Также алгоритм Шора особенно применим в аспекте защиты информации и обеспечения безопасности, что часто необходимо при работе с большими данными [6]. Крупные

технологические компании, такие как Microsoft и Google, уже ведут активные исследования в области квантово-устойчивых алгоритмов шифрования. В настоящее время эти технологии находятся на стадии теоретической проработки и тестирования, а их основная задача – успешная интеграция в существующую цифровую инфраструктуру.

Ещё одним эффективным алгоритмом является алгоритм Гровера, предназначенный для поиска информации в неструктурированных базах данных и решения задач оптимизации с квадратичным ускорением по сравнению с классическими методами. При обработке огромных массивов неструктурированных данных квантовый поиск может быть использован для ускорения таких задач, как обнаружение аномалий, фильтрация шумов или выявление ключевых паттернов. Алгоритм Гровера помогает быстрее вычислять наиболее выгодную стратегию. Volkswagen уже продемонстрировал практическое применение квантовых вычислений для оптимизации дорожного движения. Их алгоритм квантовой маршрутизации, работающий на квантовом компьютере& Квантовые технологии активно исследуются и другими автопроизводителями, включая BMW, Toyota и Ford, которые инвестируют в развитие квантовых решений для транспортной индустрии.

**Заключение.** Развитие методов анализа больших данных претерпело значительные изменения благодаря внедрению глубокого обучения и искусственного интеллекта. Современные нейросетевые модели демонстрируют высокую эффективность в обработке сложных структур данных, автоматизации принятия решений и выявлении скрытых закономерностей. Однако, с увеличением объёмов информации классические вычислительные методы сталкиваются с ограничениями в скорости и вычислительных возможностях, что делает необходимым поиск альтернативных подходов.

## Список литературы

[1] Explodingtopics.com [Электронный ресурс] – Режим доступа: https://explodingtopics.com/blog/big-data-trends.

[2] Yann Lecun, Yoshua Bengio, Geoffrey Hinton. Deep learning. Nature, 2015, 521 (7553), pp.436-444.

[3] Соловьев В.М. Квантовые компьютеры и квантовые алгоритмы. Математика. Механика. Информатика. 2016, Т. 16, вып. 1. – 104-111 с.

[4] Усова В.А., Зинкович Е.А., Пригожий М.К, Раптунович О.М. Использование нейронных сетей для проведения прогнозной аналитики. Актуальные вопросы экономики и информационных технологий, 2024. – 86-90 с.

[5] Alexobergan.ru [Электронный ресурс] – Режим доступа: https://alexobergan.ru/superobuchenie-kak-modeli-glubokogo-obucheniya-uskoryayut-proczess-obrazovaniya/.

[6] Борисов В.В., Дегтярев Д.В. Реализация алгоритма квантовой факторизации Шора. Вестник АмГУ, выпуск 103, 2023 – 3-9 с.

[7] New-science.ru [Электронный ресурс] – Режим доступа: https://new-science.ru/12-luchshih-primenenij-kvantovyh-vychislenij/.

# INNOVATIVE APPROACHES TO BIG DATA ANALYSIS: FROM THEORY TO PRACTICE

***O.M. Raptunovich***
*Senior Lecturer at the Department of Economic Informatics BSUIR*

***Y.A. Zinkovich***
*Student of the Engineering and Economics Faculty, specializing in «Information Systems and Technologies (in Economics)» BSUIR*

***M.K. Prigozhiy***
*Student of the Engineering and Economics Faculty, specializing in «Information Systems and Technologies (in Economics)» BSUIR*

**Abstract.** A review of modern innovative approaches to big data analysis using deep learning, artificial intelligence, and quantum computing has been conducted. Key algorithms and technologies ensuring efficient processing and interpretation of large volumes of information are examined. Practical applications of promising approaches in various fields of human activity are analyzed.

**Keywords**: big data, artificial intelligence, deep learning, quantum computing.

УДК 004.925

# ПРЕОДОЛЕНИЕ ОГРАНИЧЕНИЙ СОВРЕМЕННЫХ ИНСТРУМЕНТОВ ВИЗУАЛИЗАЦИИ ДАННЫХ: НОВАЯ МЕТОДОЛОГИЯ

***А.Д. Джейранян***
*Студент 4 года обучения образовательной программы «Программная инженерия» НИУ ВШЭ-Пермь*
addzheyranyan@edu.hse.ru

***Л.Н. Лядова***
*Доцент кафедры информационных технологий в бизнесе НИУ ВШЭ-Пермь, кандидат физико-математических наук, доцент*
LNLyadova@gmail.com

***А.Д. Джейранян***
*Студент 4 курса бакалавриата Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ–Пермь), образовательная программа «Программная инженерия». Область научных интересов связана с анализом и визуализацией данных, искусственным интеллектом, генеративными моделями, предметно-ориентированным моделированием.*

***Л.Н. Лядова***
*Кандидат физико-математических наук, доцент, доцент кафедры информационных технологий в бизнесе Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ–Пермь). Область научных интересов включает языки моделирования, предметно-ориентированное моделирование, языковые инструментарии, CASE-средства, системы имитационного моделирования.*

**Аннотация.** Выполнен анализ существующих методов и инструментов визуализации данных, выявлены их ограничения, связанные с необходимостью продвинутых навыков программирования или недостаточной гибкостью настройки. Предложена методология разработки инструментов визуализации данных на основе экспертных знаний, обеспечивающая возможность эффективного создания настраиваемых визуальных представлений. Методология включает три ключевых подхода: (1) разработку предметно-ориентированных языков (*DSLs*) для визуализации данных, (2) автоматизированную генерацию скриптов визуализации с использованием библиотек *Python* и больших языковых моделей (*LLMs*) и (3) расширение существующих библиотек визуализации пакетами с пользовательскими элементами. Ожидается, что предложенная методология повысит доступность и эффективность инструментов визуализации данных, снизив требования к программированию при сохранении высокой точности представления данных в различных предметных областях.

**Ключевые слова:** визуализация данных, многоаспектная онтология, искусственный интеллект, предметно-ориентированный язык, python, скрипт, библиотека.

**Введение.** Область визуализации данных вызывает возрастающий интерес у исследователей, что подтверждается увеличением числа публикаций по этой тематике в базе данных *Scopus* – с 3232 работ в 2015 году до 5724 в 2024, что соответствует росту на 77 %.

При принятии решений на основе анализа данных визуализация становится неотъемлемым компонентом, позволяющим понять сложные наборы информации, выявлять скрытые закономерности и делать обоснованные выводы. Такие инструменты

находят применение у широкого круга пользователей, включая аналитиков, учёных, руководителей предприятий, менеджеров, преподавателей, студентов и т. д.

Несмотря на возрастающее значение визуализации, многие пользователи сталкиваются с серьёзными трудностями при создании качественных и информативных представлений данных. Одной из главных проблем является распространение визуализаций низкого качества, вводящих в заблуждение и не отражающих важные аспекты данных, а порой даже искажающих исходную информацию [1]. Эта проблема широко обсуждается в онлайн-сообществах, таких как *Data is Ugly* (www.reddit.com/r/dataisugly/), где участники демонстрируют и критикуют примеры неудачных визуальных решений – от неверных подписей осей и неправильного масштабирования до избыточного использования графических элементов, которые скорее затрудняют восприятие, чем способствуют его улучшению.

Для преодоления ограничений в передаче информации о визуализируемых объектах и процессах необходимо внедрение пользовательских спецификаций визуализации [2]. Они дают возможность адаптировать диаграммы под конкретные задачи и особенности предметной области, обеспечивая более эффективное представление данных.

Существующие инструменты визуализации отличаются степенью кастомизации и уровнем реализации спецификаций. В большинстве случаев предлагается лишь базовая настройка параметров, что не всегда удовлетворяет потребности пользователей. Некоторые сценарии требуют более глубоких модификаций, включая возможность создания собственных типов диаграмм. Так, например, на рисунке 1 показаны два случая, характерных для баскетбольной статистики: первая диаграмма адаптирует турнирные деревья, а вторая иллюстрирует эффективность бросков с разных участков площадки.

а) б)

*Рисунок 1.* а – Адаптивная визуализация дерева для турнирных сеток [3]. б – Визуализация баскетбольного броска [4]

Такая глубокая настройка зачастую требует использования языков программирования. Это создает барьеры для пользователей без технической подготовки и подчёркивает необходимость *low-code* решений.

Кроме того, многие исследователи отмечают сложность создания визуализаций: традиционные методы часто требуют глубоких знаний в области программирования, что представляет значительные трудности для нетехнических специалистов [5]. Аналогичным образом, в исследовании [2] подчеркивается, что даже для опытных специалистов разработка эффективных визуальных представлений остается трудоемкой задачей.

В связи с этим возрастает потребность в новом подходе, позволяющем адаптировать диаграммы под специфику предметных областей и решаемых задач без высоких требований к навыкам программирования.

В данном исследовании сформулированы следующие исследовательские вопросы:

– (*RQ1*) какие типы диаграмм достаточны для передачи большинства идей визуализации?

– (*RQ2*) какие сильные и слабые стороны присущи существующих средствам визуализации данных?

– (*RQ3*) какие методы могут быть интегрированы для снижения требований к навыкам программирования конечных пользователей?

Цель исследования заключается в разработке собственной методологии визуализации данных, способной преодолеть ограничения современных средств, а именно, позволяющей гибко настраивать визуализацию данных под специфику решаемых задач и предметных областей, при этом минимизируя требования к навыкам программирования пользователей. Настройка будет осуществляться на основе экспертных знаний, аккумулированных в базе знаний системы – многоаспектной онтологии.

**Классификация методов визуализации.** В следующих разделах будет представлен анализ существующих типов диаграмм и инструментов визуализации данных с целью устранения существующего разрыва между пользователями без технических навыков и процессом создания настраиваемых визуализаций.

По мере развития области визуализации данных появляются новые типы диаграмм. На сегодняшний день общее количество типов диаграмм превышает 60. В связи с расширением их числа исследователи разделились на две группы.

Одна группа исследователей [6] считает, что для иллюстрации количественных данных в 90% случаев достаточно 5 типов диаграмм: круговая диаграмма, линейчатая диаграмма гистограмма, линейный график и точечная диаграмма. А остальные 10% – это использование их комбинаций, таких как линейный график с гистограммой и т. д.

В противоположность этому, другая группа исследователей считает, что традиционные типы диаграмм могут быть недостаточными для эффективной визуализации данных. Соответственно, они разрабатывают новые методы визуализации и исследуют инновационные подходы к классификации диаграмм. Среди широко известных подходов к классификации можно выделить следующие: по типу данных, по форме представления [7], по цели создания [8-9] , по интерактивности [7] и др.

Заметна разрозненность мнений (*RQ1*). Однако остаётся фактом, что разнообразие применяемых методов велико, и пользователи постоянно сталкиваются с задачами, требующими разработки новых техник визуализации данных.

**Сравнительный анализ подходов к визуализации данных.** Ряд исследований [9] показал, что инструменты визуализации данных играют решающую роль в улучшении анализа и представления данных. Перед оценкой различных инструментов необходимо установить критерии сравнения:

1 Гибкость в адаптации диаграмм к конкретным предметным областям и задачам.

2 Доступность для пользователей без знаний программирования.

Далее рассматриваются сильные и слабые стороны пяти групп средств визуализации (*RQ2*): электронные таблицы, *BI*-платформы, языки программирования, предметно-ориентированные языки и большие языковые модели.

Электронные таблицы, вероятно, являются наиболее распространенным инструментом для вычислений и анализа данных в настоящее время, поскольку они стали первыми инструментами, используемыми пользователями микрокомпьютеров. Электронные таблицы, особенно такие инструменты, как *Excel* и *Google Sheets*, широко применяются для визуализации данных благодаря их доступности, простоте использования и широкому набору стандартных типов диаграмм. Эти инструменты предлагают базовые возможности настройки, включая изменение осей, модификацию заголовков, изменение цветов и добавление подписей данных, что делает их удобными для выполнения рутинных аналитических задач [10]. Однако электронные таблицы имеют существенные ограничения, такие как базовая настройка параметров и невозможность создания новых типов диаграмм,

что снижает их эффективность при проведении анализа данных. Для устранения этих недостатков пользователи могут использовать надстройки или макросы, такие как *RExcel*, который интегрирует *Excel* с языком программирования *R*, расширяя его функциональные возможности. Однако такие решения требуют навыков программирования, что создает барьер для пользователей без технической подготовки.

Следующим традиционным подходом является использование *BI*-платформ. Лидерами в этой области, согласно *2024 Gartner Magic Quadrant for Analytics and Business Intelligence*, являются *Power BI, Tableau, Qlik* и др. Эти платформы предлагают удобные интерфейсы и интеграцию с различными источниками данных, что делает их доступными для пользователей без знаний программирования [11]. *Power BI*, например, широко используется благодаря бесшовной интеграции с инструментами *Microsoft*, но его возможности настройки для узкоспециализированных визуализаций остаются ограниченными [12-13]. Аналогично, *Tableau* предоставляет интуитивно понятные функции обработки данных, однако для создания сложных визуализаций часто требуется знание R, а некоторые параметры настройки ограничены предустановленными форматами [8].

В отличие от рассмотренных ранее двух групп, языки программирования предоставляют широкие возможности для создания эффективных и настраиваемых визуальных представлений данных. Наиболее популярными языками в этой области являются *Python, R* и *JavaScript*, каждый из которых обладает богатой системой библиотек для визуализации [9]. Кроме того, наличие открытого исходного кода позволяет пользователям разрабатывать собственные библиотеки и расширять существующие инструменты. Например, в среде R был создан специализированный пакет для интегративного пан-ракового анализа и визуализации мультиомных данных *TCGA* [14]. Несмотря на то, что языки программирования обеспечивают максимальную гибкость в настройке визуализации, их использование требует навыков кодирования. Специализированные библиотеки снижают это требование за счет предопределенных функций, однако базовое понимание программирования остается необходимым.

Предметно-ориентированные визуальные языки (*DSLs*) значительно упрощают разработку, позволяя пользователям создавать модели с помощью *drag-and-dro*p вместо написания кода. Существующие визуальные *DSLs* [15-16] преимущественно ориентированы либо на стандартные диаграммы (например, круговые диаграммы, гистограммы), либо на специализированные типы данных, такие как геопространственные данные. В то же время DSLs на основе *JSON* демонстрируют сложность нахождения баланса между удобством использования и выразительностью, подчеркивая противоречие между формальными и неформальными моделями [17]. Однако возможности адаптации DSL под специфические области остаются ограниченными, зачастую требуя профессиональных навыков программирования. Недавние разработки, такие как *ModelViz*, показывают, как *DSLs* могут интегрировать предметные знания в визуализацию, упрощая создание дашбордов для непрофессионалов при сохранении ориентации на бизнес-цели [18]. Эти достижения свидетельствуют о тенденции к более удобным, основанным на моделях подходам в разработке инструментов визуализации данных.

Применение искусственного интеллекта в визуализации данных охватывает автоматическую генерацию, анализ информации и ее графическое представление [19-21]. Современные технологии используют большие языковые модели (*LLMs*), такие как *GPT-3* и *ChatGPT*, для генерации визуализаций на основе естественного языка, что повышает точность по сравнению с традиционными NLP-интерфейсами [22]. Например, система *Chat2VIS* демонстрирует, что эффективная промпт-инженерия может улучшить точность интерпретации визуализаций, сохраняя при этом безопасность данных. Однако сами разработчики системы отмечают, что генерируемые визуализации часто недостаточно гибки для сложных аналитических задач и плохо учитывают пользовательский замысел,

особенно при неоднозначных запросах [22]. Кроме того, у языковых моделей возникают сложности с креативным дизайном визуализаций, что подчеркивает необходимость лучшей интеграции человеческого опыта в машинный процесс для создания осмысленных и контекстно значимых решений.

Таблица 1 представляет сравнительную оценку всех пяти групп инструментов на основе ранее определенных критериев. Оценки выставлены по шкале, используемой в олимпиадном стиле: --, -+, +-, ++.

Таблица 1. Сравнение подходов к визуализации данных

| Подход | Гибкость настройки диаграмм на специфику решаемых задач и предметных областей | Доступность для пользователей без знаний программирования |
|---|---|---|
| Электронные таблицы | -+ | ++ |
| *BI*-платформы | +- | ++ |
| Языки программирования | ++ | -- |
| *DSLs* | ++ | +- |
| *LLMs* | +- | +- |

Оценка существующих инструментов визуализации данных подчеркивает ключевой разрыв: ни один метод не обеспечивает полного баланса между глубиной кастомизации и доступностью:

– электронные таблицы просты в использовании, но ограничены в возможностях настройки;

– *BI*-платформы предлагают большую гибкость, но по-прежнему имеют ограничения в создании новых типов диаграмм;

– языки программирования обеспечивают максимальную настройку, но требуют навыков программирования.

– предметно-ориентированные языки балансируют между гибкостью и доступностью, но являются сложными для разработки;

– большие языковые модели могут генерировать код для визуализации, но не способны самостоятельно разрабатывать принципиально новые типы диаграмм. Их эффективное использование требует умения формулировать точные текстовые запросы.

**Гибридные подходы: интеграция искусственного интеллекта с существующими инструментами.** Для преодоления существующих ограничений применяются гибридные подходы, объединяющие искусственный интеллект с уже используемыми инструментами (*RQ3*). В частности, электронные таблицы используют большие языковые модели (*LLMs*) для автоматической генерации статичных визуализаций. Например, *Google Sheets* уже интегрировал *Gemini* для создания графиков по пользовательским запросам, однако полученные изображения статичные и неизменяемые. Аналогично, интегрированные среды разработки (*IDE*) внедряют искусственный интеллект для упрощения кодирования пользователями, не имеющими опыта программирования. Подход, основанный на предметно-ориентированном моделировании (*DSM*), в сочетании с автоматизацией на основе генеративных моделей (как показано в исследовании [23]), способен снизить трудозатраты на разработку метамоделей языков.

Хотя развитие *LLMs* демонстрирует перспективные возможности, существующие решения все еще ограничены в плане кастомизации диаграмм, что требует дальнейших исследований для их более глубокой интеграции в процессы визуализации данных.

**Предлагаемая методология**. Методология визуализации данных начинается с классификации передовых методов визуализации, основанной на анализе литературных

обзоров. Эта классификация представлена на рисунке 2, она послужит основой для разработки метамоделей базовых языков.

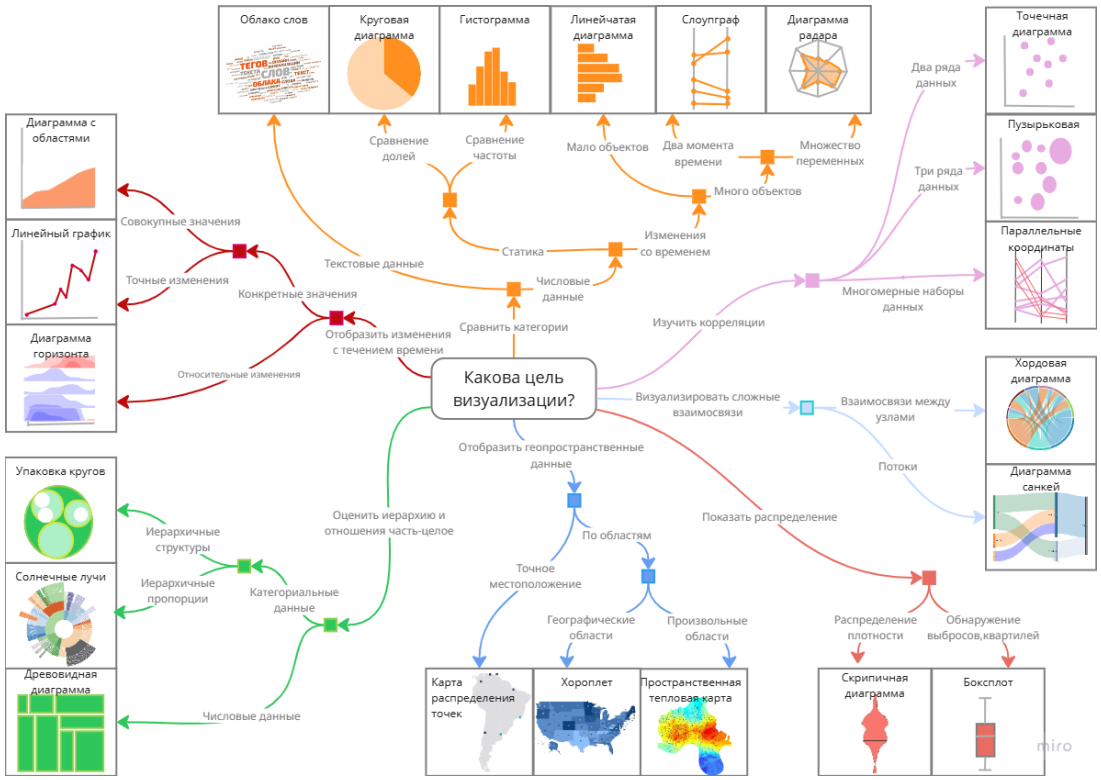



*Рисунок 2.* Классификация методов визуализации в зависимости от цели создания

Методология включает два типа задач: стандартные и нестандартные. Для стандартных задач применяется метод генерации кода на основе промтов в визуальной среде, что позволяет автоматически создавать и запускать скрипты. В случае нестандартных задач предполагается расширение существующих библиотек за счет пользовательских компонентов. Языково-ориентированный подход объединяет оба типа задач, обеспечивая гибкость и адаптивность. Общая схема методологии представлена на рисунке 3.

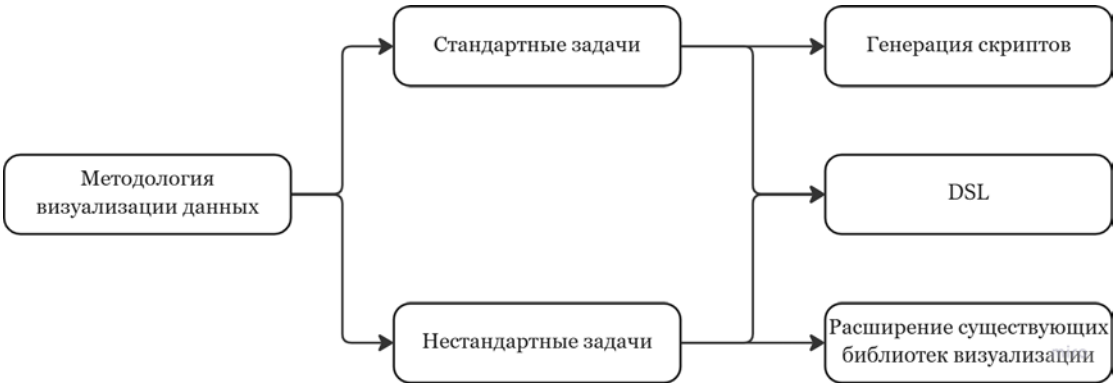



*Рисунок 3.* Схема методологии визуализации данных

**Принятые архитектурные решения.** Проектирование архитектуры системы основано на подходе, ориентированном на знания, где ядром платформы выступает база знаний. В отличие от традиционных баз данных, в качестве основы выбрана многоаспектная онтология, что обеспечивает гибкость в структурировании разнородных данных.

Онтология играет ключевую роль в хранении информации, необходимой для функционирования платформы. Она включает три основные группы онтологий:

– онтологии знаний *DSM* – содержат описание моделей, визуальных языков, предметных областей, правил трансформации и генерации *DSL* и т. д.;

– репозиторий пользовательских задач – включает онтологии пользовательских задач и применяемых методов, обеспечивая возможность многоразового использования создаваемых моделей;

– онтологии сгенерированных скриптов – охватывают онтологии промтов, результатов генерации и метаданных, что позволяет автоматизировать процесс создания и уточнения визуализаций.

Кроме того, архитектура должна интегрировать различные методы и алгоритмы, устанавливать правила *API* и обеспечивать совместимость с *DSM*-платформами.

Далее рассмотрим методы, включенные в предлагаемую методологию.

**Генерация скриптов**. Данный метод позволяет пользователям формулировать задачи визуализации на естественном языке, после чего система автоматически создает и выполняет *Python*-скрипты. Это значительно упрощает процесс работы с данными, снижая необходимость в глубоких знаниях программирования. В качестве примера можно рассмотреть генерацию визуализаций из *CSV*-файлов. В исследовании [5] была разработана платформа, которая интегрирует библиотеку *Pandas* с генеративным искусственным интеллектом и методами обработки естественного языка (NLP). В основе решения лежит модель *Generative AI*, основанная на архитектуре *GPT-3*, которая обучена на обширном корпусе материалов по анализу данных и их визуализации.

**Предметно-ориентированные языки.** Использование *DSL* реализуется двумя способами:

1 Сборка моделей на основе базовых языков – пользователи могут работать с готовыми *DSLs*, настраивая их параметры.

2 Создание новых предметно-ориентированных языков – разработка автоматизируется за счет отображения онтологии предметной области на метамодель базового языка. Данный подход подробно рассматривается в [24-26].

Рассмотрим пример разработки нового DSL для решения прикладной задачи – исследования поведения посетителей кафе в зависимости от их предпочтений в напитках и уровня шума в помещении. Предметная область: кафе, специализирующееся на подаче кофе различных сортов и других напитков. Данные собираются с помощью опросов и сенсоров, фиксирующих уровень шума, предпочтения посетителей и продолжительность их пребывания.

Онтология предметной области (объект – напитки) отображается на метамодель пузырьковой диаграммы. Результатом отображения является новый язык, метамодель которого имеет вид, представленный на рисунке 4.

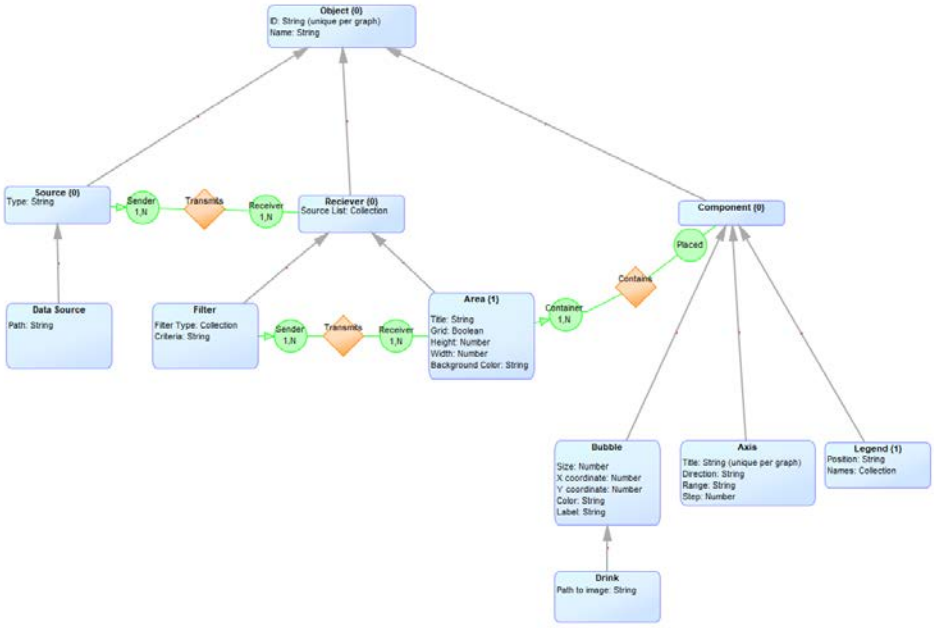



*Рисунок 4.* Метамодель языка Café Bubble Plot

Далее пользователь может создать модель на основе нового *DSL*, как представлено на рисунке 5.

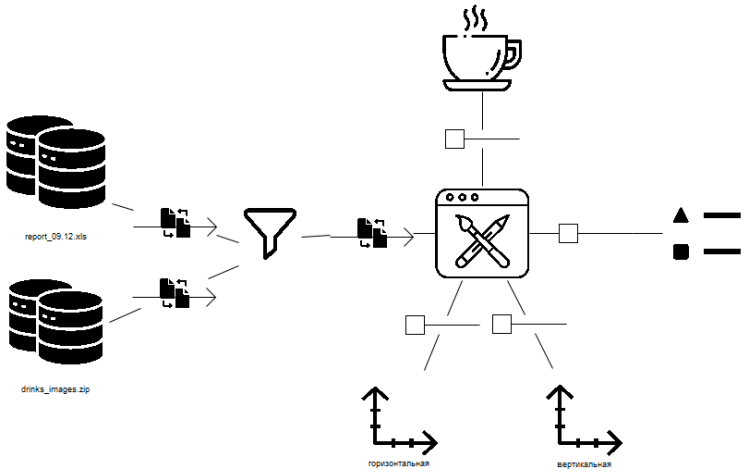



*Рисунок 5.* Модель, построенная на основе языка Café Bubble Plot

Для реализации визуализации необходимо применить к модели правила трансформации (правила генерации кода), хранящиеся в онтологии. Результирующая диаграмма показывает зависимость предпочтений посетителей от уровня шума (см. рисунок 6):

– по оси X откладывается уровень шума (в децибелах);

– по оси Y – среднее время пребывания в кафе (в минутах);

– вид эмодзи зависит от вида выбранной позиции (эспрессо, латте, капучино и т. д.);

– размер эмодзи пропорционален количеству заказов данного типа кофе.

Из диаграммы можно быстро выявить, что, например, в условиях высокого уровня шума популярны более крепкие напитки, такие как эспрессо и американо, а при среднем уровне шума посетители предпочитают задерживаться дольше, выбирая латте или капучино.

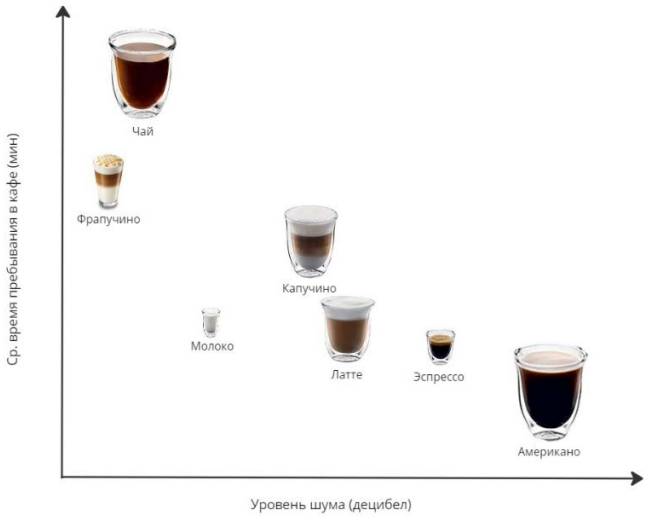



*Рисунок 6.* Диаграмма чашек

**Расширение библиотек визуализации**. Данный подход предлагает расширить существующую библиотеку визуализации, например, *Matplotlib* или *Plotly*, разработав пользовательские пакеты с новыми элементами и типами диаграмм. Процесс включает четыре этапа: создание, регистрация, загрузка и использование пакета.

Создание новых пакетов уже упрощает требования к навыкам программирования. Дополнительно в настоящее время изучается возможность предоставления визуальной среды – конструктора диаграмм, позволяющего пользователям собирать визуализации из отдельных элементов без написания кода.

**Направления дальнейшего исследования.** Дальнейшее развитие работы будет сосредоточено на проектировании архитектуры системы и реализации отдельных прототипов, ориентированных на автоматическую генерацию скриптов и расширение библиотек визуализации. С учетом перспектив расширения методологии в будущем необходимо создать гибкую архитектуру, позволяющую добавлять новые компоненты для реализации дополнительных методов, а также определить четкие требования к интеграции, обеспечивающие возможность разработки и вызова пользовательских алгоритмов.

Для оценки качества создаваемых визуализаций будут применяться установленные качественные метрики [27-28], связанные с когнитивными процессами. Среди метрик можно выделить: читаемость (ясность подписей и структуры), интерпретируемость (удобство извлечения выводов) и сопоставимость (эффективность сравнительных техник).

Кроме того, планируется исследование потенциала применения методов искусственного интеллекта для автоматизации построения предметно-ориентированных языков.

**Заключение.** В ходе исследования был проведен анализ существующих методов и инструментов автоматической генерации визуализаций. На основе этого анализа была предложена новая методология, включающая классификацию методов визуализации и три ключевых подхода: автоматическую генерацию скриптов на основе текстовых запросов к языковым моделям, разработку предметно-ориентированных языков и расширение существующих библиотек языков программирования пользовательскими элементами. Использование многоаспектной онтологии в качестве ядра системы позволяет хранить все знания, необходимые для функционирования платформы, и впоследствии интегрироваться с внешними аналитическими платформами, управляемыми знаниями (например, [29]). Результаты реализации подхода автоматизации создания DSL посредством отображения онтологии предметной области на метамодель базового языка прошли апробацию и были опубликованы в ряде статей [24-26]. Предложенная методология визуализации данных позволяет повысить гибкость и удобство инструментов визуализации, обеспечивая

возможность создания кастомизированных визуализаций без необходимости глубоких знаний в программировании.

## Список литературы

[1] Sawicki J., Burdukiewicz M. VisQualdex: a Comprehensive Guide to Good Data Visualization. Scientific Visualization. 2022;15(1):127–149. DOI: 10.26583/sv.15.1.11.

[2] Qin X., Luo Y., Tang N., Li G. Making Data Visualization More Efficient and Effective: a Survey. The VLDB Journal. 2020;29(1):93–117. DOI: https://doi.org/10.1007/s00778-019-00588-3.

[3] Tan D., Smith G., Lee B., Robertson G. AdaptiviTree: Adaptive Tree Visualization for Tournament-Style Brackets. IEEE Transactions on Visualization and Computer Graphics. 2007;13(6):1113–1120. DOI: 10.1109/TVCG.2007.70537.

[4] Beshai P. Buckets: Basketball shot visualization. 2014.

[5] Singh A., Goyal C., Parashar J. M. Democratizing Data Visualization and Insights Extraction with Pandas, Generative AI, and CSV Data. International Journal of Scientific Research in Engineering and Management. 2024;8(5):1–5.

[6] Zelazny G. Say It With Charts Complete toolkit. McGraw Hill Professional. 2006.

[7] Cuadrado-Gallego J. J., Demchenko Y., Losada M. A., Ormandjieva O. Classification and Analysis of Techniques and Tools for Data Visualization Teaching. 2021 IEEE Global Engineering Education Conference (EDUCON). 2021;1593–1599. DOI: 10.1109/EDUCON46332.2021.9453917.

[8] Cota M. P., Rodríguez M. D., González-Castro M. R., Gonçalves R. M. M. Analysis of Current Visualization Techniques and Main Challenges for the Future. Journal of Information Systems Engineering & Management. 2017;2(3). DOI: 10.20897/jisem.201719.

[9] Shakeel H. M., Iram S., Al-Aqrabi H., Alsboui T., Hill R. A Comprehensive State-of-the-Art Survey on Data Visualization Tools: Research Developments, Challenges and Future Domain Specific Visualization Framework. IEEE Access. 2022;10:96581–96601. DOI: 10.1109/ACCESS.2022.3205115.

[10] Mendez-Carbajo D., Dellachiesa A. Choice of Data Visualization Tool: FRED or Spreadsheets? International Review of Economics Education. 2023;44:100275. DOI: https://doi.org/10.1016/j.iree.2023.100275.

[11] Khatuwal V. S., Puri D. Business Intelligence Tools for Dashboard Development. 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM). 2022;128–131. DOI: 10.1109/ICIEM54221.2022.9853086.

[12] Balasubramanian V. R., Vashishtha S., Yadav N. Integrating SAP Analytics Cloud and Power BI: Comparative Analysis for Business Intelligence in Large Enterprises. International Journal of Multidisciplinary Innovation and Research Methodology. 2024;3(4):111–140.

[13] Nagy E., Ferenc G. T., Molnár G. Comparison of Data Visualization Platforms through European Traffic Data. 2024 IEEE 11th International Conference on Computational Cybernetics and Cyber-Medical Systems (ICCC). 2024;000043–000048. DOI: https://doi.org/10.1109/iccc62278.2024.10582968.

[14] Liao C., Wang X. TCGAplot: an R Package for Integrative Pan-Cancer Analysis and Visualization of TCGA Multi-Omics Data. BMC Bioinformatics. 2023;24(1). DOI: https://doi.org/10.1186/s12859-023-05615-3.

[15] Ledur C., Griebler D., Manssour I., Fernandes L. G. A High-Level DSL for Geospatial Visualizations with Multi-core Parallelism Support. IEEE 41st Annual Computer Software and Applications Conference (COMPSAC). 2017;298–304. DOI: 10.1109/COMPSAC.2017.18.

[16] Smeltzer K., Erwig M. A Domain-Specific Language for Exploratory Data Visualization. 17th ACM SIGPLAN International Conference on Generative Programming: Concepts and Experiences. 2018;1–13. DOI: 10.1145/3278122.3278138.

[17] McNutt A. M. No grammar to rule them all: A survey of JSON-style DSLs for visualization. IEEE Transactions on Visualization and Computer Graphics. 2023;29(1):160–170. DOI: https://doi.org/10.1109/tvcg.2022.3209460.

[18] Khakpour A., Vázquez-Ingelmo A., Colomo-Palacios R., García-Peñalvo F. J., Martini A. ModelViZ: A Model-Driven Engineering Approach for Visual Analytics System Design. IEEE Access. 2024;12:42667–42682. DOI: https://doi.org/10.1109/access.2024.3379268.

[19] Shi D., Xu X., Sun F., Shi Y., Cao N. Calliope: Automatic Visual Data Story Generation from a Spreadsheet. IEEE Transactions on Visualization and Computer Graphics. 2022;27(2):453–463. DOI: 10.1109/TVCG.2020.3030403.

[20] Wang X., Wu Z., Huang W., Wei Y., Huang Z., Xu M., Chen W. VIS+AI: Integrating Visualization with Artificial Intelligence for Efficient Data Analysis. Frontiers of Computer Science. 2023;17(6). DOI: https://doi.org/10.1007/s11704-023-2691-y.

[21] Wu A., Wang Y., Shu X., Moritz D., Cui W., Zhang H., Zhang D., Qu H. AI4VIS: Survey on Artificial Intelligence Approaches for Data Visualization. IEEE Transactions on Visualization and Computer Graphics. 2021;28(12):5049–5070. DOI: https://doi.org/10.1109/tvcg.2021.3099002.

[22] Maddigan P., Susnjak T. Chat2VIS: Generating Data Visualizations via Natural Language Using ChatGPT, CodeX and GPT-3 Large Language Models. IEEE Access. 2023;11:45181–45193. DOI: https://doi.org/10.1109/access.2023.3274199.

[23] Luo Y., Yang Y. Large Language Model and Domain-Specific Model Collaboration for Smart Education. Frontiers of Information Technology & Electronic Engineering. 2024;25(3):333–341. DOI: https://doi.org/10.1631/fitee2300747.

[24] Dzheiranian A. D., Ermakov I. D., Proskuryakov K. A., Lyadova L. N. Designing Data Visualization System Based on Language-Oriented Approach. Trudy ISP RAN/Proc. ISP RAS. 2024;36(2):127-140. DOI: 10.15514/ISPRAS-2024-36(2)-10.

[25] Джейранян А. Д., Ермаков И. Д., Проскуряков К. А., Лядова Л. Н. Разработка инструментов визуализации данных на основе предметно-ориентированного моделирования. GraphiCon 2024: Материалы 34-й Международной конференции по компьютерной графике и машинному зрению. 2024;300-314. DOI: 10.25206/978-5-8149-3873-2-2024-300-314.

[26] Dzheiranian A. D., Ermakov I. D., Proskuryakov K. A., Lyadova L. N. An Approach to Developing Data Visualization Tools Based on Domain Specific Modeling. Scientific Visualizationю 2024;16(4):82–101. DOI: 10.26583/sv.16.4.08.

[27] Behrisch M., Blumenschein M., Kim N. W., Shao L., El-Assady M., Fuchs J., Seebacher D., Diehl A., Brandes U., Pfister H., Schreck T., Weiskopf D., Keim D. A. Quality Metrics for Information Visualization. Computer Graphics Forum. 2018;37(3):625–662. DOI: 10.1111/cgf.13446.

[28] Midway S. R., Brum J. R., Robertson M. Show and Tell: Approaches for Effective Figures. Limnology and Oceanography Letters. 2022;8(2):213–219. DOI: https://doi.org/10.1002/lol2.10288.

[29] Zayakin V. S., Lyadova L. N., Rabchevskiy E. A. Design Patterns for a Knowledge-Driven Analytical Platform. Proceedings of the Institute for System Programming of RAS (Proceedings of ISP RAS). 2022;34(2):43–56. DOI: https://doi.org/10.15514/ISPRAS-2022-34(2)-4.

### Авторский вклад

**Джейранян Анна Даниеловна** – анализ существующих методов и инструментов визуализации данных, разработка методологии визуализации данных, классификация методов визуализации в зависимости от цели создания, описание кейса для анализа поведения посетителей кафе.

**Лядова Людмила Николаевна** – постановка задачи исследования, руководство исследованием.

## OVERCOMING THE LIMITATIONS OF MODERN DATA VISUALIZATION TOOLS: A NOVEL METHODOLOGY

***A.D. Dzheiranian***
*Undergraduate Student of Educational Program "Software Engineering" NRU HSE-Perm*

***L.N. Lyadova***
*Associate Professor of the Department of Information Technologies in Business of NRU HSE-Perm, Cand. Sci. (Phys.-Math.), Associate Professor*

**Abstract.** An analysis of existing data visualization methods and tools was conducted, revealing their limitations associated with the need for advanced programming skills or insufficient configuration flexibility. A methodology for developing data visualization tools based on expert knowledge is proposed, enabling the effective creation of customizable visual representations. The methodology comprises three key approaches: (1) the development of Domain-Specific Languages (DSLs) for data visualization; (2) the automated generation of visualization scripts using Python libraries and large language models (LLMs); and (3) the extension of existing visualization libraries with packages that incorporate custom elements. It is expected that the proposed methodology will improve the accessibility and efficiency of data visualization tools by reducing programming requirements while maintaining high accuracy in data representation across various domains.

**Keywords:** data visualization, multifaceted ontology, artificial intelligence, domain-specific language, python, script, library.

УДК 004.622:004.654

# ВЫДЕЛЕНИЕ И ПОИСК ФУНКЦИОНАЛЬНЫХ ЗАВИСИМОСТЕЙ МЕЖДУ АТРИБУТАМИ В СИСТЕМАХ BIG DATA

***А.А. Карпук***
*Профессор кафедры программного обеспечения сетей телекоммуникаций Белорусской государственной академии связи, кандидат технических наук, доцент*
a_karpuk@mail.ru

***Л.С. Лазута***
*Аспирант кафедры телекоммуникационных систем Белорусской государственной академии связи, магистр*
lenya.lazuta@mail.ru

***А.А. Карпук***

*Окончил Белорусский государственный университет, автор более 240 научных трудов, включая 2 единолично написанные монографии. Область научных интересов связана с моделированием и оптимизацией сложных систем, проектированием баз данных и хранилищ данных, оценкой качества радиосвязи и оптимизацией присвоения радиочастот в радиосетях и радиолиниях.*

***Л.С. Лазута***

*Окончил Белорусскую государственную академию связи. Область научных интересов связана с моделированием и оптимизацией сложных систем, проектированием баз данных и хранилищ данных.*

**Аннотация.** Приведен обзор методов выделения и алгоритмов поиска функциональных зависимостей между атрибутами в системах Big Data. Рассмотрены основные алгоритмы поиска функциональных зависимостей на основе анализа решетки атрибутов, на основе анализа строк таблицы и гибридные алгоритмы. Определены области применения алгоритмов каждого типа и направления дальнейших исследований по совершенствованию и разработке алгоритмов поиска функциональных зависимостей между атрибутами в системах Big Data.

**Ключевые слова:** большие данные, функциональная зависимость, выделение функциональных зависимостей, алгоритмы поиска функциональных зависимостей.

**Введение.** Современные системы Big Data могут иметь архитектуру реляционного хранилища данных (Relational Data Warehouse, RDW), озера данных (Lake Data), современного хранилища данных (Modern Data Warehouse, MDW), фабрики данных (Data Fabric), озерного хранилища данных (Data Lakehouse), сетки данных (Data Mesh) [1]. Во всех перечисленных архитектурах, кроме архитектуры Lake Data, на одном или нескольких этапах обработки данных решаются задачи очистки данных, предварительной обработки данных и нормализации данных в виде приведения данных к третьей нормальной форме. Для решения задачи нормализации данных требуется знание системы образующих структуры функциональных зависимостей (ФЗ) между данными. В настоящей работе рассматриваются методы выделения и алгоритмы поиска ФЗ между данными в системах Big Data.

**Методы выделения функциональных зависимостей между данными**. Пусть $X \subset A$ – подмножество атрибутов (признаков) системы Big Data, $Z \in A$ – некоторый атрибут. Говорят, что существует ФЗ $X \to Z$, если любой комбинации значений атрибутов из $X$ всегда соответствует единственное значение атрибута $Z$. Очевидно, что из $Z \in X$ следует, что $X \to Z$. Такая ФЗ, в которой зависимый атрибут входит в состав левой части ФЗ, называется тривиальной. В дальнейшем будем рассматривать только нетривиальные ФЗ между атрибутами. Структура ФЗ на множестве атрибутов удовлетворяет аксиомам У. Армстронга [2]. В работах по проектированию баз данных 80-х годов прошлого века структура ФЗ на множестве атрибутов предметной области задавалась путем постулирования некоторого множества ФЗ, которое называется системой образующих структуры ФЗ. П. Чен предложил описывать предметную область при проектировании баз данных в виде диаграммы «сущность-связь» (ER-диаграммы) [3]. Из ER-диаграммы можно выделить и включить в систему образующих структуры ФЗ зависимости всех атрибутов сущностей от первичных и потенциальных (альтернативных) ключей сущностей. Других нетривиальных ФЗ между атрибутами из ER-диаграммы получить нельзя. А.А. Карпук и В.М. Острейко предложили метод построения расширенной ER-диаграммы, отличающейся от диаграммы П. Чена тем, что сущности предметной области могут иметь сложную иерархическую структуру, и в диаграмме предусмотрено постулирование ФЗ и нефункциональных связей между атрибутами, существующих в отрыве от контекста [4]. Из расширенной ER-диаграммы можно дополнительно выделить и включить в систему образующих структуры ФЗ зависимости не ключевых атрибутов сложных объектов от сцепленных ключей и ФЗ между атрибутами, существующие в отрыве от контекста.

Исследования по поиску ФЗ между атрибутами возобновились в начале 21-го века в связи с появлением онтологического подхода к описанию предметной области и развитием систем Big Data. А.А. Карпук разработал метод построения системы образующих структуры ФЗ между атрибутами предметной области на основе онтологии предметной области, описанной на языке OWL-2 [5]. Этот метод выделяет ФЗ между атрибутами, входящими в один класс, входящими в класс и его подкласс, входящими в функциональные отношения между классами, входящими в определение функций, заданных на атрибутах и классах онтологии предметной области.

**Алгоритмы поиска функциональных зависимостей между данными**. Ряд авторов предложили алгоритмы поиска ФЗ между атрибутами в таблицах систем Big Data, В общем случае эта задача является NP-трудной, поскольку для таблицы, содержащей $m$ атрибутов и $n$ строк, вычислительная сложность задачи оценивается величиной $O(2^{m-2}n^2m^2)$ [6]. Существующие алгоритмы поиска ФЗ между атрибутами в таблицах можно разделить на 3 класса: алгоритмы на основе анализа решетки атрибутов, алгоритмы на основе анализа строк таблицы и гибридные алгоритмы.

Решеткой атрибутов для таблицы, содержащей $m$ атрибутов, называется ориентированный граф, состоящий из $m+1$ уровней. На уровне 0 расположена одна вершина графа, представляющая пустое множество атрибутов. На уровне 1 расположено $m$ вершин, каждая из которых представляет 1 атрибут таблицы. На уровне 2 расположено $m(m-1)/2$ вершин, каждая из которых представляет сочетание двух атрибутов таблицы. В общем виде, на уровне $1 \le k \le m$ расположено $m!/k!(m-k)!$ вершин, каждая из которых представляет сочетание из $k$ атрибутов таблицы. На уровне $m$ расположена одна вершина, представляющая все множество атрибутов таблицы. В решетке атрибутов существует направленное ребро между вершинами соседних уровней от меньшего уровня к большему уровню, если все атрибуты вершины из меньшего уровня входят в состав атрибутов

вершины из большего уровня. Каждому ребру решетки атрибутов соответствует возможная ФЗ, в левой части которой содержатся атрибуты вершины меньшего уровня, а в правой части - атрибут, полученный удалением из состава атрибутов вершины большего уровня всех атрибутов вершины меньшего уровня. В левой части рисунка 1 показана решетка атрибутов для таблицы из четырех атрибутов A, B, C, D. Вершинам A из уровня 1 и AB из уровня 2 соответствует возможная ФЗ A→B, вершинам CD из уровня 2 и BCD из уровня 3 соответствует возможная ФЗ CD→B.

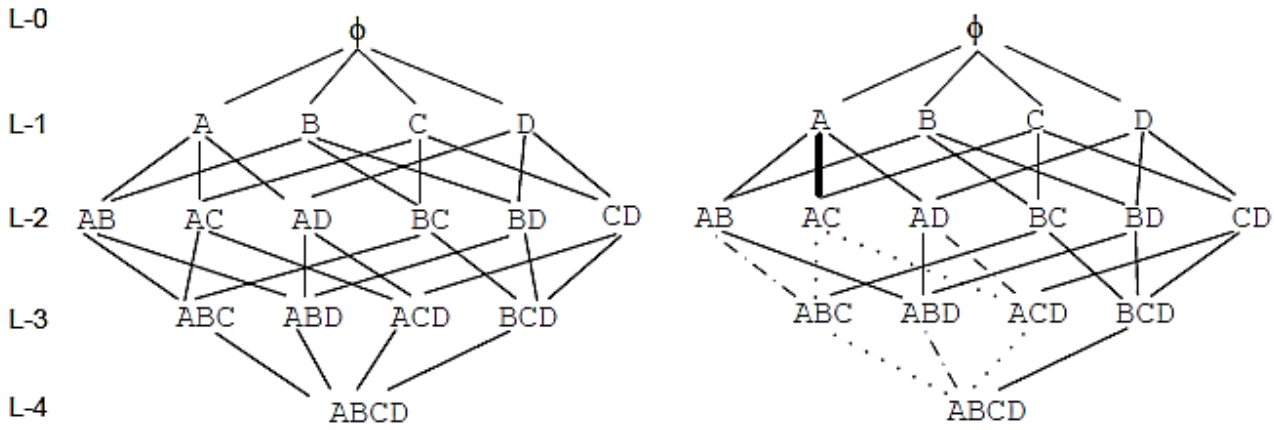



*Рисунок 1.* Решетка атрибутов для таблицы из четырех атрибутов

Идея алгоритмов поиска ФЗ между атрибутами в таблицах на основе анализа решетки атрибутов состоит в выделении возможной ФЗ, проверки возможной ФЗ и классификации ее на есть-ФЗ или нет-ФЗ, и сокращении области поиска на основе полученной классификации путем удаления некоторых возможных ФЗ или удаления некоторых вершин решетки. Если X и Y - 2 непересекающихся подмножества атрибутов, A и B - 2 атрибута, не принадлежащие X и Y, то в силу аксиом У. Армстронга справедливы следующие правила:

1) Если возможная ФЗ X→A классифицирована как есть-ФЗ, то все ФЗ XY→A тоже будут классифицированы как есть-ФЗ, и проверять их нет необходимости, а соответствующие ребра следует удалить из решетки атрибутов.

2) Если возможная ФЗ X→A классифицирована как есть-ФЗ, то вместо проверки возможной ФЗ XAY→B следует проверять возможную ФЗ XY→B, а соответствующие ребра следует удалить из решетки атрибутов.

3) Если X - ключ рассматриваемой таблицы, т. е. для любого атрибута Z, не принадлежащего X, возможная ФЗ X→A классифицируется как есть-ФЗ, то из решетки атрибутов следует удалить все вершины, в состав которых входят все атрибуты из X.

К примеру, пусть в решетке атрибутов, показанной на рисунке 1, возможная ФЗ A→C классифицирована как есть-ФЗ (соответствующее ребро выделено в правой части рисунка 1). Тогда то по правилу 1 из решетки надо удалить ребра между вершинами AB и ABC, AD и ACD, ABD и ABCD (это ФЗ AB→C, AD→C, ABD→C соответственно). По правилу 2 из решетки надо удалить ребра между вершинами AC и ABC, AC и ACD, ABC и ABCD, ACD и ABCD (это ФЗ AC→B, AC→D, ABC→D, ACD→B соответственно).

С использованием приведенных правил сокращения области поиска в решетке атрибутов работает алгоритм поиска ФЗ между атрибутами в таблицах систем Big Data, названный алгоритмом TANE [7]. В этом алгоритме обход решетки совмещен с ее

построением, начиная с уровня 1, сначала вправо, а затем вниз. За счет отсечения ребер решетки по правила 1 и 2, алгоритм находит минимальные ФЗ, не содержащие лишних атрибутов в левой части. Для классификации возможных ФЗ алгоритм TANE разбивает множество строк таблицы по классы эквивалентности с совпадающими значениями атрибутов левой части рассматриваемых возможных ФЗ, а затем проверяет значения атрибутов правых частей возможных ФЗ. Алгоритм FUN, предложенный в [8], отличается от алгоритма TANE тем, что для отсечения ребер решетки использует более жесткое правило. чем правило 2, сокращая тем самым объем вычислений.

В работе [9] предложен алгоритм поиска ФЗ между атрибутами в таблицах систем Big Data на основе анализа решетки атрибутов, названный алгоритмом FD_Mine. Этот алгоритм также обходит решетку атрибутов, начиная с уровня 1, сначала вправо, а затем вниз. Для классификации возможных ФЗ алгоритм FD_Mine использует такой же метод, как алгоритм TANE. В отличие от алгоритмов TANE и FUN, алгоритм FD_Mine использует дополнительное правило отсечения вершин и ребер решетки, основанное на классах эквивалентности подмножеств атрибутов. Два набора атрибутов считаются эквивалентными, если они функционально зависят друг от друга. После проверки каждого уровня в решетке атрибутов алгоритм FD_Mine сканирует этот уровень и обнаруженные ФЗ на предмет эквивалентности. Проверка эквивалентности производится путем построения замыкания каждого из подмножеств атрибутов относительно найденных ФЗ. Если эквивалентные подмножества атрибутов найдены, то алгоритм оставляет в решетке атрибуты одного из эквивалентных подмножеств, удаляя из решетки вершины, содержащие остальные эквивалентные подмножества, и соответствующие им ребра.

В работе [10] предложен алгоритм поиска ФЗ между атрибутами в таблицах систем Big Data на основе анализа решетки атрибутов, названный алгоритмом DFD. На первом шаге алгоритма проверяется каждый атрибут, не является ли он ключом таблицы. Если атрибут является ключом, то фиксируются ФЗ, левая часть которых состоит из этого атрибута, а в правых частях находятся все остальные атрибуты таблицы, и этот атрибут исключается из решетки. Каждый оставшийся атрибут рассматривается как возможная правая часть некоторых ФЗ. Для каждой возможной правой части ФЗ алгоритм формирует множество источников для левой части ФЗ и обходит источники, выбирая очередной источник по определенным правилам, которые в некоторых ситуациях предусматривают случайный выбор. Каждый источник (вершину решетки) и соответствующее ребро алгоритм классифицирует как зависимость, минимальную зависимость, кандидат на минимальную зависимость, не зависимость, максимальную не зависимость или кандидат на максимальную не зависимость.

Зависимость и не зависимость представляют вершины левой части, которые не являются ни минимальными, ни максимальными соответственно. Кандидат на минимальную зависимость или/максимальную не зависимость – это комбинация атрибутов, которая все еще может быть левой частью минимальной ФЗ или максимальной не-ФЗ. В случае, если текущая комбинация атрибутов уже была пройдена на более раннем этапе процесса, алгоритм DFD снова проверяет только те комбинации атрибутов, которые классифицированы как кандидаты на минимальную зависимость или максимальную не зависимость. При этом алгоритм проверяет, можно ли изменить классификацию кандидата или нет. Может случиться так, что после повторного посещения кандидата на минимальную зависимость все подмножества этой вершины были классифицированы как не зависимость, что делает кандидата минимальной зависимостью. Вот почему алгоритм DFD сохраняет посещенные вершины кандидаты в стеке трассировки. Трассировка позволяет алгоритму

откатываться по решетке, повторно посещая вершины, которые в конечном итоге могут быть классифицированы на более позднем этапе. В случае, если вершина еще не была посещена, алгоритм DFD проверяет, является ли вершина надмножеством или подмножеством ранее обнаруженной зависимости или не зависимости, и обновляет ее классификацию соответствующим образом. В противном случае алгоритм определяет, какой тип кандидата представляет комбинация вершин. Если обнаруживается ФЗ, то вершина классифицируется как кандидат на минимальную зависимость, в противном случае – как кандидат на максимальную не зависимость.

К алгоритмам поиска ФЗ между атрибутами в таблицах на основе анализа строк таблицы относятся алгоритмы FDEP [11], Dep-Miner [12], FastFDs [13] и FSC [14]. В работе [11] П.А. Флах и И. Савник предложили 3 варианта алгоритма FDEP: алгоритм анализа сверху вниз, алгоритм двунаправленного анализа и алгоритм анализа снизу вверх. При анализе сверху вниз алгоритм сначала для каждого атрибута A определяет минимальные подмножества атрибутов X, для которых ФЗ X→A выполняется в первых двух строках таблицы. Затем выполнение каждой найденной ФЗ X→A проверяется для остальных строк таблицы. Если для очередной строки таблицы ФЗ X→A не выполняется, то эта ФЗ удаляется из рассматриваемых возможных ФЗ. В результате для каждого атрибута будет получено множество минимальных ФЗ, которые выполняются для всех строк таблицы.

При двунаправленном анализе сначала выполняется анализ всех пар строк таблицы. Для каждой пары строк для каждого атрибута A алгоритм определяет максимальные подмножества атрибутов Y, для которых ФЗ Y→A не выполняется, т. е. находит максимальные нет-ФЗ. На основе найденных максимальных нет-ФЗ строится неизбыточное отрицательное покрытие ФЗ. Второй шаг алгоритма двунаправленного анализа похож на алгоритм анализа сверху вниз, и отличается от него тем, что выполнение каждой найденной по первым двум строкам ФЗ X→A проверяется не по остальным строкам таблицы, а по построенному не избыточному отрицательному покрытию ФЗ. На первом шаге алгоритма анализа снизу вверх также находятся максимальных нет-ФЗ и строится неизбыточное отрицательное покрытие ФЗ. На втором шаге алгоритма за одну итерацию на основе не избыточного отрицательного покрытия ФЗ строится положительное покрытие ФЗ, состоящее из минимальных ФЗ. Очевидно, что при достаточно большом количестве строк в таблице быстродействие алгоритма анализа снизу вверх будет выше, чем у алгоритма двунаправленного анализа.

В работе [12] С. Лопес и др. предложили алгоритм Dep-Miner, который выводит все минимальные ФЗ из наборов атрибутов, которые имеют одинаковые значения в определенных строках таблицы. Эти наборы атрибутов называются согласованными наборами, а наборы из оставшихся атрибутов называются дополнительными наборами. Алгоритм Dep-Miner состоит из пяти этапов. На этапе 1 алгоритм вычисляет урезанное разбиение таблицы для каждого атрибута, в котором каждый класс эквивалентности атрибута содержит более одной строки таблицы. Из урезанного разбиения таблицы строится множество максимальных классов эквивалентности для таблицы, которое состоит из тех классов эквивалентности для всех атрибутов, которые не являются подмножествами других классов эквивалентности. На этапе 2 алгоритма множество максимальных классов эквивалентности для таблицы используется для построения согласованных наборов атрибутов таблицы. Согласованный набор атрибутов таблицы состоит из одного или более атрибутов, которые имеют одинаковые значения в двух или более строках таблицы. На этапе 3 согласованные наборы атрибутов таблицы преобразуются в максимальные наборы, т. е. наборы атрибутов, которые не имеют

надмножества с одинаковыми значениями в двух строках таблицы. На этапе 4 алгоритм инвертирует максимальные согласованные наборы атрибутов в дополнительные наборы. На этапе 5 алгоритм вычисляет из дополнительных наборов атрибутов все минимальные ФЗ между атрибутами.

В работе [13] К. Вайс с соавторами предложили алгоритм FastFDs как усовершенствование алгоритма Dep-Miner. Алгоритм FastFDs также строится на использовании согласованных наборов атрибутов для вывода ФЗ между атрибутами. После вычисления согласованных наборов атрибутов алгоритм FastFDs следует другой стратегии для вывода минимальных ФЗ, поскольку максимизация согласованных наборов атрибутов на этапе 3 алгоритма Dep-Miner является дорогостоящей операцией, Алгоритм FastFDs вместо этого вычисляет все дополнительные наборы атрибутов непосредственно по согласованным наборам. На шаге 4 алгоритм для каждого атрибута вычисляет все минимальные дополнительные наборы атрибутов, на основе которых на шаге 5 алгоритм находит все минимальные ФЗ между атрибутами.

Одним из последних алгоритмов поиска ФЗ между атрибутами в таблицах на основе анализа строк таблицы является алгоритм FSC, предложенный С. Вань с соавторами [14] в 2024 г. Двухэтапное выполнение алгоритма FSC опирается на предварительно вычисленную вспомогательную структуру сопоставимых пар строк, которые отражают пары идентификаторов для строк, имеющих, по крайней мере, по одному атрибуту с одинаковыми значениями. На шаге 1 алгоритм FSC определяет нарушенные ФЗ и вводит выборочное сравнение с помощью сопоставимых пар строк, чтобы значительно сократить требуемое попарное сравнение. Для обработки атрибутов малой мощности разработана стратегия сжатия прямого значения и комбинации значений. На шаге 2 алгоритм FSC индуцирует требуемые ФЗ по результатам шага 1. Обширные экспериментальные результаты, полученные на синтетических и реальных наборах данных, показывают, что алгоритм FSC может эффективно обнаруживать ФЗ в таблицах систем Big Data.

Алгоритмы поиска ФЗ между атрибутами в таблицах систем Big Data на основе анализа решетки атрибутов критичны к росту количества атрибутов в таблицах и хорошо работают при небольшом количестве атрибутов (не более 20) для большого количества строк (до сотен тысяч). Алгоритмы поиска ФЗ между атрибутами в таблицах на основе анализа строк таблицы критичны к росту количества строк и хорошо работают при относительно небольшом количестве строк (не более 1000) для большого количества атрибутов (до нескольких сотен). В гибридных алгоритмах поиска ФЗ между атрибутами в таблицах систем Big Data предприняты попытки устранить эти недостатки.

Т. Папенброк и Ф. Науманн разработали гибридный алгоритм HyFD [15], который объединяет методы поиска ФЗ, эффективные для большого количества атрибутов и большого количества строк. На первом этапе алгоритм HyFD извлекает небольшое подмножество строк из входных данных и вычисляет ФЗ для этой неслучайной выборки. Поскольку на этом этапе используется только подмножество строк, он работает особенно эффективно по атрибутам. Результатом является набор ФЗ, которые либо действительны, либо почти действительны по отношению к полному входному набору данных. На втором этапе алгоритм HyFD проверяет обнаруженные ФЗ по всему набору строк и находит такие ФЗ, которые не выполняются. Этот этап эффективен по строкам, поскольку он использует ранее обнаруженные ФЗ для эффективного сокращения пространства поиска. Если проверка становится неэффективной, алгоритм может вернуться к первому этапу и продолжить работу со всеми результатами, обнаруженными на данный момент. Эта

чередующаяся двухфазная стратегия обнаружения явно превосходит все существующие алгоритмы с точки зрения времени выполнения и масштабируемости, при этом обеспечивает обнаружение всех минимальных ФЗ.

В работе [16] Ц. Вэй и С. Линк отметили, что в алгоритме HyFD переключение стратегий происходит, когда текущая стратегия работает некорректно, однако это переключение происходит без доказательств того, что другая стратегия будет работать хорошо. Более того, ни в одном из предыдущих алгоритмов нет механизма для балансировки эффективности времени выполнения и использования основной памяти. Авторы [16] предложили новый гибридный алгоритм DHyFD, в котором переход от подхода на основе атрибутов к подходу на основе строк происходит всякий раз, когда есть вероятность проверки многих ФЗ. Эта вероятность оценивается с помощью предложенной меры. Установка более низких пороговых значений для этой меры дает алгоритму возможность использовать больше ресурсов основной памяти всякий раз, когда это может повысить эффективность выполнения. Еще одним недостатком известных алгоритмов поиска ФЗ между атрибутами в таблицах систем Big Data Ц. Вэй и С. Линк считают слишком большое количество получаемых ФЗ и большое суммарное количество атрибутов в их левых частях. Действительно, ни в одном из известных алгоритмов не ставится задача минимизации количества получаемых ФЗ и количества атрибутов в них. В алгоритме DHyFD для уменьшения количества полученных ФЗ используется метод построения канонического покрытия структуры ФЗ, который фактически строит элементарный базис структуры ФЗ по алгоритму К. Делобеля и Р. Кейси [17]. В работе [18] А.А. Карпук показал, что в общем случае этот элементарный базис структуры ФЗ не является минимальным по количеству ФЗ и не является оптимальным по суммарному количеству атрибутов в ФЗ, и предложил методы построения минимального и оптимального элементарного базиса структуры ФЗ.

В заключение нашего обзора основных алгоритмов поиска ФЗ между атрибутами в таблицах систем Big Data отметим последнюю работу Т. Блейфуса и Т. Папенброка с соавторами [19], в которой они предложили новый гибридный алгоритм FDhits. В этом алгоритме интегрировано несколько методов оптимизации для эффективного обнаружения ФЗ, а именно гибридная стратегия, поиск уникальных комбинаций атрибутов через перечисление наборов совпадений (Hiting Set Enumeration), однопроходная проверка всех правых частей кандидатов на ФЗ и распараллеливание вычислений. Авторы [19] утверждают, что алгоритм FDhits может за приемлемое время найти все минимальные ФЗ в таблице, содержащей до 100 атрибутов и более 5 миллионов строк, для которой все остальные алгоритмы вообще не могут найти решение. На наш взгляд, и этот алгоритм можно улучшить, применив для поиска уникальных комбинаций атрибутов результаты работы А.А. Карпука и В.В. Краснопрошина [20].

**Заключение.** По результатам проведенного анализа основных методов выделения и алгоритмов поиска ФЗ между атрибутами в таблицах систем Big Data можно сделать вывод, что исследования в этом направлении далеки от завершения, и их следует продолжить. Очевидно, что нет смысла улучшать существующие алгоритмы поиска ФЗ между атрибутами в таблицах на основе анализа решетки атрибутов и на основе анализа строк таблицы, поскольку область их применения существенно ограничена. Следует сосредоточить усилия на совершенствовании существующих и разработке новых гибридных алгоритмов поиска ФЗ между атрибутами в таблицах систем Big Data. В первую очередь требуется реализовать в этих алгоритмах методы минимизации количества получаемых ФЗ и количества атрибутов в них, а также более эффективные методы поиска уникальных комбинаций атрибутов.

## Список литературы

[1] Serra James. Deciphering Data Architectures. Choosing Between a Modern Data Warehouse,Data Fabric, Data Lakehouse, and Data Mesh. O'Reilly 2024. – 252 p.

[2] Armstrong W.W. Dependency Structure of Data Base Relationships // Proc. IFIP Congress. – Geneva, Switzerland, 1974. – P. 580–583.

[3] Chen P.P. The Entity-Relationship Model-Toward a Unified View of Data // ACM Transactions on Database Systems. – March 1976. – Vol. 1, No 1. – P. 9–36.

[4] Карпук А.А., Острейко В.М. Построение информационно-логической модели предметной области при проектировании базы данных // Вопросы радиоэлектроники. Сер. ОТ. – 1981. – Вып. 12. – С. 23–28.

[5] Karpuk A. Bringing the Subject Domain Ontology to Optimal Canonical Form // Open Semantic Technologies for Intelligent Systems: Research Papers Collection. Issue 8 – Minsk: BSUIR, 2024. – P. 237–242.

[6] Discover Dependencies from Data - A Review / Liu J. [et al.] // IEEE Transactions on Knowledge and Data Engineering (TKDE). – Vol. 24, No 2. – 2012. – P. 251–264.

[7] TANE: An Efficient Algorithm for Discovering Functional and Approximate Dependencies / Huhtala Y. [et al.] // Computer Journal. – Vol. 42, No 2. – 1999. – P. 100–111.

[8] Novelli N., Cicchetti R. FUN: An Efficient Algorithm for Mining Functional and Embedded Dependencies // Proceedings of 8th International Conference Database Theory, ser. ICDT. – 2001. – P. 189–203.

[9] Yao H., Hamilton H.J., Butz C.J. FD Mine: Discovering Functional Dependencies in a Database Using Equivalences // Proceedings of 2002 IEEE International Conference on Data Mining, ser. ICDM. – 2002. – P. 729–732.

[10] Abedjan Z., Schulze P., Naumann F. DFD: E_Cient Functional Dependency Discovery // Proceedings of the International Conference on Information and Knowledge Management (CIKM). – 2014. – P. 949–958.

[11] Flach A., Savnik I. Database Dependency Discovery: A Machine Learning Approach // AI Communications. – 1999. – Vol. 12., No 3. – P. 139–160.

[12] Lopes S., Petit J.M., Lakhal L. Efficient Discovery of Functional Dependencies and Armstrong Relations // Proceedings of the 7th International Conference on Extending Database Technology, ser. EDB. – 2000. – P. 350–361.

[13] Wyss C., Giannella C., Robertson E. FastFDs: A Heuristicdriven, Depth-First Algorithm for Mining Functional Dependencies from Relation Instances // Proceedings of International Conference on Data Warehousing and Knowledge Discovery. – 2001. – P. 101–110.

[14] Efficient Discovery of Functional Dependencies on Massive Data / Wan X. [et al.] // IEEE Transactions on Knowledge and Data Engineering. – Vol. 36, No 1. – 2024. – P. 107–121.

[15] Papenbrock T., Naumann F. A Hybrid Approach to Functional Dependency Discovery // Proceedings of the International Conference on Management of Data, ser. SIGMOD'16. – 2016. – P. 821–833.

[16] Wei Z., Link S. Towards the Efficient Discovery of Meaningful Functional Dependencies // Information Systems. – Vol. 116. – 2023. – P. 102–224.

[17] Delobel C., Casey R.G. Decomposition of a Data Base and the Theory of Boolean Switching Functions // IBM J. Res. And Dev. – 1973. – Vol. 17, No 5. – P. 374–386.

[18] Карпук А.А. Анализ структуры функциональных зависимостей между атрибутами реляционной базы данных // Экономика и менеджмент систем управления. – 2017. – № 3 (25). – С. 64–70.

[19] Discovering Functional Dependencies through Hitting Set Enumeration / Bleifuss T. [et al.] // Proceedings of the ACM on Management of Data ( SIGMOD). – Vol. 2, No 1. – 2024. – Article 43, P. 43:1–43:24.

[20] Карпук А.А., Краснопрошин В.В. Циклы в структурах функциональных зависимостей // International Journal of Open Information Technologies. – 2017. – Vol. 5, No 7. – P. 38–44.

## Авторский вклад

**Карпук Анатолий Алексеевич** – руководство исследованием методов выделения и алгоритмов поиска ФЗ между атрибутами в таблицах систем Big Data, постановка задач исследования, оценка результатов исследования и определение направлений дальнейшей работы по совершенствованию и разработке алгоритмов поиска ФЗ между атрибутами в системах Big Data.

**Лазута Леонид Сергеевич** – анализ методов выделения и алгоритмов поиска ФЗ между атрибутами в таблицах систем Big Data, определение области применения алгоритмов поиска ФЗ между атрибутами, разработка предложений по совершенствованию алгоритмов.

# IDENTIFICATION AND SEARCH FOR FUNCTIONAL DEPENDENCIES BETWEEN ATTRIBUTES IN BIG DATA SYSTEMS

***A.A. Karpuk***
*Professor, Department of Telecommunication Network Software, Belarusian State Academy of Communications, PhD of Technical sciences, Associate Professor*

***L.S. Lazuta***
*Postgraduate student, Department of Telecommunication Systems, Belarusian State Academy of Communications, Master´s Degree*

**Abstract.** The article provides an overview of methods for identifying and algorithms for searching for functional dependencies between attributes in Big Data systems. The article considers the main algorithms for searching for functional dependencies based on the analysis of the attribute lattice, based on the analysis of table rows, and hybrid algorithms. The areas of application of each type of algorithms and directions for further research on improving and developing algorithms for searching for functional dependencies between attributes in Big Data systems are determined.

**Keywords**: big data, functional dependency, functional dependency extraction, functional dependency search algorithms.

УДК 004.67

# LEVERAGING BIG DATA AND MACHINE LEARNING FOR OPTIMIZING BUSINESS DECISIONS IN E-COMMERCE

***Eugene Alooeff***
*R&D Engineer ATEK, Bachelor of Computer Science*
alooeff@atek.dev

***Mariia Bulycheva***
*Engineer Zalando SE, Master of Mathematics*
mbulytcheva@gmail.com

***E.A. Alooeff***
*Bachelor of Computer Science, Electronic Information Systems department of Brest State Technical University. Scientific interests: cloud technology, machine learning and multi-agent modeling.*

***M.V. Bulycheva***
*Master of science in mathematics, mechanics-mathematics department of Lomonosov State University, Moscow. Scientific interests: machine learning, large scale optimization and forecasting*

**Abstract.** In today's digital economy, e-commerce companies face increasing pressure to remain agile and competitive in a rapidly evolving marketplace. The explosion of data from online transactions, customer interactions, and external sources has opened new opportunities for data-driven decision-making. This paper explores how companies can leverage big data, advanced analytics, and machine learning - particularly deep learning techniques such as transformer-based architectures - to optimize pricing, discounting, and demand forecasting strategies. By analyzing historical sales data alongside customer behavior, competitor pricing, and seasonal trends, businesses can implement dynamic and personalized pricing strategies that enhance profitability, improve customer satisfaction, and streamline operations. The paper also highlights the technical foundations of modern demand forecasting models, best practices for model training and evaluation, and the real-world challenges businesses face in implementing these solutions. Ultimately, the integration of AI and big data technologies represents a powerful shift in how strategic business and production decisions are made in the e-commerce sector.

**Keywords:** Big Data, E-Commerce, Pricing Optimization, Demand Forecasting, Price Elasticity, Customer Segmentation, Personalized Pricing, Transformer-Based Models, Deep Learning, Time-Series Forecasting.

**Introduction.** In the era of digital transformation, businesses are generating an unprecedented volume of data daily. This surge is fueled by the wide spreading of IoT devices, the explosion of social media interactions, and the ever-increasing number of online transactions. From sales transactions and customer interactions to supply chain operations and market trends, data has become a valuable asset for companies seeking to optimize decision-making. The ability to extract actionable insights from historical data enables firms to improve operational efficiency, enhance customer experience, maximize profitability, and gain a significant competitive edge. This data can be monetized, used for product development, and provides deep insights into customer behavior, all of which are crucial for strategic planning.

One of the key areas where big data and machine learning play a crucial role is in pricing and discounting strategies for e-commerce companies [1]. This area is central to revenue management and overall business strategy. By leveraging historical sales, prices, discount levels data, specific product attributes, and external factors, businesses can make data-driven pricing decisions that balance profitability and growth and succeed in delivering their promises to

shareholders while remaining competitive [2]. Effective pricing strategies are essential for maximizing revenue, optimizing inventory, and responding to market dynamics.

**The role of big data in pricing and discounting decisions.** Traditionally, pricing decisions were based on simple cost-plus models or competitor benchmarking. However, with the advent of big data and machine learning, companies can now implement dynamic pricing strategies that adapt to market conditions in real time [3].

Historical sales data is one of the most valuable assets for pricing optimization. It provides a rich source of insights that help businesses understand demand patterns, customer behavior, and price sensitivity over time. Companies use data-driven pricing strategies to maximize profitability while maintaining competitiveness. Below, we explore the key components of leveraging historical sales data for pricing optimization.

**Key data sources for pricing optimization.** To optimize pricing decisions, companies gather and analyze various types of historical data:

Table 1. Data types

| **Data type** | **Description and use case** |
|---|---|
| Transactional Data | Past sales volumes, including date, time, price and discount applied, and quantity sold. Used to track demand fluctuations and measure price elasticity. |
| Customer Data | Purchase history, segmentation, and price sensitivity. Helps in personalized pricing and targeted discounts. |
| Competitive Pricing Data | Market prices from competitors, collected via web scraping or third-party data providers. Enables dynamic pricing adjustments, helps the companies stay competitive at strategically important positions. |
| Promotional Data | Historical impact of discounts and special offers. Helps in assessing promotion campaigns' effectiveness. |
| Seasonality & External Factors | Data on holidays, weather patterns, and economic conditions affecting demand. Used for forecasting price-sensitive time periods. |

**Demand prediction and price elasticity.** One of the fundamental principles of pricing optimization is price elasticity, which measures how demand for a product changes in response to price adjustments.

1 Elastic demand: A small price decrease leads to a significant increase in sales volume. Common in price-sensitive markets (e.g., consumer electronics, fast fashion). Example: If the price of a popular brand of headphones decreases by 10%, sales might increase by 30%.

2 Inelastic demand: Price changes have little effect on sales volume. Common for premium brands or necessity goods (e.g., prescription drugs, luxury items). Example: Even if the price of a life-saving medication increases, demand will likely remain stable.

Techniques Used:

1 Regression Analysis – Models how changes in price affect sales volume. Example: Linear regression can be used to create a model that predicts sales based on price changes, while multiple regression can incorporate other factors like advertising spend.

2 A/B Testing – To test different pricing strategies, companies conduct A/B tests to measure the impact of price changes on sales and customer behavior. By splitting customers into two groups and offering different prices, businesses can analyze which pricing strategy yields higher conversions and revenue. For instance, subscription-based services like Netflix and Spotify experiment with different pricing tiers to determine the optimal balance between customer

acquisition and retention. Another approach involves segmenting similar products, also known as substitute goods, into distinct groups and applying tailored pricing strategies to each segment (if the local legislation does not allow price discrimination).

3 Demand Forecasting Models using statistical times series forecasting – Uses historical data to predict future responses to pricing changes. Example: Using ARIMA or other time-series models to forecast demand based on past sales and seasonal trends.

The utilization of deep neural networks, particularly transformer-based architectures, represents a significant leap forward in demand forecasting. To understand this powerful methodology in detail, we will explore it in the following section.

**Demand forecasting with deep learning.** Demand forecasting is a critical component in various industries, ranging from retail and e-commerce to manufacturing and logistics. Accurate predictions enable businesses to optimize inventory management, pricing strategies, and supply chain efficiency. Traditional statistical models and machine learning techniques have been widely used for demand forecasting. However, recent advances in deep learning, particularly transformer-based architectures, have demonstrated superior performance in capturing complex patterns and dependencies within time-series data.

**Why use transformers for demand forecasting?** Transformers are a type of deep neural network architecture originally designed for sequence-to-sequence modeling tasks in natural language processing [4]. They rely heavily on a mechanism called self-attention, which allows the model to weigh the importance of different elements in an input sequence when generating an output sequence. Unlike recurrent architectures, transformers process all elements in a sequence simultaneously, enabling them to model long-range dependencies more efficiently and without the limitations of sequential computation. The main idea behind using a transformer-based architecture for demand forecasting is that demand data forms a time series—and a time series is, fundamentally, a sequence. This allows us to feed a sequence of past demand values into the model and train it to output a corresponding sequence of future demand values, not just a single point forecast but an entire forecast horizon covering a specified time period [5].

Key advantages of using transformers for demand forecasting include:

1 Handling large-scale data: Transformers can process vast amounts of time-series data across multiple products and markets.

2 Capturing complex dependencies: The self-attention mechanism allows the model to focus on relevant past information when making predictions.

3 Scalability and efficiency: With parallel processing, transformers can handle high-dimensional time-series data more efficiently than RNN-based models.

**Data preparation for transformer-based forecasting.** Before training a transformer model for demand forecasting, proper data preparation is crucial. The key steps include:

1 Time-Series Aggregation: Demand data is collected at different granularities (daily, weekly, or monthly) and aggregated accordingly.

2 Feature Engineering: Additional covariates such as price discounts, stock levels, product attributes, holidays, and weather conditions can enhance predictive accuracy.

3 Handling Missing Data: Techniques such as interpolation or imputation can be used to fill gaps in historical data.

4Sales to Demand Translation: In cases where stock-outs occur, observed sales may not reflect actual demand. Estimating lost demand due to stock unavailability is important for accurate forecasting.

**Transformer-based demand forecasting model.** A typical transformer-based forecasting model consists of the following components [6][7]:

1 Encoder-Decoder Architecture: The encoder processes historical demand data, while the decoder generates future demand predictions. The attention mechanism helps identify important dependencies across time steps.

2 Positional Encoding: Since transformers do not inherently process sequential data, positional encoding is added to represent time-series order.

3 Multi-Head Attention Mechanism: This mechanism enables the model to attend to multiple relevant past points in time simultaneously.

**Training and optimization.** Training a transformer model involves several steps:

1 Loss Function Selection: Mean squared error (MSE) or custom weighted loss functions tailored to business objectives (e.g., profit maximization) are commonly used.

2 Regularization Techniques: Dropout, layer normalization, and batch normalization help prevent overfitting.

3 Hyperparameter Tuning: Parameters such as number of attention heads, hidden layers, and learning rates need to be optimized for the best performance.

4 Retraining Frequency: Regular retraining ensures that the model adapts to shifts in demand trends and external factors.

**Model evaluation.** To measure forecasting accuracy, businesses typically use metrics such as:

1 Root Mean Squared Error (RMSE): Measures the difference between predicted and actual demand.

2 Mean Absolute Percentage Error (MAPE): Evaluates the percentage deviation of forecasts.

3 Demand Bias: Identifies whether the model consistently overestimates or underestimates demand.

**Applications of transformer-based demand forecasting.**

1 Retail and E-commerce: Predicting product demand for dynamic pricing and inventory management.

2 Supply Chain Optimization: Anticipating material requirements to prevent supply disruptions.

3 Manufacturing: Forecasting raw material needs to streamline production schedules.

4 Logistics and Transportation: Demand predictions for optimizing fleet management and delivery planning.

**Time-based & seasonal pricing.** Historical data often reveals strong seasonal trends that affect demand. Businesses use these patterns to optimize prices dynamically [8].

Example: A fashion retailer sees a spike in sales of winter coats from October to February based on past sales data. They optimize pricing by:

1 Maintaining premium prices in October-November when demand is naturally high.

2 Gradually reducing prices in December-January as customers expect post-holiday discounts.

3 Applying deep discounts in February to clear remaining stock.

Techniques Used:

1 Statistical time-series forecasting (e.g. ARIMA) – Predicts seasonal demand fluctuations.

2 Moving Averages – Identifies cyclical trends in sales.

3 Fourier Transform & Decomposition – Extracts seasonal patterns from noisy sales data.

**NLP & sentiment analysis for pricing insights.** Companies also use Natural Language Processing (NLP) to analyze customer sentiment and predict price sensitivity [12][13].

Example: A retailer collects customer reviews and social media mentions to assess perceived value of products. If negative sentiment increases due to high prices, they adjust pricing accordingly.

Techniques Used:

1 Sentiment Analysis (BERT, VADER, TextBlob) [14] – Measures customer opinions on pricing.

2 Topic Modeling (LDA, Word2Vec) [15] – Identifies pricing-related discussions.

**Challenges & best practices** Despite the advantages, pricing optimization based on historical data has challenges:

Table 2. Challenges and solutions

| **Challenge** | **Solution** |
|---|---|
| Data Quality and Integration | Businesses must ensure that historical sales data is clean, consistent, structured and integrated across multiple sources |
| Customer Backlash Against Dynamic Pricing | Use transparent pricing policies and ensure fairness. Frequent price changes can lead to customer dissatisfaction if not implemented transparently |
| Overfitting in Machine Learning Models | Regularize models and test on unseen data to prevent overfittings. |
| Regulatory Compliance | Ensure compliance with antitrust laws and ethical pricing practices. In some industries, dynamic pricing is subject to legal and ethical scrutiny (e.g., price discrimination concerns). |
| Computational Complexity | Advanced AI models require significant computational power and expertise to maintain. |

**Conclusion.** The use of big data, advanced analytics, and machine learning has transformed pricing and discounting strategies across industries. By leveraging historical sales data, companies can optimize pricing in real time, enhance customer segmentation, and improve demand forecasting. By integrating machine learning, demand forecasting, and dynamic pricing algorithms, companies can increase revenue, improve customer satisfaction, and enhance operational efficiency. However, ethical considerations and transparency must remain central to avoid consumer distrust.

As technology evolves, the future of pricing strategies will increasingly rely on automated AI systems, personalized pricing models, and real-time market intelligence. Companies that invest in data-driven pricing solutions today will be better positioned to thrive in the dynamic and competitive business landscape of tomorrow.

The future of pricing optimization will likely see more automated, real-time AI systems, ensuring that businesses can stay competitive in an increasingly data-driven marketplace.

## References

[1] Cheung, Simchi-Levi, Wang (2018). Dynamic pricing and demand learning with limited price experimentation. Operations Research, 66(1), 144–162.

[2] Feng, Y., & Xiao, B. (2000). A Continuous-Time Yield Management Model with Multiple Prices and Reversible Price Changes. May 2000 Management Science 46(5):644-657.

[3] Kraus, M., Feuerriegel, S., & Oztekin, A. (2020). Deep learning in business analytics and operations research: Models, applications, and managerial implications. European Journal of Operational Research, 281(3), 628–641.

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin “Attention is all you need”

[5] Manuel Kunz, Adele Gouttes, Mateusz Koren, Tofigh Naghibi, Johannes Stephan, Mariia Bulycheva, Matthias Grzeschik, Armin Kekić, Kashif Rasul, Tim Januschowski. Deep Learning based Forecasting: a case study from the online fashion industry” (2023).

[6] Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting. International Journal of Forecasting.

[7] Li, S., Jin, X., Xuan, Y., Zhou, X., Chen, W., Wang, Y. X., & Yan, X. (2019). Enhancing the Locality and Breaking the Memory Bottleneck of Transformer on Time Series Forecasting. Advances in Neural Information Processing Systems (NeurIPS).

[8] Bandara, K., Bergmeir, C., & Smyl, S. (2020). Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach. Expert Systems with Applications, 140, 112896.

[9] Hassani, H., Huang, X., & Silva, E. S. (2020). Digitalisation and big data mining in banking. Big Data Cognitive Computing, 4(1), 5.

[10] Vellido, A., Lisboa, P. J., & Vaughan, J. (2012). Neural networks in business: A survey of applications (1999). Expert Systems with Applications, 39(12), 9290–9311.

[11] Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. Expert Systems with Applications, 36(2), 2592–2602.

[12] Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. IEEE Intelligent Systems, 28(2), 15-21.

[13] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal, 5(4), 1093–1113.

[14] Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.

[15] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(Jan), 993–1022.

## Author's contribution

**Eugene Alooeff** – conducted a general overview and reference investigation.
**Mariia Bulycheva** – conducted a transformer-based architecture analysis and implementation.

# ИСПОЛЬЗОВАНИЕ БОЛЬШИХ ДАННЫХ И МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПТИМИЗАЦИИ БИЗНЕС-РЕШЕНИЙ В ЭЛЕКТРОННОЙ КОММЕРЦИИ

***Е.А. Алуев***
*Инженер-исследователь ATEK,*
*Бакалавр технических наук*

***М.А. Булыева***
*Инженер Zalando SE,*
*Магистр математических наук*

**Аннотация.** В современной цифровой экономике компании электронной коммерции сталкиваются с растущим давлением, чтобы оставаться гибкими и конкурентоспособными на быстро меняющемся рынке. Взрывной рост данных из онлайн-транзакций, взаимодействия с клиентами и внешних источников открыл новые возможности для принятия решений на основе данных. В этой статье рассматривается, как компании могут использовать большие данные, расширенную аналитику и машинное обучение, в частности методы глубокого обучения, такие как архитектуры на основе трансформеров, для оптимизации стратегий ценообразования, скидок и прогнозирования спроса. Анализируя исторические данные о продажах наряду с поведением клиентов, ценообразованием конкурентов и сезонными тенденциями, компании могут внедрять динамические и персонализированные стратегии ценообразования, которые повышают прибыльность, улучшают удовлетворенность клиентов и оптимизируют операции. В статье также освещаются технические основы современных моделей прогнозирования спроса, передовые методы обучения и оценки моделей, а также реальные проблемы, с которыми сталкиваются компании при внедрении этих решений. В конечном счете, интеграция технологий искусственного интеллекта и больших данных представляет собой мощный сдвиг в том, как принимаются стратегические деловые и производственные решения в секторе электронной коммерции.

**Ключевые слова:** Большие данные, Электронная коммерция, Ценовая оптимизация, Прогнозирование спроса, Эластичность цен, Сегментация клиентов, Персонализированное ценообразование, Модели на основе трансформеров, Машинное обучение, Прогнозирование временных рядов.

УДК 004.021:004.75

# ОБРАБОТКА И РАСПОЗНАВАНИЕ ГЕОМЕТРИЧЕСКИХ ПАТТЕРНОВ В ИГРОВЫХ ДАННЫХ

***С. А. Берговин***
*Учащийся учреждения образования «Национальный детский технопарк», учащийся ГУО «Гимназия имени Я. Купалы» г. Мозыря*

***А. Ю. Саливон***
*Учащаяся учреждения образования «Национальный детский технопарк», учащаяся ГУО «Высоковская средняя школа» Пинского района*

***Ф.В. Усенко***
*ассистент кафедры инженерной психологии и эргономики БГУИР, магистр*
f.usenko@bsuir.by

***А.М. Прудник***
*Доцент кафедры инженерной психологии и эргономики БГУИР, кандидат технических наук, доцент*
aleksander.prudnik@bsuir.by

***С. А. Берговин***
*Учащийся гимназии имени Я. Купалы г. Мозыря, область интересов связана с информатикой и изучением точных наук.*

***А. Ю. Саливон***
*Учащаяся Высоковской средней школы Пинского района, область интересов связана с информатикой и изучением точных наук.*

***Ф. В. Усенко***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Областью научных интересов является моделирование пользователей информационных систем и разработка интерфейсов.*

***А. М. Прудник***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с взаимодействием человека с компьютером, интерфейсами информационных систем, пользовательскими интерфейсами, front-end web development.*

**Аннотация.** В работе рассматривается применение методов аналитической геометрии для реализации механики рисования заклинаний в игре Cave Enchant. Описан алгоритм распознавания геометрических фигур,

учитывающий координатные преобразования и выпуклость многоугольников. Применённый метод минимизирует влияние дрожания руки игрока и обладает высокой вычислительной эффективностью, а также предоставляет возможности его использования в других аспектах разработки игр.

**Ключевые слова: а**налитическая геометрия, декартова система координат, алгоритм распознавания фигур, игровая механика, Cave Enchant, Unity, C#, Visual Studio, виртуальная реальность, интерактивное взаимодействие.

**Введение.** Разработка игр — сложный мультидисциплинарный процесс, состоящий из множества этапов, на каждом из которых математические методы играют важную роль. Для создания простых игр могут понадобиться базовые навыки программирования и проектирования, но углубленные знания в области математики позволяют значительно повысить качество и сложность реализуемых игровых элементов.

Аналитическая геометрия – это раздел геометрии, в котором геометрические фигуры и их свойства исследуются средствами алгебры [1]. В основе метода аналитической геометрии лежит метод координат, который каждому геометрическому соотношению ставит в соответствие некое уравнение, связывающее координаты фигуры или тела.

В игре «Cave Enchant», пользователь при помощи контроллеров виртуальной реальности может рисовать заклинания для сражения с противниками (рисунок 1). Для этой механики и используются методы аналитической геометрии. Несмотря на сложность данного математического раздела, в проекте были использованы его упрощенные методы, обеспечивающие требуемую функциональность.

*Рисунок 1.* Демонстрация механики

Игра «Cave Enchant» была создана на игровом движке Unity, программный код был написан на языке C# в среде разработки Visual Studio, 3D-модели были созданы в редакторе Blender.

**Описание алгоритма.** При нажатии игроком на триггер контроллера, перед ним появляется рабочая область, предназначенная для рисования простых геометрических фигур. В механике игры предусмотрено шесть заклинаний, каждому из которых соответствует определенная геометрическая фигура: треугольник, выпуклый и невыпуклый четырёхугольники, выпуклый и невыпуклый пятиугольники, а также пятиконечная звезда. Если нарисованная пользователем фигура совпадает с одной из предопределённых форм, то после отпускания триггера контроллера создается заклинание; в противном случае оно не активируется.

**Реализация алгоритма.** Когда игрок нажимает на курок контроллера, перед ним создаётся объект SpellInputSpace, представляющий собой плоскость. Локальная система координат этого объекта установлена так, что точки с локальной координатой $z$, равной 0, лежат в этой плоскости. Также из контроллера игрока создаётся луч, который сталкивается

с этой плоскостью и создаёт в месте столкновения точку HitObject. Так как эта точка лежит на данной плоскости, то её координата *z* в системе координат этой плоскости равна нулю. Таким образом, любое положение этой точки можно представить всего в двух координатах – *x* и *y* (рисунок 2).

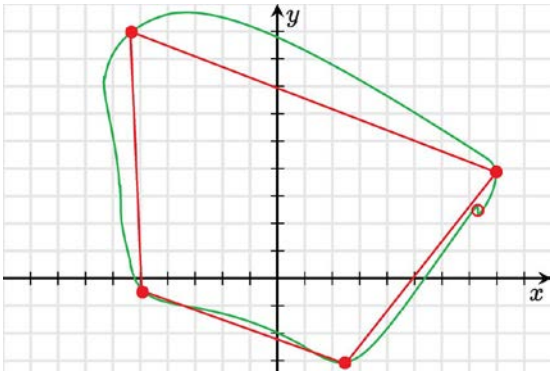



*Рисунок 2*. Визуализация алгоритма

Далее скрипт (рисунок 3) фиксирует величину угла, нарисованную игроком, и если она превышает заданное пороговое значение, то создаётся точка-угол.

```
private void CalculatingDelta()
{
    float hitDiff = hitDiffDelta;
    Vector2 hitDelta = hitObject.transform.position;
    if ((hitDelta - hitPos).sqrMagnitude > cornerSensitivity * cornerSensitivity)
    {
        hitDiffDelta = Mathf.Atan2(hitDelta.y - hitPos.y, hitDelta.x - hitPos.x);
        hitDiffSecond = hitDiffDelta - hitDiff;
    }
    float pointDistance = points.Count > 0 ? (points[points.Count - 1].transform.position - hit.point).sqrMagnitude : 999;
    if (Mathf.Abs(hitDiffSecond * Mathf.Rad2Deg) > 20 && pointDistance > cornerSensitivity*5)
    {
        hitDiffSecond = 0;
        points.Add(new SpellObject(Instantiate(pointPrefab)));
        points[points.Count - 1].transform.parent = spellSpace.transform;
        points[points.Count - 1].transform.position = hit.point;
        points[points.Count - 1].transform.parent = spellSpace.transform;
        if (points.Count > 2) Points();
    }
    hitPos = hitDelta;
}
```

*Рисунок 3.* Метод расчёта изменения угла.

Пусть ось *X* – множество всех точек, имеющих координату y, равную нулю в системе координат SpellInputSpace, а ось *Y* — множество всех точек, имеющих координату *x*, равную нулю в той же системе координат. Сначала определяется угол α между осью *X* и прямой, проходящей через точки с координатами HitObject в текущем и предыдущем кадре. Для этого используется метод Mathf.Atan2, который в отличие от стандартного метода Atan позволяет избежать ошибки, заключающейся в том, что период функции тангенса равен 180º [4]. Затем определяется, насколько быстро изменяется угол в текущем кадре, по формуле:

$$\Delta\alpha = \frac{\alpha_{\text{предыдущий кадр}} - \alpha_{\text{текущий кадр}}}{\Delta t},$$

где Δt – время, прошедшее между предыдущим и текущим кадром, взятое при помощи метода Time.DeltaTime, а Δα – скорость изменения угла.

Если значение Δα достаточно большое, то на месте HitObject ставится точка, обозначающая угол фигуры.

Поскольку движения руки игрока не могут быть абсолютно ровными и плавными, метод фиксирует изменение угла только в том случае, если HitObject прошёл достаточное расстояние. Также, когда игрок проводит прямую линию, удаляются лишние точки, созданные случайным дрожанием руки. Для этого рассматривается треугольник из трёх точек-углов. Если разность суммы длин двух его меньших сторон и длины наибольшей

стороны достаточно мала, или же если этот треугольник имеет угол близкий к 180º, то удаляется лишняя точка, и остаётся прямая линия. Данный скрипт представлен на рисунке 4.

```
private void Points()
{
    int i = points.Count - 2;
    int j=-1;
    float a = (points[i - 1].transform.position - points[i].transform.position).magnitude;
    float b = (points[i].transform.position - points[i+1].transform.position).magnitude;
    float c = (points[i - 1].transform.position - points[i+1].transform.position).magnitude;
    if (a == Mathf.Max(a, b, c) && b + c - a < a * filterSensitivity) j = i + 1;
    if (b == Mathf.Max(a, b, c) && a + c - b < b * filterSensitivity) j = i - 1;
    if (c == Mathf.Max(a, b, c) && b + a - c < c * filterSensitivity) j = i;
    if (j > -1)
    {
        Destroy(points[j].gameObject);
        points.RemoveAt(j);
        foreach (SpellObject line in lines)
        {
            Destroy(line.gameObject);
        }
        lines.Clear();
        DrawLines();
    }
}
```

*Рисунок 4.* Метод удаления лишних точек.

Далее, между точками рисуются линии. Линии и точки перекрашиваются в зависимости от того, какая фигура нарисована в данный момент.

Определение типа фигуры осуществляется с помощью метода координат. Для распознавания треугольника достаточно наличие трёх точек. Для определения остальных фигур необходимо проанализировать выпуклости многоугольника.

Многоугольник считается выпуклым, если для каждой прямой, содержащей одну из его сторон, все его вершины, не лежащие на этой прямой, расположены по одну сторону от него. В противном случае он является невыпуклым [2].

Определение пятиконечной звезды основано на следующем критерии: для каждой прямой, содержащей одну из её сторон, две вершины фигуры находятся по одну сторону от прямой, а одна вершина — по другую.

Как мы можем видеть на рисунке 5, эти утверждения верны. Например, на рисунке 5 (б) для невыпуклого четырёхугольника условие расположения вершин по одну сторону от прямой выполняется только для двух его сторон, тогда как для остальных двух — нет.

В случае пятиконечной звезды (рисунок 5 (в)) её характеристическое условие выполняется для каждой стороны: для каждой прямой, содержащей одну из сторон звезды, две вершины расположены по одну сторону, а одна вершина — по другую.

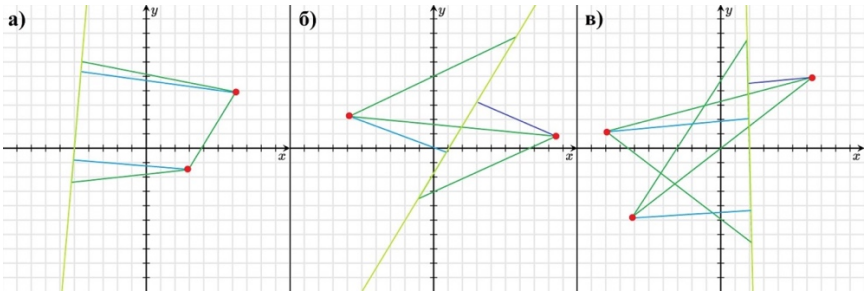



*Рисунок 5.* Положение вершин относительно прямой для а) выпуклого четырёхугольника, б) невыпуклого четырёхугольника, в) пятиконечной звезды.

Для того чтобы определить положение вершин относительно прямой, скрипт (рисунок 6) представляет каждую сторону многоугольника в виде уравнения прямой:

$$y = kx + b,$$

где $k$ – тангенс угла между этой прямой и осью $X$, а $b$ – смещение этой прямой вдоль оси $Y$.

```
case 5:
    float[] currLines5k = new float[5];
    float[] currLines5b = new float[5];
    bool pentagon = true;
    bool star = true;
    for (int i = 0; i < 5; i++)
    {
        currLines5k[i] = (points[i == 4 ? 0 : i + 1].transform.localPosition.y - points[i].transform.localPosition.y) /
            (points[i == 4 ? 0 : i + 1].transform.localPosition.x - points[i].transform.localPosition.x);
        currLines5b[i] = points[i].transform.localPosition.y - points[i].transform.localPosition.x * currLines5k[i];
        int p2 = 0;
        int[] sign = new int[3];
        for(int j =0;j<5;j++)
        {
            if (j == i || j == (i == 4 ? 0 : i + 1)) continue;
            sign[p2] = (int)Mathf.Sign(currLines5k[i] * points[j].transform.localPosition.x + currLines5b[i] - points[j].transform.localPosition.y);
            p2++;
        }
        Array.Sort(sign);
        if (sign[0] != sign[2]) pentagon = false;
        if (!((sign[0] == sign[1] && sign[1] != sign[2]) || (sign[0] != sign[1] && sign[1] == sign[2]))) star = false;
        if (!star && !pentagon) i = 5;
    }
    if (pentagon) spell = 2;
    else if (star) spell = 5;
    else spell = 4;
    break;
```

*Рисунок 6.* Алгоритм определения типа пятиугольника.

Коэффициент $k$ можно найти как:

$$\mathrm{k} = \frac{y_2 - y_1}{x_2 - x_1},$$

где $x_1$, $y_1$, $x_2$, $y_2$ – координаты двух вершин, лежащих на данной прямой.

Коэффициент $b$ можно найти, решив уравнение:

$$y_1 = \mathrm{k}x_1 + b, \qquad b = y_1 - kx_1.$$

Для определения положения точки относительно прямой подставляется координата $x$ точки в полученное уравнение, после чего из координаты $y$ этой точки вычитается вычисленное значения $y$. Знак полученной разности $\Delta y$ определяет, по какую сторону от прямой находится данная точка (рисунок 7).

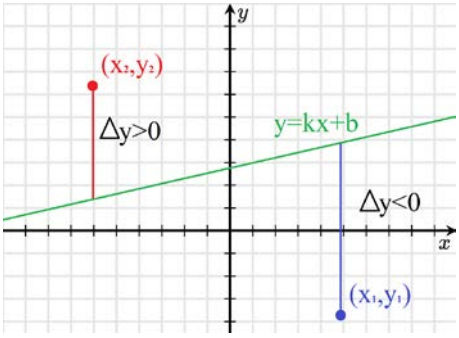



*Рисунок 7.* Положение точки относительно прямой

Существуют и другие способы определения выпуклости многоугольника [3]. Однако данный алгоритм был выбран, так как он позволяет не только определить выпуклость, но ещё и распознает пятиконечную звезду. Поскольку в рассматриваемом случае число сторон ограничено пятью, алгоритм остаётся малозатратным с точки зрения вычислительных ресурсов, так как при увеличении числа сторон он перестанет быть выполнимым.

**Заключение.** Результатом использования методов аналитической геометрии стало реализация механики рисования заклинаний в игре Cave Enchant. Данный подход обеспечил точное распознавание геометрических фигур, используемых для активации заклинаний, и эффективную обработку пользовательского ввода. Следует отметить, что эти методы находят широкое применение в других сферах разработки игр, включая

рендеринг и симуляцию физических процессов. Кроме того, они могут быть использованы для создания различных интерактивных механик, расширяя возможности игровой индустрии.

## Список литературы

[1] Ласуков Владимир Васильевич, Томский политех., Математика 1, лекция 7, «Аналитическая геометрия», https://portal.tpu.ru/SHARED/l/LASUKOV/umkd/math/math_1_6_Lecture-07.pdf.

[2] Джон М. Ли, «Axiomatic Geometry», https://books.google.ca/books?id=9Z0xAAAAQBAJ&pg=PA155&source=gbs_toc_r&cad=4#v=onepage&q&f=false

[3] Хекберт Пауль (ред.), «Graphic Gems IV», 1994 г, «I.2 Testing the convexity of the polygon» (стр. 7-15), https://archive.org/details/isbn_9780123361554/page/7.

[4] Unity Scripting API, Документация скриптинга в юнити, https://docs.unity3d.com/6000.0/Documentation/ScriptReference/index.html.

# PROCESSING AND RECOGNIZING GEOMETRIC PATTERNS IN GAME DATA

***S. A. Bergovin***
*Student of the educational Institution "National Children's Technopark", student GUO «Yanka Kupala Gymnasium», Mosyr*

***A. Y. Salivon***
*Student of the educational Institution "National Children's* Technopark", student GUO «Vysokoe school», Pinsk district

***Ph.V. Usenko***
*Assistant of the Department of Engineering Psychology and Ergonomics of BSUIR, M.Sc*

***A. M. Prudnik***
*Associate Professor of Engineering Psychology and Ergonomics Department of BSUIR, Candidate of Technical Sciences, Associate Professor*

**Abstract.** This paper discusses the application of analytic geometry methods to the implementation of spell drawing mechanics in the game Cave Enchant. An algorithm for recognizing geometric shapes that takes into account coordinate transformations and convexity of polygons is described. The method minimizes the influence of player's hand shaking and has high computational efficiency, as well as provides opportunities for its use in other aspects of game development.

**Keywords:** analytical geometry, Cartesian coordinate system, shape recognition algorithm, game mechanics, Cave Enchant, Unity, C#, Visual Studio, virtual reality.

УДК 004.8

# ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПЕРСОНАЛИЗАЦИИ АДАПТИВНОГО ДИЗАЙНА

***Т.А. Васяева***
*Декан факультета информационных систем и технологий ДонНТУ, кандидат технических наук, доцент*
tanechka.vasyaeva@yandex.ru

***М.А. Рябко***
*Магистрант кафедры АСУ, ДонНТУ*
maxryabko2706@yandex.ru

***Т.А. Васяева***
*Кандидат технических наук, доцент, декан факультета «Информационные системы и технологии», ФГБОУ ВО «Донецкий национальный технический университет». Область научных интересов связана с исследованием и разработкой методов искусственного интеллекта, организацией учебного и научно-исследовательского процессов в техническом университете.*

***М.А. Рябко***
*Магистрант кафедры автоматизированных систем управления, ФГБОУ ВО «Донецкий национальный технический университет». Область научных интересов связана с исследованием и разработкой методов адаптивного и мобильного дизайна веб-сайтов.*

**Аннотация.** В современном цифровом мире персонализация играет ключевую роль в создании эффективного пользовательского опыта. Использование машинного обучения (МО) позволяет значительно улучшить адаптивный дизайн веб-сайтов, обеспечивая более релевантный контент и интерфейс для пользователей. В данной статье рассматриваются проблемы традиционного адаптивного дизайна, методы персонализации с использованием МО, а также предлагается практическая реализация данного подхода.

**Ключевые слова:** Адаптивность, машинное обучение, веб-сайт, генетический алгоритм, нейронные сети.

**Введение.** В последние годы персонализация цифрового контента и пользовательского опыта становится одним из ключевых направлений в развитии веб-технологий. С ростом количества интернет-пользователей и разнообразия устройств, с которых осуществляется доступ к веб-сайтам, традиционные методы адаптивного дизайна перестают быть эффективными. Они учитывают лишь технические характеристики устройства, такие как размер экрана или разрешение дисплея, но не принимают во внимание индивидуальные предпочтения пользователя. Это приводит к стандартному, одинаковому для всех опыту взаимодействия, который не всегда соответствует ожиданиям конкретного пользователя.

Применение машинного обучения (МО) позволяет решить эту проблему за счёт анализа данных о поведении пользователей и предсказания их потребностей. Алгоритмы МО способны выявлять скрытые закономерности в пользовательских данных, адаптируя интерфейс веб-сайтов в режиме реального времени. Это делает возможным создание

персонализированных решений, которые не только улучшают пользовательский опыт, но и повышают показатели вовлечённости, конверсии и удовлетворённости аудитории.

**Цель данной работы.** Исследовать возможности применения МО для персонализации адаптивного дизайна веб-сайтов, проанализировать существующие методы и предложить модель для улучшения пользовательского опыта. Актуальность темы обусловлена тем, что персонализация интерфейса становится конкурентным преимуществом для компаний, работающих в цифровом пространстве. Использование методов МО в адаптивном дизайне позволяет веб-разработчикам и владельцам интернет-ресурсов предоставлять контент, максимально соответствующий потребностям пользователей, что ведёт к улучшению взаимодействия с сайтом и увеличению коммерческой эффективности.

**Постановка задачи.** Развитие цифровых технологий и рост количества интернет-ресурсов ставят перед разработчиками веб-дизайна важную задачу: создать интерфейсы, которые не только адаптируются под технические параметры устройства, но и учитывают индивидуальные предпочтения пользователя. Традиционные методы адаптивного дизайна фокусируются в основном на статических параметрах, таких как разрешение экрана, но не принимают во внимание поведенческие и контекстные факторы.

Основные задачи исследования включают:

1 Анализ традиционных методов адаптивного дизайна и их ограничений в контексте персонализации.

2 Исследование методов машинного обучения, применяемых для персонализации пользовательского опыта.

3 Разработку модели персонализированного адаптивного дизайна с применением МО.

4 Оценку эффективности предложенной модели на основе ключевых метрик, таких как вовлеченность пользователей, среднее время на сайте, коэффициент конверсии.

5 Решение данных задач позволит предложить инновационные методы персонализации интерфейсов, улучшив качество взаимодействия пользователей с цифровыми платформами.

**Описание проблемы.** Современный адаптивный дизайн сталкивается с рядом проблем, связанных с обеспечением персонализированного пользовательского опыта. Традиционные методы адаптации интерфейсов учитывают лишь технические параметры устройств, игнорируя индивидуальные предпочтения и поведение пользователей. Это приводит к созданию унифицированных интерфейсов, не всегда соответствующих ожиданиям конкретных пользователей.

Основные проблемы адаптивного дизайна:

1 Ограниченная персонализация. Традиционные подходы к адаптивному дизайну фокусируются на технических характеристиках устройств, таких как размер экрана и разрешение, не учитывая индивидуальные предпочтения пользователей. Это приводит к созданию однотипных интерфейсов, не отражающих уникальные потребности каждого пользователя.

2 Сложности в отображении контента на разных устройствах. Адаптация сложных элементов, таких как таблицы данных, на небольших экранах представляет значительные трудности. Это может ухудшать пользовательский опыт и снижать эффективность взаимодействия с веб-ресурсом.

3 Проблемы навигации и визуализации. На небольших экранах навигационные элементы и визуальные компоненты могут отображаться некорректно, что затрудняет доступ к информации и снижает удовлетворенность пользователей.

**Роль машинного обучения в решении проблем адаптивного дизайна**. Интеграция искусственного интеллекта (ИИ) и машинного обучения (МО) в процессы веб-дизайна

открывает новые возможности для персонализации и адаптации интерфейсов. Такие как, например:

1 Анализ пользовательского поведения. Алгоритмы МО могут обрабатывать большие объемы данных о поведении пользователей, выявляя скрытые закономерности и предпочтения. Это позволяет создавать интерфейсы, адаптированные под индивидуальные потребности, улучшая пользовательский опыт.

2 Автоматизация дизайна. ИИ способен автоматически генерировать дизайн-решения на основе анализа предпочтений пользователей, что ускоряет процесс разработки и обеспечивает более точное соответствие ожиданиям аудитории.

3 Персонализация контента. Использование МО позволяет предлагать пользователям контент, наиболее соответствующий их интересам, повышая вовлеченность и удовлетворенность.

**Варианты решения проблемы.** В процессе персонализации адаптивного дизайна могут использоваться различные методы машинного обучения. Рассмотрим подробнее основные подходы и их применение:

1 Кластеризация пользователей

Кластеризация позволяет группировать пользователей на основе их поведения, предпочтений и демографических характеристик. Следует рассмотреть наиболее популярные алгоритмы:

K-means – простой и эффективный метод, разделяющий пользователей на группы по схожести их данных. DBSCAN – полезен для выделения аномалий и сегментирования пользователей без необходимости задавать количество кластеров заранее. Иерархическая кластеризация – строит древовидную структуру взаимосвязей между пользователями.

2 Коллаборативная фильтрация

Метод широко используется в рекомендательных системах. Разделяется на два типа:

– пользовательская коллаборативная фильтрация – рекомендации на основе сходства между пользователями;

– контентная коллаборативная фильтрация – анализирует предпочтения пользователей на основе характеристик контента.

Применение: подбор динамических элементов интерфейса в зависимости от предпочтений пользователей, персонализированные рекомендации товаров или контента.

3 Нейронные сети и глубокое обучение

Глубокие нейронные сети (DNN) позволяют анализировать сложные паттерны поведения пользователей. Основные архитектуры:

– рекуррентные нейронные сети (RNN) – используются для анализа последовательностей действий пользователя;

– сверточные нейронные сети (CNN) – применяются в анализе изображений, полезны для персонализации графического контента;

– генеративно-состязательные сети (GAN) – позволяют генерировать персонализированные элементы дизайна.

4 Подходы на основе градиентного бустинга

Такие алгоритмы, как XGBoost, LightGBM, CatBoost, эффективно работают с табличными данными и позволяют находить скрытые зависимости в пользовательском поведении. Они применяются для прогнозирования конверсий и анализа влияния изменений в дизайне.

**Реализация задачи исследования.** Для демонстрации предлагаемого подхода была разработана модель на основе нейронной сети, которая анализирует данные о поведении пользователей и адаптирует элементы дизайна веб-сайта в реальном времени. В качестве примера использовались следующие технологии: Python, TensorFlow, Scikit-learn. Данные технологии использовались для реализации модели машинного обучения.

В рамках данной работы был реализован генетический алгоритм для оптимизации параметров персонализации адаптивного дизайна. Также разработана нейросетевая модель для предсказания вероятности конверсии пользователя на основе его поведения на сайте. Хотя интеграция в реальный веб-интерфейс не была осуществлена, данный подход демонстрирует возможность интеллектуального подбора параметров дизайна. Разберем подробнее разработки в рамках данного научного исследования.

1 Генетический алгоритм для персонализации дизайна

Генетический алгоритм (ГА) использовался для оптимизации параметров интерфейса, таких как размер шрифтов, цветовая гамма, расположение элементов и структура навигации. Основные этапы его работы:

Генерация начальной популяции – случайный набор конфигураций дизайна.

Оценка приспособленности (fitness-функция) – каждая конфигурация оценивается по ряду метрик, например, удобочитаемость, визуальное восприятие, предполагаемая вовлечённость пользователя.

Скрещивание и мутации – комбинация лучших конфигураций с небольшими случайными изменениями.

Выбор лучших решений – отбор вариантов, обеспечивающих наилучший пользовательский опыт.

2 Нейросетевая модель для предсказания конверсии

Параллельно с генетическим алгоритмом была создана нейросетевая модель, обученная на данных о пользовательском поведении. Она предсказывает вероятность конверсии (совершения целевого действия) на основе трёх параметров: времени на странице, количества кликов и глубины прокрутки.

Процесс работы модели:

– сбор и обработка данных;

– генерация случайного набора данных (1000 записей) с имитацией пользовательского поведения;

– стандартизация признаков (время на странице, количество кликов, глубина прокрутки);

– разделение данных на обучающую (80%) и тестовую (20%) выборки;

– создание и обучение нейронной сети;

– многослойная нейросеть с двумя скрытыми слоями (16 и 8 нейронов) и сигмоидным выходным слоем;

– оптимизатор Adam, функция потерь – бинарная кросс-энтропия;

– обучение модели на 20 эпохах с батчами по 32 записи;

– оценка точности и предсказания. Средняя точность модели на тестовой выборке – около 0.7.

Разберем подробнее одну из важнейших функций в рамках данной работы. Функция predict_user_experience принимает поведенческие данные пользователя и выдаёт рекомендацию: персонализированный или стандартный дизайн. Генетический алгоритм использует предсказания нейросети для выбора наиболее эффективных параметров дизайна. Например, если модель указывает, что пользователи с высоким временем на странице лучше конвертируются при определённой цветовой схеме, алгоритм закрепляет это правило.

Рассмотрим некоторые этапы в разрезе разработанного программного кода. Как было сказано ранее, для разработки использовался язык программирования Python в связке с фреймворками TensorFlow и Scikit-learn.

1 Сбор данных о поведении пользователей (время на странице, клики, взаимодействия).

Программная реализация данного пункта представлена на рисунке ниже:

```
# Генерация случайных данных
# Данные моделируют поведение пользователей на сайте: время на странице, клики, глубину прокрутки и факт конверсии.
np.random.seed(42)
n_samples = 1000
data = {
    "time_on_page": np.random.normal(60, 15, n_samples),  # Среднее время на странице
    "clicks": np.random.randint(1, 10, n_samples),  # Количество кликов пользователя
    "scroll_depth": np.random.uniform(0.2, 1.0, n_samples),  # Глубина прокрутки
    "conversion": np.random.randint(0, 2, n_samples)  # Факт конверсии (0 или 1)
}
```

*Рисунок 1.* Функция сбора данных о поведении пользователей

2 Обучение модели на исторических данных.

В данном этапе работы, программа разделяет данные на признаки и целевую переменную, затем данные нормализуются для ускорения сходимости модели и повышения точности предсказаний. В ходе работы программы, после нескольких запусков данного кода, средняя точность предсказаний была 0,7, что является приемлемым результатом для конкретного задания. После нормализации данных, они разделяются на обучающую и тестовую выборки. Затем, создается многослойная нейронная сеть для предсказания вероятности конверсии пользователя. Для компиляции модели используется оптимизатор Adam, а в качестве функции потерь используется бинарная кросс-энтропия. Следующим этапом работы является обучение модели на тренировочных данных, после чего формируется оценка точности модели на тестовой выборке.

```
# Разделение данных на признаки (X) и целевую переменную (y)
X = df.drop("conversion", axis=1)
y = df["conversion"]

# Нормализация данных для ускорения сходимости модели и повышения точности предсказаний
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Разделение выборки на обучающую (80%) и тестовую (20%)
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# Создание многослойной нейронной сети для предсказания вероятности конверсии пользователя
model = keras.Sequential([
    keras.layers.Dense(16, activation='relu', input_shape=(X_train.shape[1],)),  # Первый скрытый слой
    keras.layers.Dense(8, activation='relu'),  # Второй скрытый слой
    keras.layers.Dense(1, activation='sigmoid')  # Выходной слой с функцией активации сигмоиды
])

# Компиляция модели с использованием Adam-оптимизатора и бинарной кросс-энтропии в качестве функции потерь
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])

# Обучение модели на тренировочных данных
model.fit(X_train, y_train, epochs=20, batch_size=32, validation_data=(X_test, y_test))

# Оценка точности обученной модели на тестовой выборке
loss, accuracy = model.evaluate(X_test, y_test)
print(f"Модель обучена с точностью: {accuracy:.2f}")
```

*Рисунок 2.* Функции для обучения модели

3 Предсказание предпочтений пользователей в реальном времени.

Программная реализация данного пункта представлена на рисунке ниже:

```
# Функция для предсказания вероятности конверсии пользователя
# На основе его поведения на сайте (время на странице, количество кликов и глубина прокрутки)
def predict_user_experience(time_on_page, clicks, scroll_depth):
    user_data = np.array([[time_on_page, clicks, scroll_depth]])
    user_data_scaled = scaler.transform(user_data)  # Применяем нормализацию
    prediction = model.predict(user_data_scaled)[0, 0]  # Получаем предсказание модели
    return "Персонализированный дизайн активирован" if prediction > 0.5 else "Обычный дизайн"
```

*Рисунок 3.* Функция предсказания предпочтений пользователя

В результате работы генетического алгоритма, описанного выше, разработчик получает вариант дизайна сайта, подходящий конкретному пользователю с определенными параметры. Так, например, на данном этапе работы будет предложено два вида дизайна веб-сайта, а именно – обычный дизайн и персонализированный дизайн. На более продвинутой стадии разработки проекта возможно автоматическое перестроение дизайна веб-сайта под предпочтения конкретного пользователя.

В перспективе возможно расширение возможностей персонализации за счёт:

– интеграции с технологиями дополненной реальности (AR) для адаптации интерфейсов;

– улучшения моделей за счёт мультимодального обучения;

– использования edge computing для персонализации без передачи данных в облако.

**Заключение.** Использование методов машинного обучения позволяет значительно повысить персонализацию адаптивного дизайна веб-сайтов, улучшая взаимодействие пользователей с ресурсом. В будущем возможно дальнейшее развитие данной технологии с учетом расширенного анализа данных и более сложных моделей нейронных сетей.

## Список литературы

[1] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.

[2] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[3] Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.

[4] Chen, J., & Zhang, C. (2021). Personalized Web Experience using Deep Learning. IEEE Transactions on Knowledge and Data Engineering.

## Авторский вклад

**Васяева Татьяна Александровна** – руководство исследованием по применению машинного обучения для разработки адаптивных веб-сайтов.

**Рябко Максим Алексеевич** – постановка задачи исследования, изучение задачи исследования, разработка решения проблемы, тестирование работы программы и анализ полученных результатов.

# APPLYING MACHINE LEARNING TO PERSONALIZE RESPONSIVE DESIGN

***T.A. Vasiaeva***
*Dean of the Faculty of Information Systems and Technologies of DNTU, PhD of Technical sciences, Associate Professor*

***M.A. Riabko***
*Graduate student of the Department of Automated Control Systems, DNTU*

**Abstract.** In today's digital world, personalization plays a key role in creating an effective user experience. The use of machine learning (MO) can significantly improve the responsive design of websites, providing more relevant content and user experience. This article examines the problems of traditional adaptive design, personalization methods using MO, and also suggests a practical implementation of this approach.

**Keywords:** Adaptivity, machine learning, website, genetic algorithm, neural networks.

.

УДК 004.021:004.75

# АДАПТИВНАЯ ФИЛЬТРАЦИЯ МНОГОМЕРНЫХ ДАННЫХ ДЛЯ ОПТИМИЗАЦИИ БИЗНЕС-РЕШЕНИЙ НА ОСНОВЕ НЕЛИНЕЙНЫХ МЕТРИК ВЗАИМОЗАВИСИМОСТИ ПРИЗНАКОВ

***В.Г. Евдокимов***
*Аспирант кафедры информационных технологий автоматизированных систем БГУИР*
vigandvdk@gmail.com

***А.А. Навроцкий***
*Заведующий кафедрой информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент*
navrotsky@bsuir.by

***А.А. Навроцкий***
*Окончил Минский радиотехнический институт. Область научных интересов включает программные системы, системы компьютерного зрения и СВЧ-устройства.*

***В.Г. Евдокимов***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов включает оптимизацию алгоритмов искусственного интеллекта, применение ИИ в условиях неполноты данных и разработку адаптивных автономных интеллектуальных систем.*

**Аннотация.** В статье предлагается концептуально новый подход к адаптивной фильтрации многомерных данных, основанный на переходе от представления данных в форме традиционного гиперкуба к сферической геометрии гиперсферы и обратно. Предложенный метод потенциально может позволить эффективно идентифицировать и исключать взаимозависимые параметры путем анализа их геометрических свойств в сферическом пространстве. Рассматриваются теоретические основы идеи, математический аппарат для оценки нелинейных зависимостей между признаками и потенциальные преимущества данного подхода в различных задачах бизнес-аналитики. Обсуждаются перспективы практической реализации и адаптации метода для конкретных предметных областей.

**Ключевые слова:** многомерные данные, гиперсфера, сферическая геометрия, взаимозависимость признаков, фильтрация данных, бизнес-аналитика, OLAP.

**Введение.** Современные системы бизнес-аналитики сталкиваются с постоянно возрастающими объемами и размерностью обрабатываемых данных. По оценкам IDC, глобальный объем данных к 2026 году превысит 180 зеттабайт, причем значительная часть этого объема будет приходиться на бизнес-данные [1]. Рост размерности данных приводит к увеличению вычислительной сложности аналитических моделей и затрудняет извлечение полезной информации для принятия решений.

Традиционно данные в бизнес-аналитике представляются в виде многомерных структур – гиперкубов, где каждое измерение соответствует определенному параметру или признаку. Такое представление является основой для OLAP-систем (Online Analytical Processing) и широко используется в системах поддержки принятия решений. Однако

стандартные методы работы с гиперкубами имеют существенные ограничения при анализе сложных нелинейных взаимосвязей между признаками.

Эти ограничения связаны с тем, что евклидова геометрия гиперкуба не позволяет в полной мере отразить нелинейный характер взаимодействия признаков в реальных бизнес-процессах. Как отмечают Ли и Верлейсен, «линейные структуры представления данных неадекватны для отражения сложных нелинейных взаимозависимостей, характерных для большинства реальных бизнес-систем» [2].

В данной статье предлагается принципиально новая идея – использовать переход от евклидовой геометрии гиперкуба к сферической геометрии гиперсферы для более эффективной фильтрации параметров и последующего возврата к форме гиперкуба, но уже с исключенными взаимозависимыми параметрами. Данный подход потенциально позволяет решить проблему избыточной размерности данных и повысить эффективность бизнес-аналитики.

**Обзор существующих подходов к анализу многомерных данных**

**Традиционные методы работы с многомерными данными.** В настоящее время наиболее распространенными методами работы с многомерными данными в бизнес-аналитике являются технологии OLAP, основанные на представлении данных в виде гиперкубов. Такое представление позволяет осуществлять многомерный анализ данных, включая различные операции агрегации, детализации, секционирования и поворота.

Однако OLAP-технологии имеют существенные ограничения при анализе сложных взаимозависимостей между параметрами. Как отмечает Ван дер Маатен, «традиционные OLAP-кубы не способны эффективно выявлять нелинейные паттерны во взаимосвязях параметров, что критически важно для многих бизнес-приложений» [3].

**Методы снижения размерности.** Для решения проблемы высокой размерности данных широко используются различные методы снижения размерности, такие как анализ главных компонент (PCA), многомерное шкалирование (MDS), t-SNE и другие. Эти методы позволяют преобразовать исходные данные в пространство меньшей размерности с сохранением ключевых характеристик данных.

Однако большинство существующих методов снижения размерности имеют существенные ограничения:

– линейность преобразований – нелинейные методы снижения размерности часто создают новые признаки, которые сложно интерпретировать в терминах исходных параметров;

– потеря интерпретируемости – MQTT-сервер, который принимает информацию от издателей и передает ее соответствующим подписчикам, в сложных системах может выполнять также различные операции, связанные с анализом и обработкой поступивших данных. Разные брокеры могут соединяться между собой, если они подписываются на сообщения друг друга;

– отсутствие адаптивности – большинство методов не обеспечивают адаптивной фильтрации признаков в зависимости от их взаимозависимости.

Как отмечает Амари в своей работе по информационной геометрии, «существующие методы снижения размерности не учитывают геометрические свойства пространства признаков, что приводит к неоптимальным результатам при анализе сложных данных» [5].

**Постановка проблемы.**

Анализ существующих подходов к обработке многомерных бизнес-данных позволяет сформулировать следующие ключевые проблемы:

**Проблема идентификации взаимозависимых параметров**. В реальных бизнес-данных многие параметры могут быть взаимозависимыми, причем эта зависимость часто имеет нелинейный характер. Традиционные методы, основанные на линейной корреляции, не способны эффективно выявлять такие зависимости. Это приводит к избыточности данных и снижению эффективности аналитических моделей.

**Неэффективность евклидовой метрики для оценки нелинейных взаимосвязей.** Евклидова метрика, используемая в традиционных методах анализа данных, не адекватна для оценки нелинейных взаимосвязей между параметрами. Как показали Баттерворт и соавторы, «евклидовы расстояния могут быть крайне обманчивыми при анализе сложных многомерных данных с нелинейными взаимодействиями» [4].

**Отсутствие единого подхода к фильтрации параметров.** В настоящее время отсутствует единый подход, который бы позволял:

– эффективно выявлять нелинейные взаимозависимости между параметрами;

– адаптивно фильтровать избыточные параметры;

– сохранять интерпретируемость результатов анализа;

– обеспечивать вычислительную эффективность при работе с большими данными.

Предлагаемая в данной статье идея трансформации данных из гиперкуба в гиперсферу с последующей фильтрацией и обратным преобразованием направлена на решение именно этих проблем.

**Концептуальное описание предлагаемого подхода - идея перехода от гиперкуба к гиперсфере и обратно.** Центральная идея предлагаемого подхода заключается в трансформации представления данных из традиционного гиперкуба в сферическое пространство (гиперсферу), где нелинейные взаимосвязи между параметрами могут быть более эффективно выявлены и проанализированы. После выявления и исключения взаимозависимых параметров данные преобразуются обратно в форму гиперкуба, но уже с оптимизированной размерностью.

Этот процесс можно представить в виде следующих этапов:

– формирование исходного гиперкуба данных – данные организуются в многомерную структуру, где каждое измерение соответствует определенному признаку или параметру;

– трансформация в сферическое пространство – каждый параметр представляется как точка на единичной гиперсфере $S^{(n-1)}$, где n – исходная размерность данных;

– анализ геометрических свойств в сферическом пространстве – выявление нелинейных взаимозависимостей между параметрами на основе геометрических характеристик (сферических избытков, углов, расстояний);

– фильтрация взаимозависимых параметров – исключение избыточных параметров на основе выявленных зависимостей;

– обратное преобразование к форме гиперкуба – построение оптимизированного гиперкуба с исключенными взаимозависимыми параметрами.

**Сферическая геометрия как инструмент анализа нелинейных взаимосвязей.** Сферическая геометрия предоставляет уникальные возможности для анализа нелинейных взаимосвязей между параметрами. В отличие от евклидовой геометрии, где прямые линии соответствуют кратчайшим расстояниям, в сферической геометрии кратчайшими путями являются дуги больших кругов. Это принципиально меняет характер метрических отношений между точками.

Ключевым элементом анализа в сферическом пространстве является сферический избыток треугольника, образованного тремя точками на сфере (формула 1):

$$E(p_i, p_j, p_k) = \alpha + \beta + \gamma - \pi, \quad (1)$$

где $\alpha,\ \beta,\ \gamma$ - углы сферического треугольника, образованного параметрами $p_i,\ p_j,\ p_k$.

Сферический избыток служит мерой нелинейности взаимодействия между параметрами. Чем больше сферический избыток, тем сильнее нелинейный характер взаимосвязи.

**Формализация критериев взаимозависимости параметров.** Для формализации критериев взаимозависимости параметров вводится понятие радиуса кривизны локального пространства параметров (формула 2):

$$r(\rho) = \sqrt{[A_{sph}/E]}, \quad (2)$$

где $A_{sph}$ – площадь сферического многоугольника, образованного взаимодействующими параметрами, E – сферический избыток многоугольника.

Параметры считаются взаимозависимыми, если разница их радиусов кривизны меньше определенного порогового значения (формула 3):

$$\frac{|r(\rho_i) - r(\rho_j)|}{\max(r(\rho_i),\ r(\rho_j))} < \varepsilon, \quad (3)$$

где $\varepsilon$ – пороговый параметр, определяющий степень чувствительности метода к выявлению взаимозависимостей.

**Процедура фильтрации параметров гиперкуба.** На основе выявленных взаимозависимостей производится фильтрация параметров гиперкуба. Для этого вводится понятие базовых и производных параметров:

– базовые параметры – минимальный набор параметров, необходимый для полного описания данных с учетом линейных зависимостей;

– производные параметры – параметры, которые могут быть выражены через базовые с учетом нелинейных зависимостей.

Процедура фильтрации заключается в выявлении и исключении производных параметров, которые могут быть выражены через базовые. Математически это можно представить следующим образом (формула 4):

$$N_{filtered} = N_{base} + \sum_{k=2}^{k_{crit}} F(k, r, \varepsilon), \quad (4)$$

где $N_{base}$ – число базовых параметров, F(k, r, ε) – функция, определяющая число независимых параметров порядка k после фильтрации с радиусом кривизны r и порогом $\varepsilon$.

**Алгоритм фильтрации взаимозависимых параметров.** Алгоритм фильтрации взаимозависимых параметров можно представить в виде следующих шагов:

1 Определение базовых параметров
2 Инициализация множества фильтрованных параметров
3 Анализ комбинаций параметров
   a. для каждой комбинации вычислить сферический избыток E;
   b. для каждой комбинации вычислить радиус кривизны r.
4 Кластеризация параметров по радиусу кривизны
5 Фильтрация взаимозависимых параметров
6 Формирование оптимизированного гиперкуба

Этот алгоритм обеспечивает эффективную фильтрацию параметров с учетом их нелинейных взаимозависимостей и позволяет существенно снизить размерность данных без значимой потери информации.

**Потенциальные приложения и преимущества подхода**

**Применение в бизнес-аналитике.** Предлагаемый подход имеет широкий спектр потенциальных приложений в бизнес-аналитике:

– оптимизация OLAP-кубов – снижение размерности OLAP-кубов без значимой потери аналитической информации;

– улучшение прогностических моделей – исключение избыточных параметров повышает точность и стабильность прогнозов;

– оптимизация клиентской аналитики – выявление ключевых параметров, определяющих поведение клиентов;

– анализ цепочек поставок – идентификация критических параметров в сложных логистических системах;

– финансовый анализ – выявление ключевых факторов риска и доходности в финансовых инструментах.

**Преимущества перед существующими методами.** Предлагаемый подход обладает следующими преимуществами перед существующими методами:

– учет нелинейных взаимосвязей – в отличие от линейных методов, таких как PCA, предлагаемый подход эффективно выявляет нелинейные зависимости между параметрами;

– сохранение интерпретируемости – в отличие от многих нелинейных методов снижения размерности, результаты анализа сохраняют прямую интерпретацию в терминах исходных параметров;

– адаптивность – метод автоматически адаптируется к характеристикам конкретного набора данных;

– математическая обоснованность – подход имеет строгое математическое обоснование в рамках сферической геометрии и теории информации.

**Ограничения и направления дальнейшей разработки.** Несмотря на потенциальные преимущества, предлагаемый подход имеет определенные ограничения:

– вычислительная сложность – анализ в сферическом пространстве может требовать значительных вычислительных ресурсов, особенно для данных высокой размерности;

– эмпирический подбор параметров – эффективность метода зависит от правильного выбора параметров, таких как пороговое значение ε;

– необходимость экспериментальной проверки – теоретические преимущества требуют подтверждения на реальных бизнес-данных.

Направления дальнейшей разработки включают:

– разработку эффективных алгоритмов вычисления сферических метрик для больших данных;

– создание методов автоматического подбора оптимальных параметров метода;

– экспериментальную валидацию на различных типах бизнес-данных;

– интеграцию с существующими инструментами бизнес-аналитики.

**Заключение.** В данной статье предложена инновационная идея адаптивной фильтрации многомерных данных на основе перехода от представления данных в форме гиперкуба к сферической геометрии гиперсферы и обратно. Этот подход позволяет эффективно выявлять и исключать взаимозависимые параметры, учитывая их нелинейные взаимосвязи, что потенциально может значительно повысить эффективность бизнес-аналитики.

Ключевыми элементами предложенного подхода являются:

– трансформация представления данных из гиперкуба в гиперсферу;

– анализ нелинейных взаимосвязей на основе сферической геометрии;

– фильтрация взаимозависимых параметров с использованием критерия близости радиусов кривизны;

– обратное преобразование к гиперкубу с оптимизированной размерностью.

Данный подход имеет прочную математическую основу и потенциально может преодолеть ограничения существующих методов анализа многомерных данных. Однако его практическая эффективность требует дальнейшей экспериментальной проверки и разработки эффективных алгоритмов реализации.

Перспективы дальнейших исследований включают разработку конкретных алгоритмов и программных инструментов для реализации предложенного подхода, его интеграцию с существующими платформами бизнес-аналитики и валидацию на различных типах бизнес-данных.

## Список литературы

[1] World Bank. Cloud Computing: Growth Drivers, Main Players, and Key Trends. Chapter 4 in Advancing Cloud and Data Infrastructure Markets. Washington, DC: World Bank, 2021. – 49 с.

[2] Holmström, Lasse. Review of Nonlinear Dimensionality Reduction by John A. Lee and Michel Verleysen. International Statistical Review, February 2008, 76(2):308–309.

[3] Van Der Maaten L., Hinton G. Visualizing Data using t-SNE // Journal of Machine Learning Research. 2008. Vol. 9. P. 2579-2605

[4] Butterworth R., Piatetsky-Shapiro G., Simovici D.A. On Feature Selection through Clustering // IEEE International Conference on Data Mining. 2015. P. 581-584.

[5] Amari S., Nagaoka H. Methods of Information Geometry. American Mathematical Society, 2007. – 206 p.

**Авторский вклад**

**Навроцкий Анатолий Александрович** – руководство исследованием и постановка задачи. Общее руководство научным проектом, формулировка научной проблемы.

**Евдокимов Виталий Геннадьевич** – разработка концепции, алгоритмического обеспечения и анализ практического применения.

# ADAPTIVE FILTERING OF MULTIDIMENSIONAL DATA FOR BUSINESS DECISION OPTIMIZATION BASED ON NONLINEAR METRICS OF FEATURE INTERDEPENDENCE

***V.G. Evdokimov***
*Postgraduate student of the Department of Information Technologies of Automated Systems, BSUIR*

***A.A. Navrotsky***
*Head of the Department of Information Technologies of Automated Systems, BSUIR, PhD in Physics and Mathematics, Associate Professor*

**Abstract.** The article proposes a conceptually new approach to adaptive filtering of multidimensional data, based on the transition from data representation in the form of a traditional hypercube to the spherical geometry of a hypersphere and back. The method allows to effectively identify and exclude interdependent parameters by analyzing their geometric properties in spherical space. The theoretical foundations of the idea, mathematical apparatus for evaluating nonlinear dependencies between features, and potential advantages of this approach in various business analytics tasks are considered. The prospects for practical implementation and adaptation of the method for specific subject areas are discussed.

**Keywords**: multidimensional data, hypercube, hypersphere, spherical geometry, feature interdependence, data filtering, business analytics, OLAP.

УДК 621

# РАЗВИТИЕ ПОДХОДА К ОЦЕНКЕ ПАРАМЕТРОВ МЕХАНИЧЕСКОЙ СИСТЕМЫ МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

***Д.А. Кечик***
*Научный сотрудник Лаборатории цифровых прикладных технологий НГУ, м.т.н.*

***А.С. Кожухов***
*Научный сотрудник Лаборатории цифровых прикладных технологий НГУ, к.ф.-м.н.*

***Ю.Д. Кацер***
*Научный сотрудник Лаборатории цифровых прикладных технологий НГУ, м.т.н.*

***Д.А. Кечик***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов цифровой обработки сигналов, математической статистики и машинного обучения.*

***А.С. Кожухов***
*Окончил Новосибирский государственный университет. Область научных интересов связана с разработкой методов и алгоритмов математической статистики, машинного обучения и искусственного интеллекта.*

***Ю.Д. Кацер***
*Окончил МГТУ имени Баумана. Область научных интересов связана с разработкой методов и алгоритмов математической статистики, машинного обучения и искусственного интеллекта.*

**Аннотация.** В работе рассматриваются методы математического моделирования роторного оборудования и оценки его параметров по сигналам вибрации для его диагностирования. Выявлены достоинства и недостатки нового подхода к моделированию, предложены и обоснованы пути его дальнейшего развития.

**Ключевые слова:** АЧХ, гармонический баланс, вибрационная диагностика, машинное обучение, метод наименьших квадратов, метод полных наименьших квадратов.

**Введение.** Методы неразрушающего контроля, к которым относится и вибродиагностика, эффективно применяются для снижения затрат на обслуживание оборудования. Для повышения точности диагностики вращающегося оборудования нужно различать дефекты, возникающие в процессе его эксплуатации, и оценивать степень их опасности. Значительную часть дефектов большинства узлов оборудования можно обнаружить и различить при помощи анализа амплитудного спектра вибрации. Данный метод основан на периодичности возникающих в работающем оборудовании сил [1].

Дефекты валов и их соединений приводят к появлению колебаний на частотах, кратных частоте вращения вала. Обычно присутствуют первые 3-4 гармоники, реже 10 гармоник оборотной частоты. По наличию гармонических составляющих вибрации на расчётных частотах – по превышению их амплитудами определённых порогов, делают вывод о наличии неисправности. Но сделать вывод о виде дефекта валов или их сборки можно опираясь на диагностические правила, которые обычно рассматривают соотношения амплитуд гармоник оборотной частоты. Автоматизировать принятие решения

о наличии и виде неисправности возможно путём выбора амплитудных порогов, опираясь на знание кинематической схемы оборудования и расчёты или опыт эксплуатации. При этом, амплитуды гармонических составляющих отклика системы, т.е. вибрации, зависят как от амплитуд возбуждающих сил, так и от параметров механической системы. Это сильно затрудняет выставление амплитудных порогов: близость резонансной частоты к частоте одной из гармоник может привести к усилению отклика на данной частоте и постановке ложного диагноза.

Гораздо сложнее диагностика оборудования, работающего в нестационарном режиме, такого как ветрогенераторы или газотурбинные двигатели. При значительном изменении частоты вращения вала, амплитуды составляющих вибрации вследствие неравномерности АЧХ тоже становятся непостоянными. При этом, амплитуды возбуждающих сил постоянны, либо зависят от частоты вращения вала по известному закону.

Изменение состояния механической системы отражаются либо как возникновение в системе сил или изменение их амплитуды, либо как изменение механических параметров самой системы. Последнее приводит к изменению амплитудно-частотной характеристики, или к более сложным, нелинейным и параметрическим эффектам. Амплитуды сил, как и механические параметры работающего оборудования, недоступны для наблюдения или вычисления. Поэтому, было предложено оценивать соотношения амплитуд действующих в системе сил и использовать их для диагностирования. Далее рассмотрим методы моделирования механических систем, в том числе нелинейных, метод оценки соотношения амплитуд, и пути развития данного подхода.

**Физическое моделирование агрегата.** Главным образом, механические системы моделируются при помощи дифференциального уравнения Лагранжа, которое связывает перемещение и две его производные (скорость и ускорение) с воздействующими на систему силами. В случае наличия в системе геометрической нелинейности, неисправность моделируется как воздействие на линейную систему периодических сил на кратных частотах [1–5]. Реальный агрегат может моделироваться как система с распределёнными параметрами [6], так и состоящая из нескольких элементов со сосредоточенными параметрами [1–5]. В обоих случаях, агрегат понимается как линейная стационарная система, входным воздействием на которую являются силы, возникающие вследствие наличия дефекта, а откликом – вибрация. Такая трактовка возможна в случае наличия в системе геометрической нелинейности: жёсткость и затухание не зависят от перемещения или от времени, но дополнительные силы возникают вследствие деформации системы. Система с расцентрованными валами моделируется как линейная механическая система, в которой действуют силы на гармониках оборотной частоты [1–5]. Такая система может быть описана при помощи её амплитудно-частотной характеристики (АЧХ). Эта характеристика зависит от параметров её элементарных звеньев - их жёсткости, упругости, массы, которые как правило неизвестны. Подобные модели используются для расчёта амплитуд вибрации при различных величинах дефекта и частотах вращения вала. Для этого дифференциальное уравнение решается любым методом интегрирования.

Схожим образом моделируется механическая система и в более общем случае: тогда в уравнение добавляются слагаемые, в которых жёсткость какого-либо элемента зависит от угла поворота вала или от перемещения узла. Такие системы возбуждаются внешними силами другой природы, например, силами дисбаланса. Периодическим изменением жёсткости моделируются трещины вала [7–9]: вал имеет минимальную жёсткость при раскрытии трещины под действием внешней силы, и максимальную - при её закрытии. В таком случае, имеет место параметрическое возбуждение колебаний на разных частотах. Задевания ротора [10] или ослабления [11, 12] его посадки моделируются как зависимость жёсткости от перемещения: жёсткость соединения ротора с опорой резко увеличивается при его столкновении с ней. Таким образом, появляется нелинейная упругость. В обоих случаях, нельзя описывать систему в терминах АЧХ: ведь её отклик на гармоническое

воздействие содержит несколько гармоник. В таком случае, рассчитывают перемещение при различных частотах вращения вала. Из него - получают спектральный состав вибрации, орбиту движения центра вала (при наличии двух координат перемещения), собственные частоты конструкции, а также их смещения и появления их гармоник вследствие появления дефекта [7, 11, 12].

Перемещение получают как интегрированием ДУ, так и методом гармонического баланса [7, 9]. Данный метод предполагает приравнивание комплексных амплитуд сил на каждой частоте к комплексным амплитудам перемещения на тех же частотах. Метод позволяет не только получить аналитические выражения для амплитуд отклика, но и выделить стабильные и переходные режимы колебаний [13, 14]. При наличии нелинейностей, может возникать явление нелинейного резонанса. В таком случае, при одних и тех же значениях частоты вращения вала в области такого резонанса может наблюдаться одно из двух значений амплитуды колебаний в устоявшемся режиме. Смена режима колебаний происходит при изменении частоты вала через переменный режим. Метод гармонического баланса позволяет построить такую частотную характеристику, которая бы учитывала все возможные режимы колебаний.

Ввиду недоступности для измерения входного воздействия, оценить непосредственно частотную характеристику или все её параметры на основании одной лишь вибрации невозможно. Для проведения таких оценок используют либо моделирование методом конечных элементов, либо используя заданное механическое воздействие на агрегат [15]. Однако, далеко не всегда оказывается возможным подобное точное моделирование или возбуждение конструкции внутри сложного агрегата. Желательно также проводить диагностику агрегата в ходе его работы. Тогда, вместо зависимости отношения выходной амплитуды ко входной от частоты, оценивается зависимость СКЗ широкополосной вибрации или одной из гармонических составляющих. Построение таких зависимостей используется в двух целях: для устранения влияния неравномерности АЧХ агрегата на уровень вибрации или её гармонических составляющих, либо для обнаружения изменений состояния оборудования.

Если частоты рассматриваемых компонентов вибрации расположены на равномерном участке АЧХ, то можно принять, что их амплитуды равны амплитудам действующих в системе сил, умноженным на постоянный коэффициент. В случае близости частоты механического резонанса к частоте одной из гармоник, её амплитуда может изменяться в десятки раз. Это может привести к неверным выводам, даже если частота вала не меняется в рабочем режиме агрегата. Диагностика оборудования с переменной частотой вращения вала дополнительно затрудняется нестационарностью амплитуд гармонических компонентов вибрации. Поэтому целью работы является разработка и реализация метода анализа вибрационных сигналов для повышения точности автоматической диагностики по амплитудным спектрам вибрации.

В работе [16] оценивается зависимость СКЗ фильтрованной в трёх частотных диапазонах вибрации от частоты вращения вала. В результате ряда измерений вибрации при переменной известной в каждый момент частоте вала, получаются оценки СКЗ в коротком временном окне, и каждой такой оценке соответствует средняя частота вала. Эмпирическая зависимость аппроксимируется полиномом, коэффициенты которого запоминаются и используются в дальнейшем для устранения зависимости СКЗ вибрации в каждой полосе частот от частоты вала. Метод используется для обработки нестационарных вибросигналов методами спектрального анализа вибрации.

Для поиска резонансных областей применяется [17] построение зависимости СКЗ колебаний на оборотной частоте от частоты вращения вала. Метод не предусматривает построение математической модели. Предполагается, что информация о выделенных визуальным осмотром графиков резонансных областях будет использоваться в целях виброзащиты оборудования.

Изменение состояния оборудования оценивается [18] по изменению графиков зависимости СКЗ первой гармоники вибросигнала от частоты вращения вала, снятых в разные периоды времени. Предполагается прежде всего экспертный анализ таких зависимостей. Обнаруженные отличия зависимостей связываются с изменением состояния оборудования, а не с каким-либо его дефектом.

На данный момент, область остаётся мало разработанной. Это может объясняться трудностями, связанными с отсутствием возможности наблюдать входное воздействие. Отсутствует физическая и математическая модель, которая связывает изменение состояния оборудования и параметров такой зависимости. Существующих моделей, описывающих зависимости СКЗ от частоты вала, недостаточно для повышения точности диагностирования оборудования. Математический аппарат для проведения моделирования также нуждается в доработке.

В предыдущих наших работах [19, 20] был предложен способ оценки параметров модели, которая описывает внешние гармонические силы и линейную механическую систему, на которую они воздействуют. Предложено оценивать полюса передаточной функции и произведения её нулей на комплексные амплитуды сил. Из таких оценок можно получать соотношения амплитуд сил на разных гармониках частоты вала, что в дальнейшем может использоваться для диагностирования роторного оборудования.

Для этого составляется уравнение авторегрессии, затем применяется метод наименьших квадратов:

$$\dot{X}\, s^n = -b_{n-1} s^{n-1}\, \dot{X} - \ldots - b_0 s^0\, \dot{X} + \dot{F}\, a_m s^m + \ldots + \dot{F}\, a_0 s^0 \tag{1}$$

Отсюда в матричном виде:

$$\begin{gathered} \boldsymbol{b} = [B \mid A]\boldsymbol{x}, B = [\boldsymbol{B}_R^T \mid \boldsymbol{B}_I^T]^T, \boldsymbol{B}_R = \Re\{\dot{\boldsymbol{B}}\}, \boldsymbol{B}_I = \Im\{\dot{\boldsymbol{B}}\}, A = [\boldsymbol{A}_R^T \mid \boldsymbol{A}_I^T]^T, \\ \boldsymbol{A}_R = \Re\{\dot{\boldsymbol{A}}\}, A_I = \Im\{\dot{\boldsymbol{A}}\}, \dot{\boldsymbol{A}} = [s^m, \ldots, s^0], \dot{\boldsymbol{B}} = X[s^n, \ldots, s^0] \end{gathered} \tag{2}$$

где $s = j\omega$, $\omega$ – циклическая частота гармонической силы и отклика, $X$ – комплексная амплитуда отклика. Вектор $\boldsymbol{x}$ содержит полюса частотной передаточной функции системы и произведения её нулей с амплитудой гармонической силы. Матрица наблюдений $[B|A]$ формируется из строк, соответствующих мнимым и действительным частям комплексных амплитуд отклика. Для получения точных оценок, целесообразно использовать большой объём измерений и технологии обработки информации, используемые в *Big Data*. Также необходимо учитывать, что строки матрицы должны быть линейно независимы, что обеспечивается в случае, когда измерения проводятся на разных частотах вращения вала.

Достоинства данного способа:

- модель отражает физическое устройство агрегата;
- возможна оценка параметров нескольких перекрывающихся резонансных пиков по измерениям на отдельных точках на частотной оси.

Столбцы матрицы наблюдений содержат ошибки оценки частоты гармоник. Столбцы, зависимые от комплексных амплитуд составляющих вибрации, содержат также ошибки оценки амплитуд. В таком случае оценки, полученные методом наименьших квадратов, являются смещёнными. При этом применение метода полных наименьших квадратов, обычно используемого в таких случаях, привело к ухудшению результата, которое было меньше, если предварительно столбцы матрицы наблюдений отмасштабировать. Предполагается, что причина наблюдаемого эффекта в наличии корреляции ошибок в столбцах матрицы наблюдения.

**Дальнейшее развитие работы.** Для устранения указанных недостатков, а также применения подхода для более широкого набора случаев, предлагаются следующие этапы разработки.

Систему (2) можно использовать для получения оценок при условии, что частота вращения вала изменяется очень медленно на протяжении временного окна, на котором строят оценку амплитуд спектральных составляющих, либо при получении мгновенных оценок амплитуд и частот составляющих вибрации. Предлагается использовать мгновенные оценки гармонических параметров вибрации, полученной в ходе разгона и выбега оборудования для получения оценок параметров механической системы.

Аппроксимация системы уравнений (2) системой, в которой требуемые параметры (резонансные частоты и добротности элементов) входят в явном виде. Чтобы учесть более общий случай нелинейных систем, аппроксимирующая функция должна учитывать асимметрию резонансного пика и резкую смену режима колебаний при изменении частоты вращения вала.

Разработка алгоритма оценки параметров системы, основанного на максимальном правдоподобии, что позволит учесть известное распределение ошибок коэффициентов матрицы наблюдений и их корреляцию. На основании известного распределения оценок параметров регрессии, возможно разработать статистические тесты для проверки гипотез о состоянии системы.

Для выявления в системе нелинейных эффектов, предлагается разработать следующие статистические тесты: проверять наличие отклонения формы резонансного пика, статистически значимые отличия от нуля коэффициентов асимметрии; проверять наличие гистерезиса при нелинейном резонансе по наличию статистически значимых отличий коэффициентов формы, полученных при разгоне и выбеге агрегата.

Для автоматизации выявления и различения дефектов по сигналам вибрации, разработать метод расчёта амплитудных порогов. Для распространённых типов нелинейностей составить уравнение гармонического баланса либо модель геометрической нелинейности, которые связывают амплитуды гармоник вибрации с величинами дефекта (смещением валов, размером трещины и т.п.). По известным распределениям оцениваемых параметров, полученных на предыдущем этапе, составить статистические тесты для различения и оценки размеров дефектов механического оборудования.

**Заключение.** Рассмотрен метод оценки некоторых параметров механических систем и соотношения амплитуд гармонических сил, возникающих вследствие неисправности. Предложены пути улучшения данного метода и развития подхода к диагностированию роторного оборудования.

## Список литературы

[1] Haifei, W. Dynamic analysis of coupling misalignment and unbalance coupled faults / W. Haifei, J. Gong // Journal of Low Frequency Noise, Vibration and Active Control. – 2019. – Т. 38. – С. 14.

[2] Gibbons, C.B. Coupling Misalignment Forces / C.B. Gibbons // Proceedings of the Fifth Turbomachinery Symposium. – Gas Turbine Laboratories, Texas A&M University, College Station, Texas, 1976. – P. 111-116.

[3] Dynamic Response of a Rotating Assembly under the Coupled Effects of Misalignment and Imbalance / M. Desouki [et al.] // Shock and Vibration. – 2020. – Vol. 2020. – P. 26.

[4] Lees, A.W. Misalignment in rigidly coupled rotors / A.W. Lees // Journal of Sound and Vibration. – 2007. – Vol. 305, № 1-2. – P. 261-271.

[5] Sekhar, A.S. Effects of coupling misalignment on vibrations of rotating machinery / A.S. Sekhar, B.S. Prabhu // Journal of Sound and Vibration. – 1995. – Vol. 185, № 4. – P. 655-671.

[6] Tuckmantel, F.W.D.S. Vibration signatures of a rotor-coupling-bearing system under angular misalignment / F.W.D.S. Tuckmantel, K.L. Cavalca // Mechanism and Machine Theory. – 2019. – Vol. 133. – P. 559-583.

[7] Sinou, J.-J. The influence of cracks in rotating shafts / J.-J. Sinou, A.W. Lees // Journal of sound and vibration. – 2005. – Т. 285, № 4-5. – С. 1015-1037.

[8] Prabhakar, S. Crack versus coupling misalignment in a transient rotor system / S. Prabhakar, A.S. Sekhar, A.R. Mohanty // Journal of Sound and Vibration. – 2002. – Vol. 256, № 4. – P. 773-786.

[9] Sinou, J.-J. A non-linear study of a cracked rotor / J.-J. Sinou, A.W. Lees // European Journal of Mechanics-A/Solids. – 2007. – Т. 26, № 1. – С. 152-170.

[10] Effects of misalignment on the nonlinear dynamics of a two-shaft rotor-bearing-gear coupling system with rub-impact fault / X. Lu [et al.] // J. vibroeng. – 2017. – Vol. 19, № 8. – P. 5960-5977.

[11] Nonlinear vibration characteristics of a rotor system with pedestal looseness fault under different loading conditions / H. Ma [et al.] // . VOLUME. – Vol. 15, № 1.

[12] Liang, Q. Dynamic Performance of Time-Domain Piecewise Linear Stiffness System / Q. Liang, L. Li // Journal of Engineering Mechanics. – 2021. – Т. 147. – С. 04021037.

[13] HarmonicBalance.jl: A Julia suite for nonlinear dynamics using harmonic balance / J. Košata [et al.] // SciPost Phys. Codebases. – 2022. – P. 6.

[14] Harmonic Balance Methods: A Review and Recent Developments / Z. Yan [et al.] // Computer Modeling in Engineering & Sciences. – 2023. – Vol. 137, № 2. – P. 1419-1459.

[15] Дмитриев, Н.В. Анализ методов для определения параметров амплитудно-частотной характеристики балки / Н.В. Дмитриев, Д.А. Козлов // Инженерный вестник Дона. – 2023. – № 2. – С. 10.

[16] Кечик, Д.А. Компенсация изменения мощности виброакустического сигнала при изменении скоростного режима оборудования / Д.А. Кечик // Доклады БГУИР. – 2020. – Т. 18, № 5. – С. 27-34.

[17] Алейников, Д.П. Исследование и визуализация амплитудно-частотных характеристик обрабатывающих центров / Д.П. Алейников, А.В. Лукьянов, Д.А. Лукьянов // Современные технологии. Системный анализ. Моделирование. – 2017. – № 1 (53). – С. 58-66.

[18] Бранцевич, П.Ю. Оценка технического состояния механизмов с вращательным движением на основе анализа вибрационных характеристик пусков и выбегов / П.Ю. Бранцевич. – 2021. – 236 с.

[19] Д.А. Кечик, А.С.Кожухов, Ю.Д.Кацер, Применение машинного обучения для оценки параметров АЧХ по отклику, Сборник трудов Международной научно-практической конференции «Методы обработки сигналов в диагностике сложных систем», Минск, 2024 (в печати)

[20] Д.А. Кечик, А.С.Кожухов, Ю.Д.Кацер, Повышение точности автоматического диагностирования вращающегося оборудования по амплитудным спектрам вибрации методами машинного обучения, Прикладная механика и техническая физика, Новосибирск, 2025 (в печати)

## Авторский вклад

**Кечик Даниил Александрович** – разработка математической модели, оценка её применимости.

**Кожухов Антон Сергеевич** – постановка задачи исследования, анализ полученных результатов.

**Кацер Юрий Дмитриевич** – руководство направлением исследования, оценка результатов, редактура статьи.

# DEVELOPMENT OF AN APPROACH TO ESTIMATING THE PARAMETERS OF A MECHANICAL SYSTEM USING MACHINE LEARNING METHODS

***D.A. Kechik***
*Researcher at the NSU Laboratory of Digital Applied Technologies, M.Sc.*

***A.S. Kozhukhov***
*Researcher at the NSU Laboratory of Digital Applied Technologies, Ph.D.*

***Yu.D. Katser***
*Researcher at the NSU Laboratory of Digital Applied Technologies, M.Sc.*

**Abstract.** The paper discusses methods for mathematical modeling of rotary equipment and estimating its parameters based on vibration signals for its diagnosis. The advantages and disadvantages of the new approach to modeling are revealed, and the ways of its further development are proposed and justified.

**Keywords**: frequency response, harmonic balance, vibration diagnostics, machine learning, least squares method, total least squares method.

УДК 004.021:004.75

# ПОСТРОЕНИЕ ГРАФИЧЕСКОЙ ОБОЛОЧКИ ДЛЯ ПРОВЕДЕНИЯ ЭКСПЕРИМЕНТОВ НА БАЗЕ СИСТЕМЫ ЛОГИЧЕСКОЙ ОПТИМИЗАЦИИ ФУНКЦИОНАЛЬНО-СТРУКТУРНЫХ ОПИСАНИЙ ДИСКРЕТНЫХ УСТРОЙСТВ

***И.П. Логинова***
*Старший научный сотрудник ОИПИ НАНБ,*
*кандидат технических наук, доцент*
irilog@mail.ru

***И.П. Логинова***

*Окончила Белорусский государственный университет, кандидат технических наук, доцент. Работает в ОИПИ НАН Беларуси в должности старшего научного сотрудника и в Белорусском государственном университете информатики и радиоэлектроники в должности доцента.*

*Круг научных интересов: программирование, логическое проектирование и верификация цифровых схем, виртуализация, реализация параллельных алгоритмов.*

**Аннотация.** В статье описывается функционирование графического интерфейса, предназначенного для визуального формирования схемы вычислительного эксперимента в виде цепочек действий, предлагаемых элементами управления интерфейса. Визуальная среда программного средства позволяет проводить разнообразные эксперименты с использованием набора проектных операций системы логической оптимизации функционально-структурных описаний дискретных устройств FLC-2. Средствами графического интерфейса формулируется схема эксперимента, в процессе составления которой проводится выбор: программ и объектов, участвующих в эксперименте, Затем пользователь запускает составленную схему эксперимента, отслеживает результат встроенными инструментами визуализации и верификации. В состав программного средства включена также подсистема пакетирования, в которой средствами интерфейса формируются задания для параллельного выполнения логической оптимизации в объектах, собранными в пакеты.

**Ключевые слова:** многоядерный процессор, параллельные процессы, планировщик задач, логическая оптимизация, визуализация.

**Введение.** Важным этапом в автоматизированных системах проектирования является этап проектирования, в котором проводится синтез логические схем, реализующих функциональные комбинационные блоки цифровых СБИС в различных технологических базисах. Первым и особенно трудоемким этапом синтеза схемы является технологически независимая оптимизация, которая в дальнейшем и определяет такие характеристики логической схемы: площадь (число транзисторов), быстродействие, энергопотребление.

Система логической оптимизации функционально-структурных описаний цифровых устройств FLC-2 была описана в работах [1]. В качестве исходных данных для этой системы выступает функциональное либо иерархическое функционально-структурное описание проектируемой логической схемы. Результатом является оптимизированное описание, на основе которого можно получить логическую схему меньшей сложности и большего быстродействия. В основе системной организации FLC-2 лежит принцип модульной

организации выполнения проектных процедур. Взаимодействие модулей осуществляется на языке управления консольными приложениями, выполняющими проектные процедуры [1]. На разных этапах проведения оптимизации логическая схема представляется в виде различных описаний, преобразование которых осуществляется методами детализации, экстракции, оптимизации. При проведении этих преобразований учитываются характеристики и состояние логической схемы. С учетом этих атрибутов формируются цепочки проектных процедур и операций, представляющих маршрут проектирования, где программные модули реализации проектных процедур унифицированы по способу вызова и возврата, что позволяет организовывать их автоматический вызов. Рост размерности СБИС, уменьшение технологических норм приводит к усложнению процесса проектирования и подводит к необходимости организации параллельных вычислений в САПР, как одного из основных способов уменьшения времени проектирования. Возможность распараллеливания прикладных задач логического проектирования определяется рядом факторов: реализацией параллельных алгоритмов для решения прикладных задач; возможностью построения эффективных параллельных программ, реализующих эти алгоритмы; наличием технологий для реализации параллельных программ на многопроцессорной (многоядерной) архитектуре. Как показали многочисленные исследования [2-4], разработка параллельных алгоритмов для большинства логико-комбинаторных задач проектирования цифровых схем, в принципе, невозможна. Тем не менее, один из альтернативных вариантов организации параллельных вычислений на многоядерной вычислительной системе, не требующий разработки параллельных алгоритмов, реализован в системе логической оптимизации функционально-структурных описаний дискретных устройств FLC-2 [5].

В рамках системы FLC-2 реализована интеллектуальная поддержка проектирования, основанная на применении продукционно-фреймовой модели представления знаний [6], имеются свои механизмы управления процессами проектирования (механизмы продукций и стратегий), которые позволяют сократить число однотипных проектных действий, выполняемых проектировщиком, в ряде случаев значительно улучшить качество решений за счет применения комбинированных алгоритмов оптимизации. Расширение возможностей системы FLC-2 путем включения в состав проектных процедур оптимизации, которые выполняются с использованием параллельных вычислений, позволяет существенно сократить время проектирования, и, таким образом, повысить размерность решаемых задач [6].

В основу организации параллельных вычислений в FLC-2 лежит одновременный запуск на ядрах, так называемых, процессов исполнения. В каждом из процессов выполняется процедура (например, проводится оптимизация) с отдельными частями описания схемы, представленной в исходном проекте. Разбиение на части исходного проекта производится применением ряда операций. При необходимости проводятся проектные операции по выявлению и устранению информационных зависимостей между полученными в результате разбиения частями исходного объекта, что приводит к их преобразованию в независимые объекты. Независимость частей проекта , в контексте параллельной реализации означает, что входные аргументы для работы с одной частью проекта не определяются результатами вычислений с другими частями проекта. В итоге, появляется возможность проведения параллельных вычислений с частями проекта при выполнении некоторых проектных операций без блокировок и синхронизации [7].

Организация параллельной работы в системе FLC-2 представлена в работах [5-8], выполнена с использованием технологии OpenMP, которая является стандартом для реализации многопоточных приложений и содержит набор различных директив компилятора, функций соответствующей библиотеки и переменных окружения [9].

Следует отметить, что в системе FLC-2 на всех этапах проектирования проводятся преобразования только с одним проектом, который представлен описанием одной логической схемы. В этой системе отсутствует возможность одновременного проведения

однотипных проектных операций с несколькими описаниями логических схем. Потребность в массовом проведении однотипных проектных операций одновременно для нескольких описаний логических схем привело к запросу на разработку программной среды, в которой можно было бы проводить вычислительные эксперименты с разными проектами, с разными программами, представляющими проектные операции оптимизации на разных операционных платформах. При этом такая проектная среда должна предоставлять средства верификации и визуализации результатов вычислительных экспериментов и обладать удобной графической оболочкой с с интеллектуальной поддержкой организации экспериментов.

**Архитектура графического интерфейса визуальной среды.** Поэтому для системы FLC-2 был разработано автономное программное средство, графическая оболочка, которого представляет визуальную среду с многофункциональным графическим интерфейсом, предназначенную для проведения экспериментов. Возможности интерфейса позволяют проводить эксперименты для потока проектов, формировать пакеты, проводить в параллельные вычисления. Программное средство (далее, ПС PLUS) позволяет исследовать временные показатели выполнения различных программ оптимизации, получать оценки загрузки вычислительных ядер процессора при проведении параллельных вычислений, проводить верификацию. На рисунке 1 представлен состав функциональных модулей, управлением которыми осуществляют средства ПС PLUS.

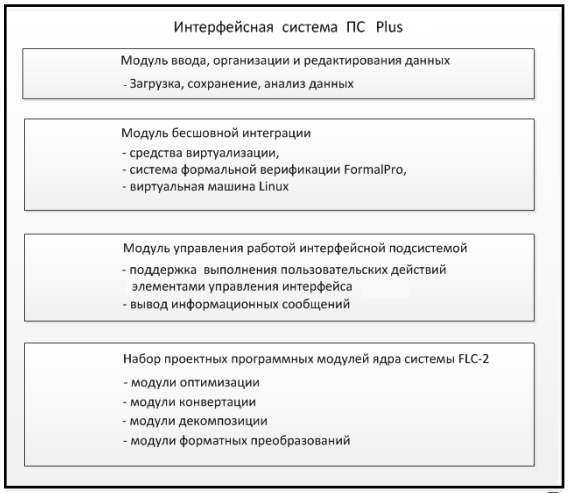



*Рисунок 1.* Общая структура программного средства Plus

Интерфейс ПС PLUS предоставляет инструментарий для организации пакетирования, когда пакеты представляют собой группы заданий, каждое из которых реализует свою операцию оптимизации с одним из проектов, включенных в пакет. Пакет формируется в виде коллекции независимых объектов, получаемых в результате декомпозиции описаний логических схем, представленных в каждом проекте пакета. Преобразование в независимые объекты проводится посредством процедур, входящих в ядро системы FLC-2. В качестве исходных данных, как для системы FLC-2, так и для визуальной среды PLUS выступают функциональные либо иерархические функционально-структурные описания проектируемых логических схем на языке SF [3]. Схема на языке SF определяется иерахической последовательностью функционально-структурных описаний подсхем (блоков). Функциональные описания листовых блоков дерева иерархии описания представляют собой либо логические уравнения (LOG-формат), либо матричные формы представления систем булевых функций в виде ДНФ (SDF-формат). Изменение иерархии SF-описаний и форматные преобразования позволяет переходить к реализации описания схемы в виде одноуровневой структуры описаний подсхем (2-Connect формат) [10]. При

проведении такого преобразования устраняются (элиминируются) все промежуточные внутренние переменные.

Графическая оболочка программного средства PLUS представлена экранной формой в виде окна, обеспечивает пользователю возможность проводить эксперименты, при этом использует функциональные и проектные модули системы FLC-2, которые включены в состав ядра программного среды. На рисунке 2 показан вид окна интерфейса ПС PLUS.

*Рисунок 2.* Главное окно графического интерфейса программного средства PLUS

Многофункциональный графический интерфейс позволяет существенно упрощать, и, в некоторой степени, автоматизировать проведение экспериментов с программами оптимизации путем визуально-интерактивного формирования элементами управления интерфейса цепочек действий. Одни цепочки действий позволяют конструировать пакет из разных исходных описаний логических схем. Другие цепочки действий определяют работу с набором функциональных модулей, управляющими преобразованиями и форматами данных. В окне интерфейса, представленном на рисунке 2, элементы управления сгруппированы по четырем секциям (*a*, *b*, *c*, *d*). Цепочки действий в каждой секции далее рассматриваются по отдельности.

**Организация визуально-интерактивной навигации в графическом интерфейсе.** Пользователю предоставляется следующий функционал: выбор ОС (Windows или Linux), выбор различных программ логической оптимизации, формирование пакета, выполнение задания для одного проекта и заданий для всех проектов пакета в параллельном и последовательном режимах, отслеживание результатов оптимизации с запуском программ сбора статистики, визуализации и верификации. При выполнении параллельных вычислений визуализация трассы параллельных процессов, позволяет увидеть порядок выполнения процессов и распределение вычислительной нагрузки по ядрам процессора.

Навигация, осуществляемая посредством последовательной активацией элементов управления графического интерфейса ПС PLUS, предоставляет следующие возможности:

– позволяет проводить эксперименты с программами оптимизации в двух ОС;

– доступ к различным форматам и конвертерам функционально-структурных описаний дискретных устройств системы FLC-2;

– запускать исполнение программ оптимизации в параллельных процессах, в том числе из коллекции программ оптимизации, находящихся в открытом доступе;

– не отвлекаться на задачи чтения и сохранения результатов;

– получать оценку времени выполнения операций оптимизации в протоколе или в трассе параллельных процессов, в том числе получить визуальный представление этой оценки;

– сравнить эффективности программ оптимизации, выполняемых в различных ОС;

– импортировать и использовать файл конфигурации, системы FLC-2 в котором хранятся параметры конфигурации, например, параметры программ оптимизации;

– формировать структуру и порядок выполнения пакета заданий.

На рисунке 3 приведена схема организации работы в каждой секции интерфейса.

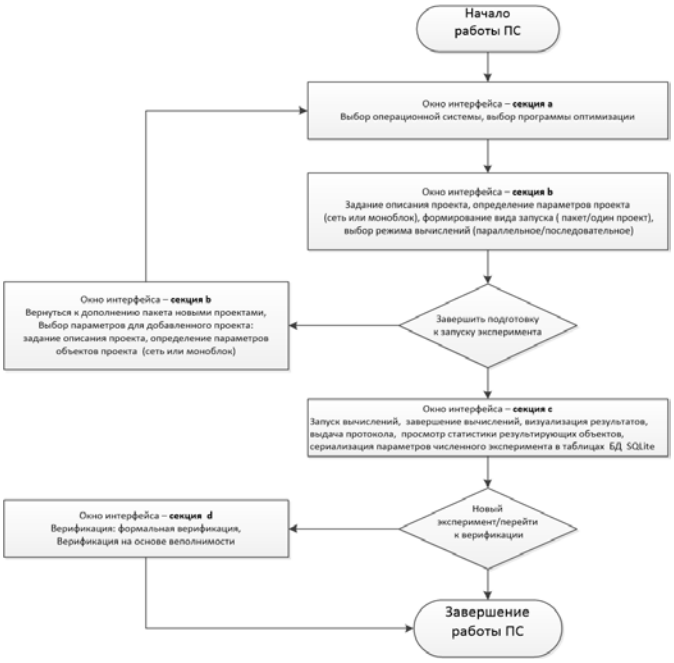



*Рисунок 3.* Общая схема работы с данными в окне интерфейса PLUS

На рисунке 2 представлено состояние графического интерфейса ПС PLUS, когда ни один из элементов управления (радиокнопки, кнопки, выпадающие списки) еще не был выбран (не актуализирован). Актуализация в контексте навизации по интерфейсу означает, что при выборе одного из элементов предлагаются возможные варианты последующих действий путем подсвечивания элементов, обозначающих последующие действия. Подобным образом осуществляется визуальная навигация при составлении той или иной схемы проводимого эксперимента. Работа в среде PLUS начинается с актуализации элементов управления в секции *a*. Соответствующий фрагмент окна представлен на рисунке 4. На фрагменте элементы управления обозначены номерами, что позволяет формальным образом описать порядок действий с элементми. Переход от одного элемента управления интерфейса к другому можно записать схематически так: если за выбором элемента управления $i$ следует выбор актуализированного (доступного) элемента управления $j$, то навигация от $i$ к $j$ .записывается как $i \rightarrow j$.

Подготовка к проведению эксперимента начинается с выбора операционной системы – PC (Windows), Virtual Machine (Linux), Hibrid Computer (GRID cluster). В секции *a* есть элементы, которые задают два выпадающих списка: *a*6 – список доступных программ оптимизации для Windows и для Linux; *a*7 – список имен виртуальных машин (ВМ) Linux. Выбор программы оптимизации подтверждается нажатием кнопки *a*8. Выбор виртуальной машины подтверждается нажатием кнопки *a*9. После этого становятся доступными элементы: список выбора программы и список выбора ВМ. На рисунке 4 также представлены три варианта навигации в этой секции полученных посредством актуализации элементов

управления для: 1) проведения эксперимента на платформе Windows; 2) проведения эксперимента на платформе Linux (виртуальная машина); 3) Переход к секции *d* для проведения верификации результатов проведенных экспериментов.

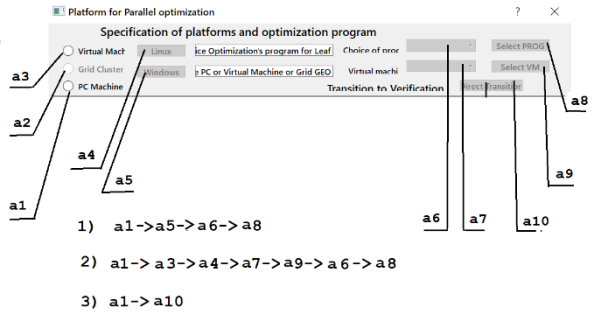



*Рисунок 4.* Элементы управления и варианты навигации по элементам управления в секции *a*

После выбора ОС и конкретной программы оптимизации становятся доступными элементы управления во второй секции (секции *b*). Навигация по элементам этой секции определяет: место расположения папки проекта, структуру sf-описания исходного проекта, режим вычислений, предоставление статистики об объектах проекта, подключение файла конфигурации (по умолчанию или импорт из системы FLC-2). При проведении эксперимента, как для одного проекта, так и для пакета, местом хранения исходной и результирующей информации является папка проекта. В папке проекта собираются исходные и оптимизированные описания логических схем, протоколы с результатами оптимизации, протоколы верификации, скрипты заданий для пакетирования. Актуализация элементов управления в этой секции начинается с выбора места расположения проекта. Место размещения проекта в файловой системе и имя папки проекта определяется пользователем. Фрагмент окна с секцией *b* и обозначением порядкового номера каждого элемента управления представлен на рисунке 5.

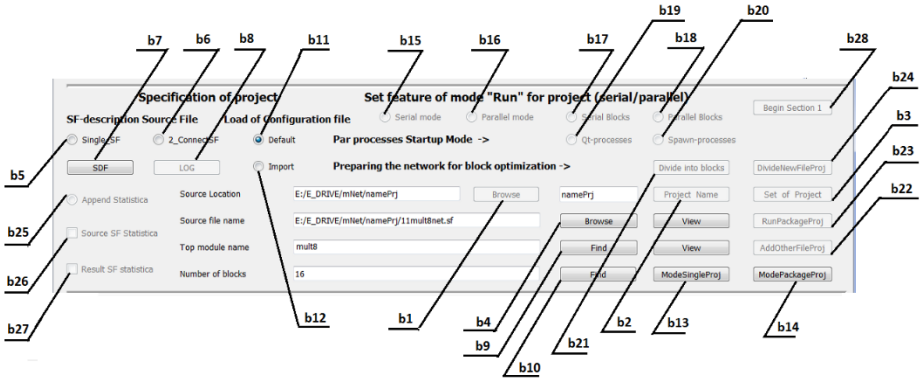



*Рисунок 5*. Обозначение элементов управления во второй секции интерфейса

Место размещения проекта и имя папки проекта определяются пользователем, посредством последовательным прохождением элементов: *b*1 (выбор места размещения проекта), *b*2 (активация окна для ввода имени папки проекта) и *b*3 (фиксация имени папки проекта). После выбора места расположения и задания имени папки проекта (действия *b*1→*b*2→*b*3), при актуализации элемента *b*4 (кнопка «Browse») осуществляется поиск исходного sf-описания проекта. Имя файла с этим описанием и путь к месту его расположения отображается в соответствующем окне просмотра. После нажатия элемента *b*4 актуализируется группа элементов управления в левой части секции *b*, последовательным выбором которых уточняется структура и формат исходного sf-описания проекта.

Таким образом, начало навигации по элементам секции $b$ выглядит следующим образом: $b1 \rightarrow b2 \rightarrow b3 \rightarrow b4$. Далее, выбором элементов $b5 \rightarrow b7$ уточняется структура исходного проекта – sf-описание схемы представлено одним блоком в формате SDF. Выбор элементов $b5 \rightarrow b8$ определяет следующую структуру проекта – sf-описание схемы представлено одним блоком в формате LOG. Если провести выбор элементов $b6 \rightarrow b8$, то он определяет структуру описания схемы как 2-Connect сеть в формате LOG. Выбор $b6 \rightarrow b7$ определяет структуру описания схемы как 2-Connect сеть в формате SDF. Исходное описание проекта можно просмотреть (имеются соответствующие кнопки «View»). Только после уточнения структуры исходного проекта актуализируются кнопки $b9$ и $b10$. Нажатие этих кнопок позволяет в соответствующих полях просмотра визуализировать – имя головного блока и число независимых блоков, которым представлено описание исходного проекта. По завершении действий с кнопками $b9$ и $b10$ актуализируются элементы $b11$ и $b12$, которые определяют способ подключения файла конфигурации config.ini ($b11$ – по умолчанию) или ($b12$ – импорт из системы FLC-2). После определения файла конфигурации появляется возможность выбрать: продолжать эксперимент с одним проектом или организовать пакет с заданным числом проектов. Эту возможность предоставляет актуализация элементов $b13$ («ModeSingleProj» – запуск оптимизации для одного проекта) и $b14$ («ModePackageProj» – формирование пакета для проведения оптимизации с несколькими проектами). Задание и уточнение структуры описания при работе с одним проектом определяются следующими вариантами навигации по элементам управления секции $b$, которые представлены на рисунке 6.

1) $b1 \rightarrow b2 \rightarrow b3 \rightarrow b4 \rightarrow b5 \rightarrow b7 \rightarrow b9 \rightarrow b10 \rightarrow b11(b12) \rightarrow b13$ – блок, формат SDF
2) $b1 \rightarrow b2 \rightarrow b3 \rightarrow b4 \rightarrow b5 \rightarrow b8 \rightarrow b9 \rightarrow b10 \rightarrow b11(b12) \rightarrow b13$ – блок, формат LOG
3) $b1 \rightarrow b2 \rightarrow b3 \rightarrow b4 \rightarrow b6 \rightarrow b7 \rightarrow b9 \rightarrow b10 \rightarrow b11(b12) \rightarrow b13$ – сеть, формат SDF
4) $b1 \rightarrow b2 \rightarrow b3 \rightarrow b4 \rightarrow b6 \rightarrow b8 \rightarrow b9 \rightarrow b10 \rightarrow b11(b12) \rightarrow b13$ – сеть, формат LOG

*Рисунок 6* . Варианты навигации по секции $b$ для подготовки данных при запуске одного проекта

**Порядок действий при проведении эксперимента с одним проектом.** При нажатии $b13$ актуализируется подмножество из четырех радиокнопок, определяющих режим вычислений. Выбор одной радиокнопки из этого подмножества осуществляет подготовку к запуску выполнения логической оптимизации:

1 $b15$ («serial mode») – выполнение программы оптимизации в последовательном режиме для sf-описания исходного проекта, представленного в виде одного блока;

2 $b16$ («parallel mode») – выполнение программы оптимизации в параллельном режиме для sf-описания исходного проекта, представленного в виде 2-Connect сети. В противном случае выдается предупреждение о недоступности этой кнопки для программы оптимизации, которая была выбрана в секции $a$, при представлении исходного проекта в виде одного блока;

3 $b17$ («serial block») – выполнение оптимизации в обычном режиме для sf-описания исходного проекта, представленного описанием в 2-Connect формате;

4 $b18$ («parallel block») – выполнение оптимизации в параллельном режиме для sf-описания исходного проекта, представленного описанием сети в 2-Connect формате. Фрагмент третьей секции (секции $c$) окна интерфейса представлен на рисунке 7.

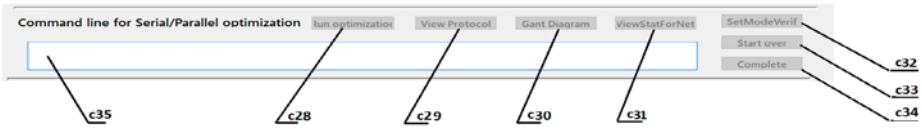



*Рисунок 7*. Назначение элементов управления в третьей секции (*c*) интерфейса

На рисунке 7 указаны номера для обозначения каждого элемента управления данной секции. В секции расположено окно просмотра, в котором отображается командная строка запуска выбранной программы оптимизации. После завершения действий по уточнению структуры объекта (один проект), которые были доступны при выборе *b*13, и выбора одного из режимов вычислений (*b*15|*b*16|*b*17|b18) актуализируется элемент в третьей секции *c*28 («Run optimization»), нажатие которого приводит к запуску программы оптимизации. В окне просмотра (*c*35) отображается командная строка с параметрами запускаемой программы оптимизации. При завершении вычислений актуализируются элементы *c*29 («View protocol»), *c*30 («Gant Diagram») и *c*31 («ViewStatNew») – выбор просмотра протокола, визуализации и статистики. В этой секции предусмотрен элемент *c*35 («Complete») – завершение работы PLUS. На рисунке 8 приведены варианты навигации в секции *b* для проведения эксперимента с одним проектом, представленном или в виде одного блока или в виде 2-Connect-сети.

1 *b*1→*b*2→*b*3→*b*4→*b*5→*b*7→*b*9→*b*10→*b*11(*b*12)→*b*13→(*b*15|*b*16)→*c*28 – блок, формат SDF
2 *b*1→*b*2→*b*3→*b*4→*b*5→*b*8→*b*9→*b*10→*b*11(*b*12)→*b*13→(*b*15|*b*16)→*c*28 – блок, формат LOG
3 *b*1→*b*2→*b*3→*b*4→*b*6→*b*7→*b*9→*b*10→*b*11(*b*12)→*b*13→(*b*17|*b*18)→*c*28 – сеть, формат SDF
4 *b*1→*b*2→*b*3→*b*4→*b*6→*b*8→*b*9→*b*10→*b*11(*b*12)→*b*13→(*b*17|*b*18)→*c*28 – сеть, формат LOG

*Рисунок 8*. Завершение навигации действием, запускающим эксперимент с одним проектом

Если необходимо провести вычисления с той же программой оптимизации, но с исходными данными другого проекта, проводится выбор элемента *c*33 («Start over»). В таком случае при сохранении установок первой секции проводится заново актуализация элементов в секции *b*, согласно маршрутам навигации, приведенным на рисунке 8, начиная с определения места расположения нового проекта. Если надо провести верификацию результатов эксперимента, то выбор элемента *c*32 («SetModeVerif») актуализирует элементы четвертой секции (секции *d*) для работы с программами верификации.

В вычислительном ядре визуальной среды PLUS представлены: набор модулей, реализующих процедуры логической оптимизации в системе FLC-2, параллельные версии программ оптимизации (Minim, bdd_builder, BoolNetOpt), исполняемые коды таких известных программ оптимизации, как Espresso и ABC [5]. Выбор элемента *c*28 запускает выполнение модулей оптимизации с ключевыми параметрами, формируемыми структурой исходного описания и условиями запуска конкретной программы.

**Замечание 1.** Если исходное описание схемы представлено сетью в 2-Connect формате, кнопки *b*17 и *b*18 не будут актуализированы, пока не будет проведено выбором

элемента *b*21 («Divide into block») разделение сети на отдельные файлы с sf-описаниями независимых блоков (3-й маршрут на рисунке 8).

**Замечание 2.** Кнопки *b*25, *b*26, *b*27 актуализируются по завершению задания исходных данных, и могут быть выбраны при необходимости получения полной статистики для исходного и оптимизированного sf-описаний проекта. Статистика отображается в протоколе при завершении процесса оптимизации.

**Замечание 3**. Программы, включенные в вычислительное ядро визуальной среды PLUS и выполняемые в ОС Windows, программы, выполняемые в ВМ с ОС Linux (на гостевой виртуальной машине), находятся в общем выпадающем списке секции *a*, который открывается при нажатии элемента *a*8.

**Замечание 4.** Актуализация группы элементов *b*15, *b*16, *b*17, *b*18 определяется применимостью выбранной программы оптимизации для описания схемы с конкретной структурой и форматом. Например, если при выборе распараллеленной версии программы раздельной минимизации MinimSepPar ошибочно было указано действие *b*15 («serial mode»), выдается предупреждение с подсказкой выбора нужного для этой программы действия, как показано на рисунке 9. Совершив действие, указанное в этом предупреждении, проведение эксперимента с этой программой можно продолжить.

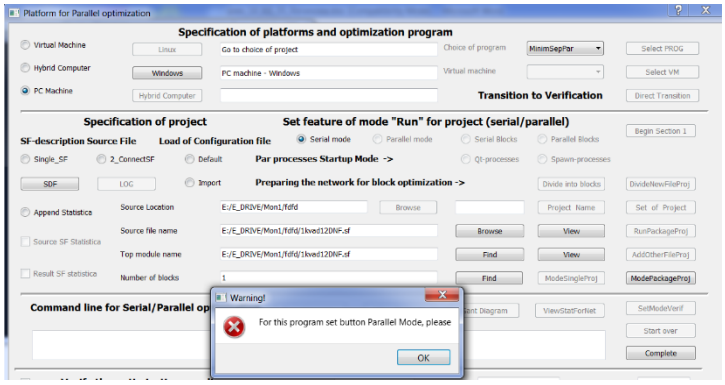



*Рисунок 9*. Предупреждение о неправильном выборе режима вычислений

**Порядок действий при проведении эксперимента с пакетом проектов.** При выборе элемента *b*14 («ModePackageProj») осуществляется подготовка по организации экспериментов в подсистеме пакетирования. Формирование пакета запускается, ставшим доступным, элементом *b*22 («AddOtherProj»). При этом в секции *b* очищаются все поля с данными предыдущего проекта, затем проводятся действия согласно маршрутам навигации, представленными на рисунке 8, но с формированием тех параметров исходных данных и режимов вычислений, которые соответствуют проекту, присоединяемому к пакету. Для добавленного проекта можно поменять и программу, которая будет работать с данными этого проекта, так как становится доступным элемент *a*8 в первой секции.

Каждый раз добавление очередного проекта в пакет предшествует выбор элемента *b*22 («AddOtherProj»). Формирование нужного числа проектов в пакете завершается действием *b*23 («RunPackageProj»), которая была доступна на всем этапе формирования пакета. В результате действия *b*23 актуализируется подмножество из четырех элементов, определяющих режим вычислений и параметры запускающей программы для пакета:

1 *b*17 («serial block») – выполнение в обычном режиме программ оптимизации, назначенных для проектов, которые собраны в сформированный пакет;

2 *b*18 («parallel block») – выполнение в параллельном режиме программ оптимизации назначенных для проектов, которые собраны в сформированный пакет;

3 *b*19 («Qt-processes») – запуск программ оптимизации, которые назначены для проектов в пакете, командой «start»;

4 *b*20 («Spawn-processes») – запуск программ оптмизации, которые назначены для проектов в пакете, командой «spawn».

**Замечание 5.** Если описание проекта, добавляемого в пакет, представлено 2-Connect сетью, то элементы *b*22 (добавление нового проект в пакет) и *b*23 (завершение формирования пакета) не будут актуализированы, до тех пор, пока не будет проведено разделение 2-Connect сети предыдущего проекта на файлы с описаниями отдельных блоков. Таким образом, при представлении описания нового проекта в виде 2-Connect сети только после выбора элемента *b*24 («Divide into block») актуализируются кнопки *b*22 и *b*23.

**Замечание 6.** Выбор элементов управления *b*24 (при работе в пакетном режиме) и *b*21 (при работе с одним проектом) заключается в выполнении следующих действий:

– разделение исходной 2-Connect- сети на блоки, проведение анализа блоков, которые имеют промежуточные переменные, исключение промежуточные переменные в этих блоках;

– подготовка и пребразование структуры блоков, унификации имен блоков;

– занесение sf-описаний блоков в отдельные файлы, перенос файлов в папку проекта.

На рисунке 10 представлена схема всех связей элементов управления в главном окне графического интерфейса визуальной среды PLUS с указанием порядка действий и маршрутов навигации (работа с секцией верификации, в данной статье не рассматривается). Если провести анализ формирования всех возможных маршрутов навигации по элементам интерфейса, можно придти к следующим выводам:

1 При работе в визуальной среде PLUS требуется лишь понимание того, с какими проектами проводятся эксперименты, и какие программы оптимизации для этих проектов надо запустить. Навигация по элементам интерфейса с интеллектуальной поддержкой требует одного – следовать указаниям доступных элементов управления интерфейса, согласно выбранному пользователем цели эксперимента.

2 Вся логика взаимодействия системных, программных модулей визуальной среды, функциональных модулей процедур оптимизации системы FLC-2, средств виртуализации и верификации, отделено от пользователя графической оболочкой с многофункциональным интерфейсом. Особенностью интерфейса является: «умное» управление действиями пользователя, сериализация параметров численного эксперимента, формирование таблиц базы данных в формате SQLite ]14] с информацией о прошедших экспериментах, проведение параллельных вычислений с возможностью визуального анализа результатов параллельных процессов, как в экспериментах с одним проектом, так и с проектами, собранными в пакет.

Таким образом, реализации вычислительных экспериментов с программами логической оптимизации согласно рассмотренным выше схемам, в определенной мере представляют собой потоки выполнения (workflow) в нотации сеанса эксперимента, которые позволяют компоновать различные вычислительные задачи в последовательные и параллельные структуры, направленные на достижение цели проводимого эксперимента. Потоки выполнения может быть описаны графом, где вершины являются действиями, выполняемые над данными, а ребра – потоками данных. Такая модель позволяет абстрагироваться от тонкостей запуска отдельных приложений и функциональных модулей. В оболочке PLUS в качестве действий выступают элементы управления, описанные в предыдущих разделах. Визуальная оболочка, встроенная в PLUS, позволяет

пользователю проектировать будущий workflow и задавать характеристики его выполнения.

Таким образом, пользователь может проводить вычисления, создавая конструкции, которые отражают семантику вычислительных процессов. К тому же, одной из главных особенностей PLUS является то, что программы оптимизации могут запускаться в параллельном режиме, что позволяет пользователю в полной мере задействовать возможности многоядерной вычислительной архитектуры. В случае запроса на параллельное выполнение, происходит автоматическая декомпозиция исходных данных, распределение звдвч кооперативным планировщиком по ядрам процессора, выполнение оптимизации, сбор и слияние в оптимизированное описание логической схемы, в соответствии с форматами системы FLC-2.

Из всех многочисленных связей программных модулей и действий, представленных на рисунке 10, помечены цветом те сущности, которые необходимы для достижения цели проводимых вычислений и которые должны четко определяться действиями пользователя, как показано ниже на рисунке 11.

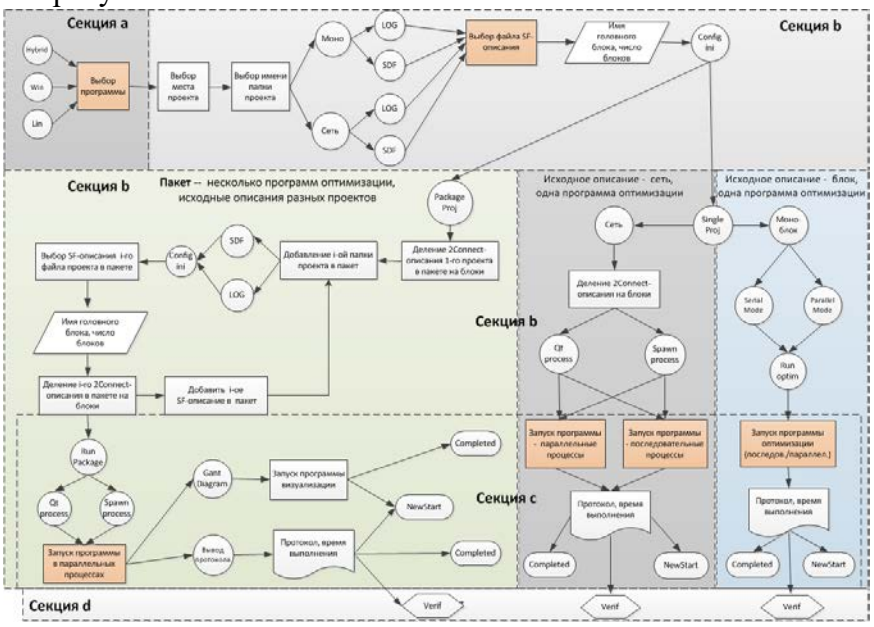



*Рисунок 10*. Графовая модель workflow эксперимента в нотации визуальной оболочки PLUS

Остальные элементы workflow, входящие в маршрут навигации формируются с обеспечением интеллектуальной поддержки программными средствами интерфейса PLUS в зависимости от структуры проекта, типа оптимизации и типа вычислений.

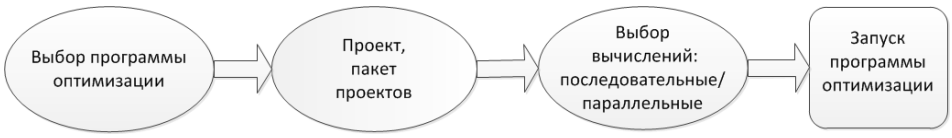



*Рисунок 11*. Действия пользователя при проведении вычислительного эксперимента

**Организация визуализации при проведении параллельных вычислений в ПС Plus.** Средства интерфейса визуальной среды PLUS предоставляют возможность визуализации трассы параллельных вычислений при проведении эксперимента с пакетами или с проектом, в котором описание логической схемы представленно в виде сети в 2-Connect формате. Визуальный анализ позволяет оценить эффективность распределения параллельных процессов

по ядрам вычислительной системы для получения характеристик производительности, оценить оптимальность процедур разбиения исходных проектов на независимые блоки. Присутствие такого средства в системе PLUS можно обосновать следующими доводами. Эффективность параллельной программы зависит от большого числа параметров: как от архитектуры вычислительной системы, алгоритмов прикладной задачи, так и от библиотек поддержки различных технологий параллелизма. Помимо вопросов, связанных с решением вышеперечисленных проблем, существует проблема, заложенная в самой природе применения параллельных вычислений для решения логико-комбинаторной задач, – недетерминированность выполнения [12]. Параллельная программа на одних и тех же входных данных может давать разные результаты. Поэтому для валидации полученных параллельных решений в PLUS включена подсистема верификации результатов оптимизации [13].

В третьей секции графического интерфейса предоставляется возможность действием элемента *c*30 получить визуальное представление метрик, которые в процессе выполнения параллельной программы записываются в файлы, называемые трассами (trace file). Данные метрики отображаются на визуальные абстракции, заимствованные из статистической графики, в частности, в виде горизонтальных диаграмм с накоплением [15]. Визуализация трассы, в которой зафиксированы последовательность событий и их временные отметки, возможна после завершения программы, когда актуализируется элемент управления *c*30.

**Особенности параллельных вычислений в визуальной среде PLUS.** В основе организации параллельной вычислений в визуальной среде PLUS, также как и в системе FLC-2, положено выполнение действий: разделение (декомпозиция) объекта проектирования на части, выявление и устранение информационных зависимостей между частями, распределение работы с этими частями по вычислительным элементам (по ядрам процессора). Поскольку любая выполняемая программа образует процесс, то для запуска определенного числа независимых процессов (определяется числом декомпозируемых блоков объекта проектирования) возможна достаточно простая организация параллельного выполнения процессов, если между ними нет межпроцессных коммуникаций. Это осуществляется путем вставки директив OpenMP в код программы, которая производит запуск выполнение процессов. В программе, запускающей параллельные процессы, создается множество так называемых «задач» путем вставки соответствующей директивы OpenMP [9]. Каждая задача производит запуск одного процесса. Задачи динамическим образом распределяются по ядрам (включая логические процессоры, если они есть). Данная схема удобна в том смысле, что: а) она не зависит от численного метода, реализованного в проектной операции; б) в процессах, выполняемых параллельно можно запускать программы, представленные только исполняемыми кодами [8]. При таком подходе организация параллельных вычислений возможна по принципу: один процесс исполнения (код какой-либо программы, разные входные данные) – одно ядро процессора.

**Примеры визуализации, полученные при проведении разных экспериментов.** На рисунке 12 приведены фрагменты файлов конфигурации для четырех проведенных экспериментов, в которых описаны: имя и место расположения исходных проектов, имя программ, работающих с каждом из проектов. Файл конфигурации подается как один из аргументов командной строки для программы запуска параллельных вычислений. Для оптимизации в экспериментах 1-4 использовались программы, которые указаны в файле конфигурации.

| Файлы конфигурации для экспериментов 1-4 | | | |
|---|---|---|---|
| Эксперимент 1 | Эксперимент 2 | Эксперимент 3 | Эксперимент 4 |
| E:/All_Exper2/Net/prim1<br>MinimSepPar<br>24 | E:/AllExper2/maNet/v111<br>MinimABC<br>28<br>E:/AllExper2/maNet/v222<br>MinimABC<br>28<br>E:/AllExper2/maNet/v333<br>MinimABC<br>26<br>E:/AllExper2/maNet/v444<br>MinimABC<br>16 | E:/AllExper2/maNet/ff111<br>ForEsp<br>28<br>E:/AllExper2/maNet/ff222<br>ForEsp<br>28<br>E:/AllExper2/maNet/ff333<br>ForEsp<br>26<br>E:/AllExper2/maNet/ff444<br>ForEsp<br>16 | E:/AllExper2/maNet/vv111<br>MinimABC<br>28<br>E:/AllExper2/maNet/vv222<br>ForEsp<br>28<br>E:/AllExper2/maNet/vv333<br>MinimSepPar<br>26<br>E:/AllExper2/maNet/vv444<br>MinimABC<br>16 |

*Рисунок 12*. Файлы конфигурации для запуска параллельных процессов в экспериментах 1-4

Файл конфигурации подается как один из аргументов командной строки для программы запуска параллельных вычислений. Для оптимизации в экспериментах 1-4 использовались программы, которые указаны в файле конфигурации.

**Эксперимент 1.** Визуализация результатов параллельного выполнения программы MINIM [5] для 2-Connect-сети, которая разделена на 24 блока.

**Эксперимент 2-3**. Визуализация результатов параллельного выполнения двумя программами (ABC и Espresso) четырех проектов, собранных в пакет.

**Эксперимент 4.** Визуализация результатов параллельного проведения оптимизации разными программами в каждом из четырех проектов, собранных в одном пакет.

Вычисления проводились на могоядерном процессоре (8 ядер плюс 8 логических ядер при включении гипертрейдинга). Таким образом, в первом эксперименте выполнение программы MINIM проходит в 24-х процессах, которые распределяются по 16-ти ядрам процессора. После завершения работы получена трасса, в которой представлены значения метрик, показывающих время начала и завершения процесса, длительность выполнения программы, запускаемой в данном процессе. В трассе для каждого процесса приводятся: номер потока исполнения (номер ядра), номер блока данных с которым проводится вычисления. Для визуального оценки распределения вычислений между ядрами процессора проводится построение графического отображения трассы в виде горизонтальной столбиковой диаграммы с накоплением (модифицированной диаграммы Ганта) [15].

Использование этого инструмента визуализации в контексте отображения результатов параллельных вычислений, можно использовать не только для оценки загрузки ядер процессора, но также позволяет увидеть возможные проблемные места при организации параллельной работы с использованием различных алгоритмов разбиения и оптимизации. При построении диаграммы сначала проводится фильтрация событий по потокам исполнения, по результатам которой формируется подмножества событий, относящихся к каждому из ядер. Затем в каждом из полученных подмножеств, проводится сортировка процессов по возрастанию времени от начала и до окончания последнего процесса в данном подмножестве. Поскольку в трассе отдельного процесса указывается время выполнения программы, именно эти значения будут определять размеры интервалов на диаграмме. На рисунках 13 и 14 приведены диаграммы Ганта, построенные по трассам экспериментов 1 и 2.

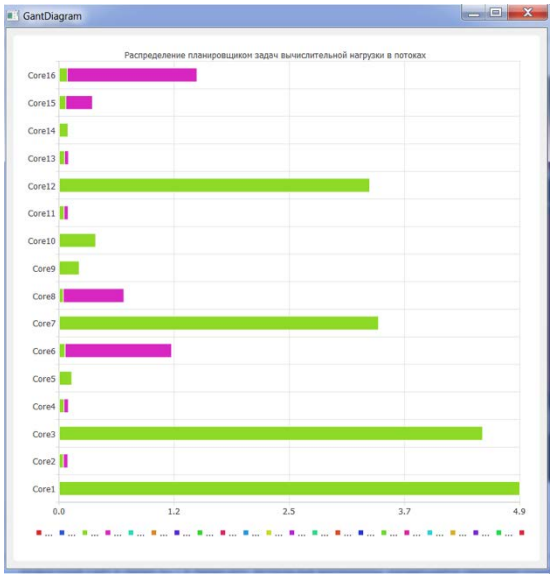



*Рисунок 13.* Диаграмма распределения процессов по ядрам для одного проекта

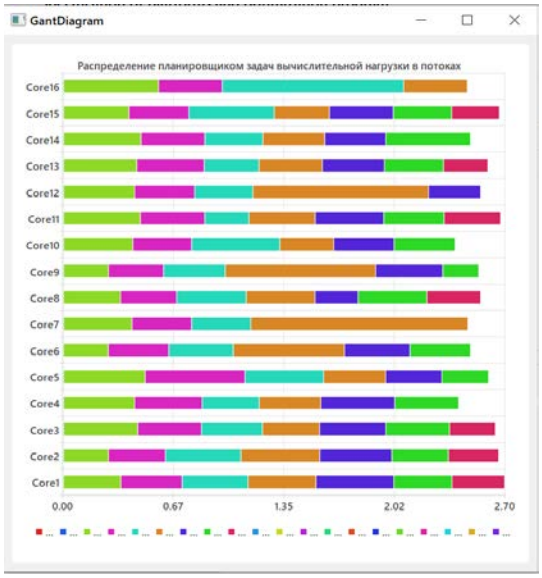



*Рисунок 14.* Программа ABC. Диаграмма распределения процессов пакета по ядрам

**Эксперимент 2.** Параллельное выполнение четырех проектов, собранных в пакет. Для всех 98-ми блоков выполнялась программа ABC, алгоритмы которой эффективно работают при минимизации ДНФ булевых функций большой размерности [5]. По завершению всех параллельных процессов получен файл трассы, визуализация которого приведена на рисунке 14.

Длина каждого интервала на диаграмме определяется временем работы программы оптимизпции в одном процессе. Разные значения этой величины определяется параметрами исходных данных, которые представляют содержимое каждого блока. Нумерация блоков в пакете определяется порядком расположения исходных объектов в файле конфигурации,

**Эксперимент 3.** При проведении эксперимента использована программа Espresso [5]. Параллельное выполнение организовано для пакета, собранного в том же порядке, как и для эксперимента 2. В 98-ми процессах проходило выполнение программы минимизации Espresso для каждого блока. По завершению всех параллельных процессов, получены файл трассы с указанием порядка выполнения процессов в ядрах. На рисунке 15 приведена диаграмма Ганта, построенная по данным файла трассы, полученной в эксперименте 3.

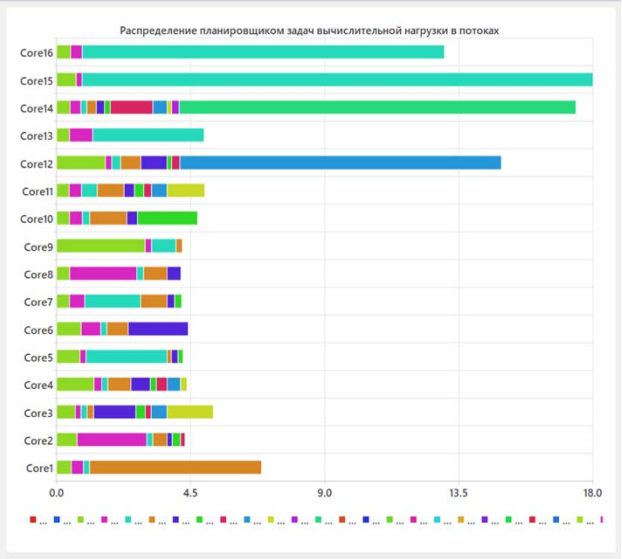



*Рисунок 15.* Программа Espresso. Диаграмма распределения процессов пакета по ядрам

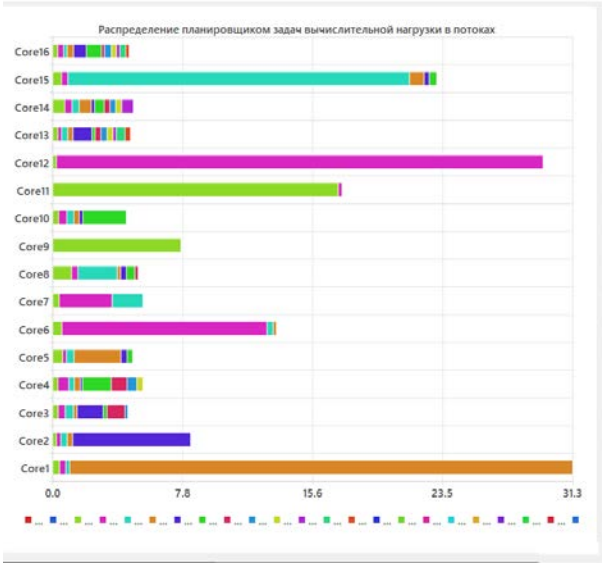



*Рисунок 16.* Диаграмма распределения процессов пакета по ядрам при выполнении разных программ

**Эксперимент 4.** Визуализация результатов параллельного проведения вычислений несколькими программами (ABC, Espresso, MINIM) для четырех проектов представлена на рисунке 16. Во время работы 98-ми параллельных процессов проходило выполнение оптимизации с той группой блоков, для которой назначена каждая программа оптимизации.

**Заключение.** Графические интерфейсы являются важнейшим элементом человеко-компьютерного взаимодействия. Они обеспечивают формирование входных данных для программ и визуализацию результатов. Система FLC-2 прелагает удобный интерфейс для разработчика, но при этом лишена встроенных средств одновременной организации вычислительных экспериментов с раздичными описаниями логических схем. Построение автономной визуальной среды PLUS, интегрированной по данным и функциональным рещениям расширяет возможности системы FLC-2. В настоящей работе представлен поэтапный процесс работы в графической оболочке для постановки численных экспериментов. Особенности интерфейса визуальной среды расширяют и дополняют возможности системы FLC-2, повышают уровень автоматизации при выполнении численных экспериментов, позволяют увидеть схему эксперимента в целом, снижает количество ошибок и повышает оперативность принятия проектных решений.

## Список литературы

[1] Бибило, П. Н. Система логической оптимизации функционально-структурных описаний цифровых устройств на основе продукционно-фреймовой модели представления знаний / П. Н. Бибило, В. И. Романов // Проблемы разработки перспективных микро- и наноэлектронных систем. – 2020. – Вып. 4. – С. 9–16.

2] Banerjee, P. Parallel Algorithms For VLSI Computer-Aided Design / P.Banerjee – Prentice Hall, Englewoods Cliffs, NJ, 1994. – 699 p.

[3] Черемисинов, Д.И. Проектирование и анализ параллелизма в процессах и программах / Д.И. Черемисинов. – Минск: Беларус. навука, 2011. – 300 с.

[4] Черемисинов, Д.И. Использование параллельных вычислений при автоматизированном проектировании СБИС / Д.И. Черемисинов, Л.Д. Черемисинова // Проблемы разработки перспективных микро- и наноэлектронных систем – 2016. Сборник трудов / под общ. ред. академика РАН А.Л. Стемпковского. – М.: ИППМ РАН, 2016. Часть – С. 2-8.

[5] Бибило, П.Н. Экспериментальное сравнение эффективности программ минимизации систем булевых функций в классе дизъюнктивных нормальных форм / П.Н. Бибило, И.П. Логинова // Информатика. – 2022. – № 2(19). – С. 26–55.

[6] Бибило, П. Н. Система логической оптимизации функционально-структурных описаний цифровых устройств на основе продукционно-фреймовой модели представления знаний / П. Н. Бибило, В. И. Романов // Проблемы разработки перспективных микро- и наноэлектронных систем. – 2020. – Вып. 4. – С. 9–16.

[7] Логинова, И. П. Построение и визуальный анализ параллельных решений в системе логической оптимизации функционально-структурных описаний дискретных устройств / И. П. Логинова // Компьютерное проектирование в электронике = Electronic Design Automation: сб. трудов Междунар. науч.-практ. конф. (Республика Беларусь, г. Минск, 28 ноября 2024 г.) / редкол.: В. Р. Стемпицкий [и др.]. – Минск: БГУИР, 2024. – С. 108 –111.

[8] Логинова, И.П. Анализ быстродействия при реализации параллельных вычислений для решения задач оптимизации в системе логического проектирования / И. П. Логинова // IX Международная научно-практическая конференция «BIG DATA and Advanced Analytics» (BIG DATA 2023): Сборник научных статей в 2 ч. Ч. 1, (Республика Беларусь, Минск, 17-18 мая 2023 г.). – Минск: БГУИР, 2023. – С. 365–378.

[9] E Ayguadé, E. The design of Open MP tasks / E. Ayguade [et al.] // IEEE Transactions on Parallel and Distributed Systems. – 2009. – Vol. 20, no. 3. – P. 404 –418.

[10] Торопов, Н. Р. Преобразование многоярусной комбинационной сети в двухъярусную / Н. Р. Торопов // Логическое проектирование : сб. науч. тр. – Минск : Ин-т техн. кибернетики НАН Беларуси, 2000. – Вып. 5. – С. 4–14.

[11] Логинова И.П. Организация параллельных вычислений на базе программного комплекса PLUS /И.П. Логинова // XII Международная научно-техническая конференция «Информационные технологии в промышленности, логистике и социальной сфере» (ITI*2023): тезисы докладов, 21-22 сентября 2023 г. – Минск : ОИПИ НАН Беларуси, 2023. – С. 103–107.

[12] Черемисинова, Л.Д. Программные средства верификации описаний комбинационных устройств в процессе логического проектирования / Л.Д. Черемисинова, Д.Я. Новиков // Программная инженерия, 2013, № 7, с. 8–15.

[13] Логинова, И.П. Формальная верификация результатов выполнения параллельных программ логической оптимизации / И. П. Логинова // Информационные технологии и системы 2022 (ИТС 2022) – Information Technologies and Systems 2022 (ITS 2022): материалы междунар. науч. конф. (Республика Беларусь, Минск, 30 октября 2021 года). – Минск: БГУИР, 2021. – С. 85-86.

[14] Шлее, М. Qt 5.10. Профессиональное программирование на C++ / М. Шлее. – СПб.: БХВ-Петербург, 2018. – 1072 с.

[15] Логинова, И. П. Визуализация трассы параллельной программы в системе проектирования / И. П. Логинова // Информационные технологии и системы 2024 (ИТС 2024) – Information Technologies and Systems 2024 (ITS 2024): материалы междунар. науч. конф., Минск, 20 нояб. / редкол. : Л. Ю. Шилин [и др.]. – С. 89 – 90

# THE BUILDING A GRAPHICAL SHELL FOR PERFORMING EXPERIMENTS BASED ON A SYSTEM FOR LOGICAL OPTIMIZATION OF FUNCTIONAL AND STRUCTURAL DESCRIPTIONS OF DISCRETE DEVICES

***I.P. Loginova***

*Senior researcher, United Institute of Informatics Problems of the National Academy of Sciences of Belarus, PhD, associate professor*

**Abstract.** The article describes the functioning of a graphical interface designed to visually form the scheme of a computational experiment in the form of chains of actions proposed by the interface control elements. The visual environment of the software allows making for a variety of experiments using a set of design operations of a system for logical optimization of functional and structural descriptions of discrete devices FLC-2. By means of the graphical interface, an experimental scheme is formulated, in the process of which a choice is made: programs and objects participating in the experiment, User runs the experiment scheme, monitors the result with built-in visualization and verification tools. The software includes a bundling subsystem in which interface tools generate parallel tasks for logical optimization in objects assembled into packages.

**Keywords:** multicore system, parallel process, task scheduler, speedup of performance, logical optimization, visualisation.

УДК 336.74

# АНАЛИЗ МЕДИЦИНСКИХ ДАННЫХ С ПОМОЩЬЮ АЛГОРИТМА APRIORI НА ЯЗЫКЕ R

***В.С. Лянцевич***
*Студент кафедры математического и информационного обеспечения экономических систем*
*УО ГрГУ им. Я.Купалы,*
valeria1lyantsevich@gmail.com

***Н.В. Марковская***
*Доцент кафедры математического и информационного обеспечения экономических систем*
*УО ГрГУ им. Я.Купалы,*
n.markovskaya@grsu.by

***В.С. Лянцевич***
*Студентка 4 курса Гродненского государственного университета имени Янки Купалы кафедры математического и информационного обеспечения экономических систем.*

***Н.В. Марковская***
*Доцент кафедры математического и информационного обеспечения экономических систем УО ГрГУ им.Я.Купалы, кандидат физико-математических наук, доцент.*

**Аннотация.** Статья рассматривает использование алгоритма Apriori для выявления ассоциативных правил в медицинском наборе данных опухоли молочной железы. Анализируются особенности структур данных, проводится их изменение для применения алгоритма, оценивается реализация Apriori. Исследование проведено на основе преобразованного набора данных, выявлены ключевые закономерности, а также проанализировано влияние параметров алгоритма на качество извлекаемых правил. Сделаны выводы о применимости алгоритма Apriori в задачах анализа медицинских данных, предложены рекомендации по его настройке для различных типов данных.

**Ключевые слова:** ассоциативные правила, алгоритм Apriori, медицинские данные, опухоль молочной железы.

**Введение.** Обнаружение ассоциативных правил – это метод, направленный на выявление логических закономерностей между элементами. Он гарантирует нахождение сильных правил в базе данных с помощью указанных показателей.

Правила ассоциаций представляют выражениями: «если "некоторое условие", то "некоторое следствие или результат"». На практике их записывают определённой математической записью: {некоторое условие} => {некоторое следствие или результат}.

Главная проблема поиска ассоциативных правил – большое количество правил, которые появляются при анализе. Для уменьшения количества были придуманы ограничения: они сохраняют только те правила, значения мер которых превышают заданные пороги. К ним относятся поддержка, доверие, лифт и уверенность.

Один из способов поиска правил ассоциации – алгоритм Apriori. Это классический метод анализа данных, который находит скрытые закономерности и зависимости между различными элементами.

Apriori традиционно применяют для анализа покупательской корзины, поиска связей между приобретаемыми товарами. Однако эта статья посвящена использованию алгоритма

для медицинской сферы: найти характеристики, которые будут сопровождать различные опухоли.

**Основная часть.** Исследование проводится с использованием набора данных Breast Cancer Wisconsin (Diagnostic) Data Set. Наша цель – определить признаки, которые чаще всего встречаются вместе с конкретным видом образования.

Breast Cancer Wisconsin (Diagnostic) Data Set содержит идентификатор, поставленный диагноз и десять действительных характеристик [1]:

1 Радиус – среднее значение расстояний от центра до точек по периметру.

2 Периметр – общее расстояние между точками.

3 Площадь измеряется простым подсчетом количества пикселей внутри змеи и добавлением половины пикселей по периметру.

4 Компактность получается из объединения периметра и площади с использованием формулы 1 периметр$^2$/площадь. Это безразмерное число увеличивается с нерегулярностью границы.

5 Гладкость количественно определяется путём измерения разницы между длиной радиальной линии и средней длиной линий, её окружающих.

6 Вогнутостью является выраженность вогнутых частей контура.

7 Вогнутые точки – это характеристика, похожая на вогнутость, но измеряет только количество, а не величину вогнутостей контура.

8 Симметрия измеряется путём нахождения большой оси или самой длинной хорды, проходящей через центр.

9 Фрактальная размерность клетки аппроксимируется с помощью «приближения береговой линии», описанного Мандельбротом. Периметр ядра измеряется с помощью все более крупных «линеек». По мере увеличения размера линейки, уменьшая точность измерения, наблюдаемый периметр уменьшается.

10 Текстура ядра клетки измеряется путем нахождения дисперсии интенсивностей серой шкалы в составляющих пикселях.

В наборе содержится информация о 357 доброкачественных образованиях и 212 – о злокачественных. Для каждой записи определены среднее значение, экстремальное (наихудшее) значение и стандартное отклонение каждой характеристики [2].

Характеристики представлены в виде численных значений. Однако алгоритм Apriori предназначен для поиска ассоциативных правил в категориальных данных. Следовательно, для последующего применения нужно преобразовать числовые данные.

Для реализации алгоритма будут использоваться средние значения признаков. По этой причине из файла данных удалим ненужные столбцы и изменим названия других.

```
breastc <- subset(breastc, select = -c(id, X, perimeter_se, area_se, smoothness_se, compactness_se, concavity_se, concave.points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worst, area_worst, smoothness_worst, compactness_worst, concavity_worst, concave.points_worst, symmetry_worst, texture_se, radius_se, fractal_dimension_worst)).
```

```
breastc <- breastc %>% rename(radius = radius_mean, texture = texture_mean, perimeter = perimeter_mean, area = area_mean, smoothness = smoothness_mean, compactness = compactness_mean, concavity = concavity_mean, concave_points = concave.points_mean, symmetry = symmetry_mean, fractal_dimension = fractal_dimension_mean)
```

Далее данные разделим на три группы, выделяя три категории: с низкими, средними и высокими значениями. В языке R это можно осуществить с помощью функции cut(). Она автоматически определяет минимальное и максимальное значение переменной, а затем делит диапазон на равные части.

В результате вместо чисел будут использоваться соответствующие слова. После этого приводим данные к виду транзакций.

```
breastc_discr <- breastc %>%
+       mutate(across(where(is.numeric), ~ cut(., breaks = 3, labels = c("Low", "Medium", "High"))))
```

```
breastc_trans <- as(breastc_discr, "transactions").
```

Количество доброкачественных и злокачественных наблюдений в наборе данных отличается. Поэтому поиск правил проводится раздельно.

Для злокачественного случая пороговое значение поддержки будет равно 0,1, а доверия – 0,8. Следствием будет выступать злокачественная опухоль, условием – признаки. Показатель доверия отражает частоту выполнения правила, поэтому отсортируем результаты по этому критерию.

```
rules_m <- apriori(breastc_trans,
      parameter = list(support = 0.1, confidence = 0.8, minlen = 2),
      appearance = list(rhs = "diagnosis=M", default = "lhs"))
inspect(head(sort(rules_m, by = "confidence"), 5))
plot(rules_m)
```

```
     lhs                          rhs             support confidence  coverage
[1] {area=Medium,
     concavity=Medium}       => {diagnosis=M} 0.1019332          1 0.1019332
[2] {area=Medium,
     concave_points=Medium} => {diagnosis=M} 0.1318102          1 0.1318102
[3] {texture=Medium,
     area=Medium}            => {diagnosis=M} 0.1265378          1 0.1265378
[4] {area=Medium,
     symmetry=Medium}        => {diagnosis=M} 0.1212654          1 0.1212654
[5] {area=Medium,
     smoothness=Medium}      => {diagnosis=M} 0.1458699          1 0.1458699
```

*Рисунок 1.* Ассоциативные правила со злокачественным диагнозом в правой части

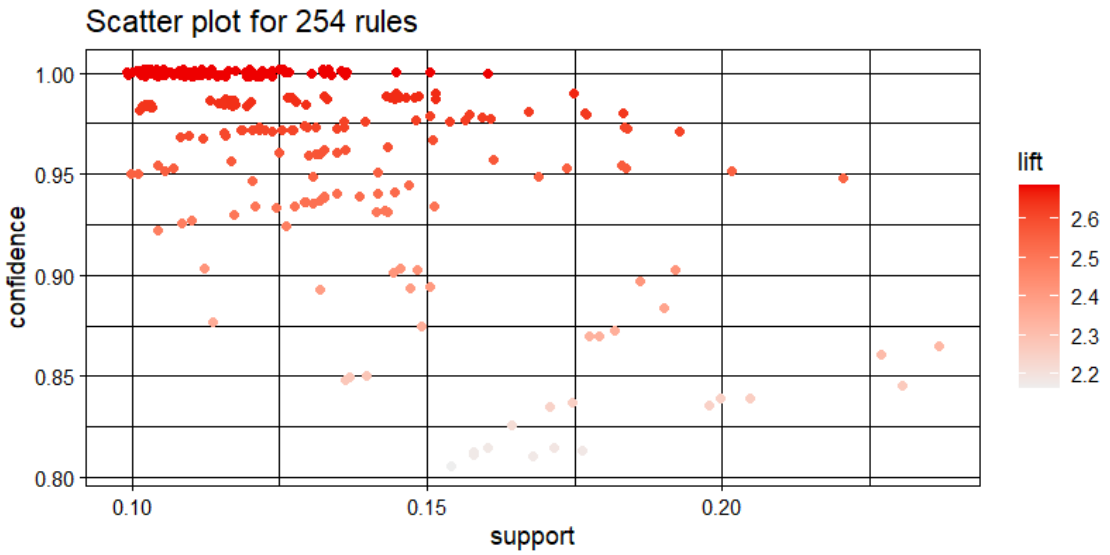



Рисунок 2. Диаграмма рассеяния ассоциативных правил со злокачественным диагнозом в правой части

Итого получилось пять важных правил.

1 Правило {среднее значение площади, среднее значение вогнутости} => {злокачественная опухоль} подразумевает, что сочетание площади и выраженности вогнутых частей контура указывает на заболевание.

2 Правило {среднее значение площади, среднее значение вогнутых точек} => {злокачественная опухоль} значит, что величина площади и количество вогнутых частей контура характерны для злокачественных образований.

3 Правило {среднее значение площади, среднее значение текстуры} => {злокачественная опухоль} выражает влияние площади и стандартного отклонения значений по серой шкале на конечный диагноз.

4 Правило {среднее значение площади, среднее значение симметрии} => {злокачественная опухоль} указывает на взаимосвязь между признаками условия и признаком следствия.

5 Правило {среднее значение площади, среднее значение гладкости} => {злокачественная опухоль} означает, что средние величины площади и локального изменения длин радиусов негативно сказываются на характере образования.

У полученных правил ассоциации значения доверия равны единице, то есть правило полностью достоверно. Хотя значения поддержки невелики, можно сделать вывод, что такого вида правило будет встречаться в каждой десятой записи.

Рисунок 1 показывает, что достаточно средних значений характеристик для выявления злокачественности опухоли. На Рисунке 2 видно, что, хотя правила встречаются редко, они достоверны и признаки будут встречаться вместе часто.

Высокие значения признаков будут только подтверждать неблагоприятный диагноз. Тем не менее, найдём такие правила, где слева будет хотя бы один показатель с высоким значением. Для этого изменим порог поддержки на 0,01. Здесь рассмотрим десять правил.

```
rules_m <- apriori(breastc_trans,
        parameter = list(support = 0.01, confidence = 0.8, minlen = 2),
        appearance = list(rhs = "diagnosis=M", default = "lhs"))
rules_m_high <- subset(rules_m, subset = lhs %pin% "High")
inspect(head(sort(rules_m_high, by = "confidence"), 10))
```

```
     lhs                                rhs           support    confidence coverage
[1]  {area=High}                     => {diagnosis=M} 0.01405975 1          0.01405975
[2]  {compactness=High}              => {diagnosis=M} 0.02284710 1          0.02284710
[3]  {radius=High}                   => {diagnosis=M} 0.03866432 1          0.03866432
[4]  {concave_points=High}           => {diagnosis=M} 0.03866432 1          0.03866432
[5]  {perimeter=High}                => {diagnosis=M} 0.03866432 1          0.03866432
[6]  {area=High, concavity=High}     => {diagnosis=M} 0.01054482 1          0.01054482
[7]  {radius=High, area=High}        => {diagnosis=M} 0.01405975 1          0.01405975
[8]  {area=High, concave_points=High} => {diagnosis=M} 0.01405975 1          0.01405975
[9]  {perimeter=High, area=High}     => {diagnosis=M} 0.01405975 1          0.01405975
[10] {area=High, compactness=Medium} => {diagnosis=M} 0.01054482 1          0.01054482
     count
```

*Рисунок 3.* Ассоциативные правила со злокачественным диагнозом в правой части и хотя бы одним высоким показателем в левой части

Рисунок 3 показывает, что высокие значения характеристик являются отличительной чертой злокачественного течения болезни. Видно, что первые пять правил состоят из одного параметра в левой части: площади, компактности, радиуса, вогнутых точек и периметра. Этого достаточно, чтобы установить злокачественный диагноз.

1 Правило {высокое значение площади, высокое значение вогнутости} => {злокачественная опухоль} показывает, как признаки из левой части воздействуют на результат правой части.

2 Правило {высокое значение площади, высокое значение радиуса} => {злокачественная опухоль} подтверждает, что такие величины площади и радиуса свидетельствуют о злокачественности.

3 Правило {высокое значение площади, высокое значение вогнутых точек} => {злокачественная опухоль} работает аналогичным образом, как и предыдущие правила.

4 Правило {высокое значение площади, высокое значение периметра} => {злокачественная опухоль} содержит в левой части одинакового рода характеристики, которые негативно сказываются на течении болезни.

5 Правило {высокое значение площади, среднее значение компактности} => {злокачественная опухоль} интересует больше остальных. Видно, что показатели признаков отличаются, но величина компактности никак не влияет на правдивость правила.

Значения доверия этих ассоциативных правил всё ещё равны единице в отличие от поддержки. Дело в меньшем количестве высоких значений. Проблема легко решается добавлением записей с высокими показателями.

Теперь перейдём к ситуации, когда опухоль доброкачественная. Таких случаев больше, поэтому значение поддержки будет равно 0,3, а значение доверия останется прежним.

```
rules_b <- apriori(breastc_trans,
      parameter = list(support = 0.3, confidence = 0.8, minlen = 2),
      appearance = list(rhs = "diagnosis=B", default = "lhs"))
inspect(head(sort(rules_b, by = "confidence"), 5))
plot(rules_b)
```

```
    lhs                      rhs             support confidence  coverage
[1] {texture=Low,
     perimeter=Low,
     compactness=Low,
     concave_points=Low} => {diagnosis=B} 0.3725835  0.9860465 0.3778559
[2] {texture=Low,
     perimeter=Low,
     compactness=Low,
     concavity=Low,
     concave_points=Low} => {diagnosis=B} 0.3725835  0.9860465 0.3778559
[3] {texture=Low,
     perimeter=Low,
     area=Low,
     compactness=Low,
     concave_points=Low} => {diagnosis=B} 0.3725835  0.9860465 0.3778559
[4] {texture=Low,
     perimeter=Low,
     area=Low,
     compactness=Low,
     concavity=Low,
     concave_points=Low} => {diagnosis=B} 0.3725835  0.9860465 0.3778559
[5] {radius=Low,
     texture=Low,
     compactness=Low,
     concave_points=Low} => {diagnosis=B} 0.3637961  0.9857143 0.3690685
```

*Рисунок 4.* Ассоциативные правила с доброкачественным диагнозом в правой части

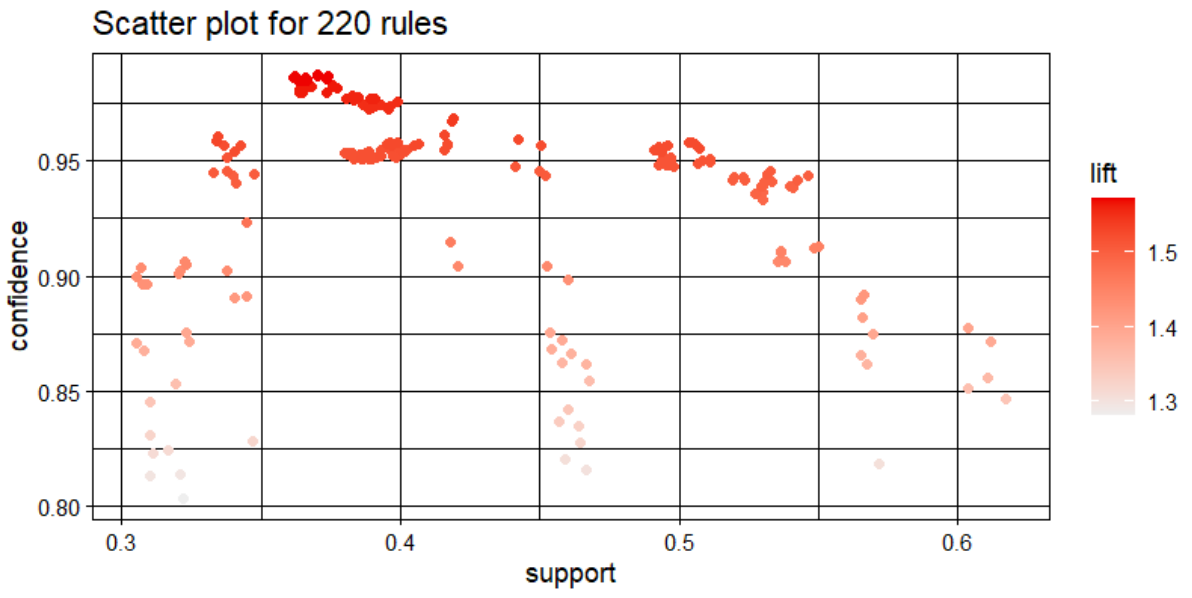



*Рисунок 5.* Диаграмма рассеяния ассоциативных правил с доброкачественным диагнозом в правой части

Получены пять важных правил.

1 Правило {низкое значение текстуры, низкое значение периметра, низкое значение компактности, низкое значение вогнутых точек} => {доброкачественная опухоль} занимает первое место и указывает на связь низких показателей с качеством заболевания.

2 Правило {низкое значение текстуры, низкое значение периметра, низкое значение компактности, низкое значение вогнутости, низкое значение вогнутых точек} => {доброкачественная опухоль} позволяет понять, что на конечный диагноз повлияли представленные характеристики.

3 Правило {низкое значение текстуры, низкое значение периметра, низкое значение площади, низкое значение компактности, низкое значение вогнутых точек} => {доброкачественная опухоль} означает, что низкие величины периметра, площади, компактности, стандартного отклонения значений по серой шкале и количество вогнутых точек негативно сказываются на характере образования.

4 Правило {низкое значение текстуры, низкое значение периметра, низкое значение площади, низкое значение компактности, низкое значение вогнутости, низкое значение вогнутых точек} => {доброкачественная опухоль} выражает влияние признаков на доброкачественность опухоли.

5 Правило {низкое значение радиуса, низкое значение текстуры, низкое значение компактности, низкое значение вогнутых точек} => {доброкачественная опухоль} значит, что сочетание радиуса, текстуры, компактности и количества вогнутых частей контура с их значениями чаще всего гарантирует отсутствие злокачественности.

Значения доверия ассоциативных правил близко к единице, когда как значения поддержки выше в сравнении со злокачественными случаями. На это влияет большее количество доброкачественных диагнозов. Такого вида правила будут встречаться в тридцати процентах случаев.

На Рисунке 4 видно, что на доброкачественность опухоли одновременно влияют несколько признаков. Из Рисунка 5 понятно, что у правил с заданными порогами высокая достоверность и положительная корреляция признаков.

Как и в случае со злокачественной ситуацией, попробуем поискать правила с хотя бы одним средним значением параметра. Значение порога поддержки уменьшится до 0,2, доверия – до 0,5. Также определим ограничение на максимальную длину, равную 4.

```
rules_b <- apriori(breastc_trans,
   parameter = list(support = 0.2, confidence = 0.5, minlen = 2, maxlen = 4),
   appearance = list(rhs = "diagnosis=B", default = "lhs"))
rules_b_med <- subset(rules_b, subset = lhs %pin% "Medium")
inspect(head(sort(rules_b_med, by = "confidence"), 5))
```

```
    lhs                                                    rhs           support   confidence coverage
[1] {texture=Low, perimeter=Low, smoothness=Medium}      => {diagnosis=B} 0.2302285 0.9424460  0.2442882
[2] {radius=Low, texture=Low, smoothness=Medium}         => {diagnosis=B} 0.2284710 0.9420290  0.2425308
[3] {texture=Low, smoothness=Medium, concave_points=Low} => {diagnosis=B} 0.2530756 0.9411765  0.2688928
[4] {radius=Low, compactness=Low, symmetry=Medium}       => {diagnosis=B} 0.2214411 0.9402985  0.2355009
[5] {perimeter=Low, compactness=Low, symmetry=Medium}    => {diagnosis=B} 0.2214411 0.9402985  0.2355009
```

*Рисунок 6.* Ассоциативные правила с доброкачественным диагнозом в правой части и хотя бы одним средним показателем в левой части

Получилось пять правил.

1 Правило {низкое значение текстуры, низкое значение периметра, среднее значение гладкости} => {доброкачественная опухоль} со средним показателем гладкости отражает влияние таких характеристик, как текстура, периметр, гладкость, и их значений на благоприятный диагноз.

2 Правило {низкое значение радиуса, низкое значение текстуры, среднее значение гладкости} => {доброкачественная опухоль} похоже на предыдущее и показывает воздействие признаков из левой части на правую.

3 Правило {низкое значение текстуры, среднее значение гладкости, низкое значение вогнутых точек} => {доброкачественная опухоль} значит, что полученные величины гладкости, стандартного отклонения значений по серой шкале и количества вогнутых точек влияют на качество образования.

4 Правило {низкое значение радиуса, низкое значение компактности, среднее значение симметрии} => {доброкачественная опухоль} содержит другую характеристику со средним значением – симметрию. При совпадении признаков из правила, высока вероятность, что опухоль окажется доброкачественной.

5 Правило {низкое значение периметра, низкое значение компактности, среднее значение симметрии} => {доброкачественная опухоль} подразумевает благоприятный диагноз в большинстве случаев при наблюдении указанных параметров.

Из полученных выше правил только характеристики симметрия и гладкость принимают средние значения, а остальные низкое. Вполне возможно, что при использовании других значений поддержки и доверия было бы выделено большее количество правил и результат получился бы более разнообразным. Видно, что у данных правил достаточно высокая вероятность появления, согласно которой они встречаются в каждой пятой записи. Их достоверность ожидаемо ниже, хотя всё ещё достаточно высока.

**Заключение.** Таким образом, в статье была рассмотрена реализация алгоритма Apriori на языке R для анализа медицинского набора данных Breast Cancer Wisconsin (Diagnostic) Data Set с численными значениями, которые были преобразованы в категориальные для дальнейшей работы.

Преобразование данных в подобных наборах необходимо для корректной работы алгоритма. Иначе Apriori будет искать правила с аналогичными числовыми значениями в записях. Пороги будут минимальными, потому что точное совпадение величин площади, периметра или любого другого признака маловероятно. С преобразованием, где числовые значения разбиты на три категории, легко выявить ассоциативные правила, полезные для анализа взаимосвязей характеристик в медицинских данных.

Применение алгоритма Apriori в медицинской сфере демонстрирует его потенциал в выявлении скрытых закономерностей, которые можно использовать для улучшения диагностики, прогнозирования и персонализации лечения. Важно учитывать, что точность и интерпретируемость полученных результатов во многом зависят от качества исходных данных и параметров алгоритма – поддержки и доверия.

Найденные ассоциативные правила могут помочь выявить, что, например, средние и высокие значения характеристик коррелируют с постановкой злокачественного диагноза, тогда как низкие значения чаще указывают на доброкачественную природу образования.

Если диагнозы одного вида превалируют над другими, полезно сделать нахождение ассоциативных правил для каждого случая отдельно. Для исследования конкретных случаев с определённой группой признаков можно выполнить сортировку по необходимым критериям.

Другие результаты исследований с помощью алгоритма Apriori можно найти в статьях [3-5].

## Список литературы

[1] Интернет ресурс «Kaggle» [Электронный ресурс] / Breast Cancer Wisconsin (Diagnostic) Data Set. – 2017. – Режим доступа: https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data. – Дата доступа: 17.03.2025.

[2] Street, W. N. Nuclear feature extraction for breast tumor diagnosis / W. N. Street, W. H. Wolberg, O. L. Mangasarian // Electronic imaging. - 1993. - С. 3- 5.

[3] Лянцевич, В. С. Использование алгоритма Apriori для анализа данных выбросов загрязняющих веществ в атмосферу / В. С. Лянцевич; науч. рук. Н. В. Марковская // Студенческий вестник. – 2024. – Т. Ч.7. – № 42(328). – С. 66-67.

[4] Лянцевич, В. С. История создания и разработка алгоритма Apriori / В. С. Лянцевич; науч. рук. Н. В. Марковская // Молодежь и наука XXI века: реалии и перспективы: Междунар. науч.-практ. конф., Астана, 12 апреля 2024 г. – Астана: ИПЦ; Esil University; 2024. – С. 771-772.

[5] Лянцевич, В. С. Библиотеки языка Python для анализа данных с помощью алгоритма Apriori / В. С. Лянцевич; науч. рук. Н. В. Марковская // От Альфа к Омеге...: сб. материалов XIV Респ. науч.-практ. конф.-конкурса науч.-исслед. работ учащихся средних, средних спец. учебных заведений и студентов вузов; От Альфа к Омеге...; / Гродн. гос. ун-т им. Янки Купалы; гл. ред. А. В. Кузьмич; редкол.: А. В. Кузьмич [и др.]. – Гродно: ГрГУ им. Янки Купалы, 2024. – С. 71-72.

## Авторский вклад

**Лянцевич Валерия Сергеевна** – выполнение исследования, написание программного алгоритма, анализ полученных результатов.

**Марковская Наталья Вацлавовна** – постановка задачи исследования, анализ полученных результатов.

# MEDICAL DATA ANALYSIS USING THE APRIORI ALGORITHM IN R

***V.S. Lyantsevich***
*Student of the Department of Mathematical and Information Support of Economic Systems, YKSUG*

***N.V. Markovskaya***
*Associate professor of the Department of Mathematical and Information Support of Economic Systems, YKSUG, PhD of Mathematical Sciences, Associate Professor*

**Abstract.** The article considers the use of the Apriori algorithm for identifying association rules in a medical breast tumor dataset. The features of data structures are analyzed, they are modified for the application of the algorithm, and the implementation of Apriori is assessed. The study is based on the transformed dataset, key patterns are identified, and the influence of the algorithm parameters on the quality of the extracted rules is analyzed. Conclusions are made about the applicability of the Apriori algorithm in the tasks of analyzing medical data, and recommendations are offered for its configuration for various types of data.

**Key words:** association rules, Apriori algorithm, medical data, breast tumor.

УДК 621.396

# ПРИМЕНЕНИЕ МУЛЬТИФОКУСИРОВАННЫХ АНТЕНН В ЗАДАЧАХ МЕДИЦИНСКОЙ ДИАГНОСТИКИ

***Ю.Е. Седельников***
*Профессор кафедры Радиофотоники и микроволновых технологий КНИТУ-КАИ, доктор технических наук, профессор*
sedhome2013@yandex.ru

***Д.А. Веденькин***
*Доцент кафедры Радиофотоники и микроволновых технологий КНИТУ-КАИ, кандидат технических наук, доцент*
denis_ved@mail.ru

***А.Ф. Гильфанова***
*Аспирант кафедры Радиофотоники и микроволновых технологий КНИТУ-КАИ*
gilfanova-almira@mail.ru

***Ю.Е. Седельников***
*Окончил Казанский авиационный институт. Область научных интересов связана с оценкой электромагнитной совместимости, разработкой и анализом антенн и СВЧ устройств и их технических приложений.*

***Д.А. Веденькин***
*Окончил Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ. Область научных интересов связана с оценкой электромагнитной совместимости, разработкой и анализом антенн и СВЧ устройств, методами обработки информации и статистической теорией радиотехнических систем.*

***А.Ф. Гильфанова***
*Окончила Казанский национальный исследовательский технический университет им. А.Н. Туполева-КАИ. Область научных интересов связана с анализом антенн и СВЧ устройств, методами передачи и обработки информации в телекоммуникационных сетях.*

**Аннотация.** Рассмотрен способ фокусировки электромагнитной энергии в несколько точек в области зоны ближнего излученного поля за счет работы системы на нескольких частотах. Представлены результаты моделирования антенной решетки, сфокусированной в две точки пространства. Продемонстрированные иллюстрации показывают возможность формирования нескольких максимумов энергии поля в заданных областях пространства. Подобные антенные системы могут с применением эффекта мультифокусировки может усовершенствовать измерения температуры в задачах диагностики различных патологий.

**Ключевые слова:** сфокусированные антенны, область фокусировки, мультифокусировка, зона ближнего излученного поля, распределение поля.

**Введение.** В настоящее время все больший интерес в теории и технике антенн вызывает область ближнего излученного поля, а также возможность формирования сфокусированного излучения на расстояниях близких к геометрическим размерам апертуры. Это обусловлено тем, что системы, работающие в области ближнего излученного поля, широко распространены в различных областях жизни человека, в том числе в медицине [1]. Диагностика по средством измерения температуры тела, а также температуры внутренних тканей биологического объекта является одной из важных задач. На сегодняшний день существует множество работ, в которых изучались различные методы повышения эффективности диагностики с помощью радиотермометрии: авторы

исследовали влияние типа антенны, размеров апертуры, а также использования принципа фокусировки на качество измерений [2-3]. В данной работе используется другой подход к усовершенствованию измерения температуры внутренних тканей биологических объектов – применение эффекта мультифокусировки.

**Мультисфокусированные антенны.** В качестве оптимизации диагностики различных патологий в данной работе предлагается применение мультисфокусированных антенных систем. Такие системы способны формировать несколько максимумов излучения в зоне ближнего излученного поля (рисунок 1), что открывает новые возможности в задачах терапевтического воздействия на ткани биологических объектов [4-5], а при реализации соответствующих фазовых задержек осуществлять одновременную радиотермометрию нескольких пространственно-разнесенных областей конечных размеров в тканях биологического объекта.

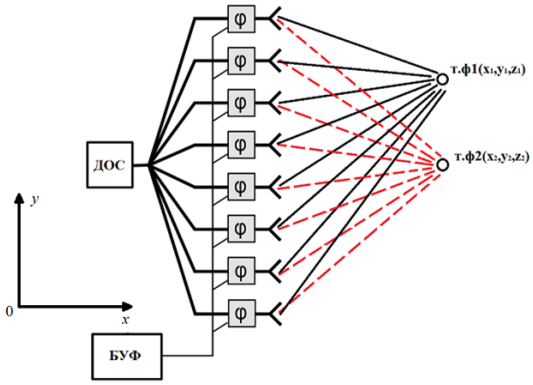



*Рисунок 1.* Мультисфокусированная антенная решетка с многочастотным излучением, ДОС – диаграммо-образующая схема, БУФ – блок управления фазами

Рассмотрим модель антенны, сфокусированной в две точки пространства $F_1$ = (9 см, −0.9 см) и $F_2$ = (9 см, 0.9 см). Это становится возможным при следующих параметрах моделирования: частоты $f_1$ = 95 ГГц и $f_2$ = 105 ГГц, количество излучателей — 21, размер апертуры — 6 см, расстояния до точек фокусировки — 9 см, расстояние между точками фокусировки — 1.8 см.

При заданных параметрах системы получаем следующее пространственное распределение областей приема энергии электромагнитных полей в задаче радиотермометрии (рисунок 2). На представленном рисунке имеются два ярко выраженных максимума, которые соответствуют заданным точкам фокусировки $F_1$ и $F_2$.

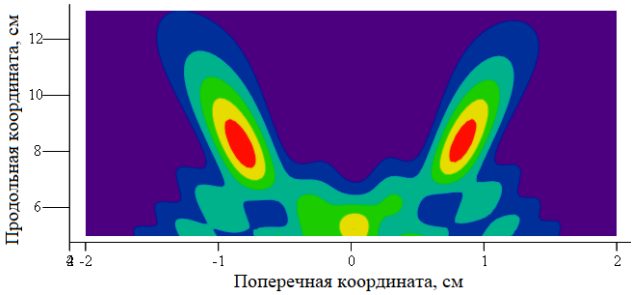



*Рисунок 2.* Пространственное распределение энергии электромагнитного поля, сфокусированного в две точки пространства $F_1$ и $F_2$ на частотах $f_1$ и $f_2$, соответственно

Рассмотрим картину распределения областей приема энергии электромагнитных полей вдоль оси, параллельной линейной антенне и проходящей через обе точки фокусировки. Данная картина представлен на рисунке 3, где отчетливо видны два максимума энергии поля в точках фокусировки $F_1$ и $F_2$.

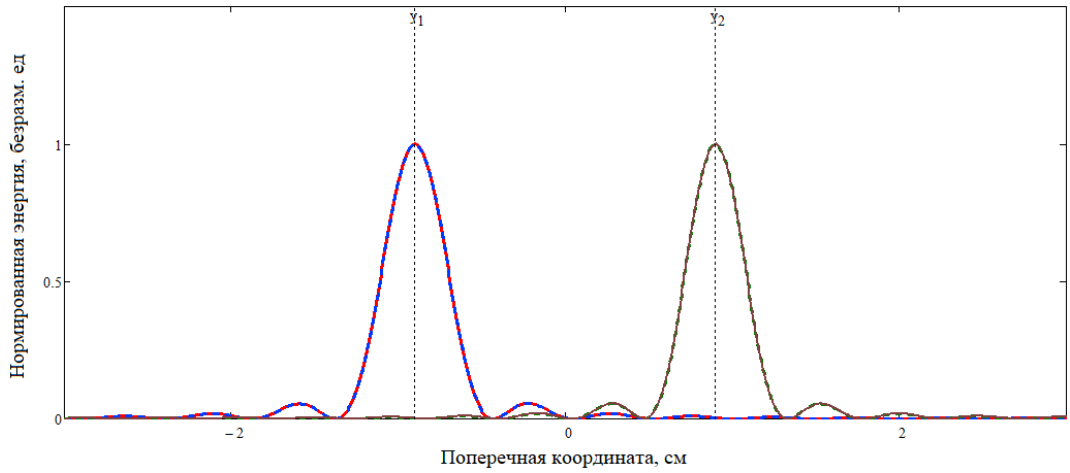



*Рисунок 3.* Распределение энергии электромагнитных полей по поперечной координате, синим с красным цветами обозначено распределение области приема энергии электромагнитного поля на частоте $f_1$ и зеленым обозначено распределение области приема энергии электромагнитного поля на частоте $f_2$

Рассмотрим возможность осуществления радиотермометрии областей конечных размеров, располагающихся на различных фокусных расстояниях (рисунок 4).

*Рисунок 4.* Пространственное распределение областей приема энергии электромагнитных полей при ассиметричном расположении точек фокусировки

В этом случае можно отметить, что в точке $F_1$ = (6 см, −0.9 см) возможность приема энергии электромагнитного полея выше, чем в точке $F_2$ = (9 см, 0.9 см). Этот эффект можно обосновать тем, что расстояние от апертуры до точки $F_1$ меньше, чем до точки $F_2$, и следовательно принимаемая энергия электромагнитного поля в области $F_1$ выше.

**Заключение.** Возможность приема энергии электромагнитных полей из сразу нескольких заданных областей пространства открывает новые возможности применения сфокусированных антенн, в частности в области медицинской диагностики. Таким образом, применение мультисфокусированных антенных систем в области радиотермометрии может позволить оптимизировать процесс измерения температуры тела и внутренних тканей биологических объектов, что позволит повысить эффективность диагностики различных патологий.

## Список литературы

[1] Седельников Ю. Е., Тестоедова Н. А. Антенны, сфокусированные в зоне ближнего излученного поля: монография. Красноярск: Сиб. гос. аэрокосмич. ун-т., 2015. – 308 с.

[2] Седельников Ю. Е., Кубланов В. С., Потапова О. В. Сфокусированные антенны-аппликаторы в задачах диагностической радиотермометрии // Журнал радиоэлектроники, 2018, № 7. – С. 1-7.

[3] Седельников Ю. Е., Потапова О. В., Садыков А. Р., Скачков В. А. Сфокусированные антенны в задачах контактной радиотермометрии // Журнал радиоэлектроники, 2021, № 3. – С. 1-26.

[4] Гильфанова А. Ф. Формирование мультисфокусированного излучения // Прикладная электродинамика, фотоника и живые системы - 2024: Сборник тезисов докладов XI Международной молодежной научно-технической конференции молодых ученых, аспирантов и студентов, Казань, 11–12 апреля 2024 года. Казань: ИП Сагиев А.Р., 2024. – С. 69–70.

[5] Седельников Ю. Е., Веденькин Д. А., Гильфанова А. Ф. Применение метода Ильина-Морозова и в задачах формирования мультисфокусированного электромагнитного излучения // BIG DATA и анализ высокого уровня: Сборник научных статей IX международной научно-практической конференции: в 2х частях, Минск, 17–18 мая 2023 года. Минск: Белорусский государственный университет информатики и радиоэлектроники, 2023. – С. 261–266.

## Авторский вклад

**Седельников Юрий Евгеньевич** – руководство исследованием по оценке качества работы мультисфокусированных антенн.

**Веденькин Денис Андреевич** – постановка задачи исследования, описание принципа работы сфокусированных апертур в зоне ближнего излученного поля, анализ полученных результатов.

**Гильфанова Альмира Фанисовна** – моделирование антенной системы с использованием эффекта мультифокусировки, формирование структуры статьи.

# APPLICATION OF MULTIFOCUSED ANTENNAS IN MEDICAL DIAGNOSTIC PROBLEMS

***Y.E. Sedelnikov***
*Professor of Radiophotonics and Microwave Technologies Department of KNRTU-KAI, Doctor of Technical Sciences, Professor*

***D.A. Vedenkin***
*Associate Professor of Radiophotonics and Microwave Technologies Department KNRTU-KAI, PhD, Associate Professor*

***A.F. Gilfanova***
*Postgraduate student of Radiophotonics and Microwave Technologies Department of KNRTU-KAI*

**Abstract.** The method of focusing electromagnetic energy to several points in the region of the near radiated field zone due to the operation of the system at several frequencies is considered. The results of modeling of the antenna array focused in two points of space are presented. The demonstrated illustrations show the possibility of forming several field energy maxima in given regions of space. Such antenna systems can improve temperature measurements in diagnostics of various pathologies by using the effect of multifocusing.

**Keywords**: focused antennas, focusing area, multifocusing, near radiated field zone, field distribution.

УДК 004.021:004.75

# ОПТИМИЗАЦИЯ АЛГОРИТМОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ РАСПОЗНАВАНИЯ БЕЛОРУССКИХ ДОРОЖНЫХ ЗНАКОВ В УСЛОВИЯХ ДИСБАЛАНСА ДАННЫХ

***Л.Ю. Шилин***
*декан факультета информационных технологий и управления БГУИР, доктор технических наук, профессор*
dekfitu@bsuir.by

***К.Д. Сморщок***
*студент факультета информационных технологий и управления БГУИР*
s.kirill150282@mail.ru

***Д.Д. Хмыз***
*ассистент кафедры вычислительных методов и программирования БГУИР*
d.khmyz@bsuir.by

***Л.Ю. Шилин***
*Окончил Минский радиотехнический институт. Является автором более 170 научных и научно-методических работ. Проводит научные исследования по анализу и синтезу дискретных систем фазовой синхронизации.*

***К.Д. Сморщок***
*Обучается на кафедре информационных технологий автоматизированных систем Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой алгоритмов машинного обучения для решения задач компьютерного зрения.*

***Д.А. Хмыз***
*Окончила Белорусский национальный технический университет, специальность «профессиональное обучение(информатика). Область научных интересов связана с разработкой алгоритмов машинного обучения для решения задач компьютерного зрения.*

**Аннотация.** Разработана система автоматического распознавания белорусских дорожных знаков с использованием модели YOLO (You Only Look Once). Проведен анализ данных, включая сбор, предобработку и балансировку классов. Исследованы методы оценки качества модели, такие как кроссвалидация, метрики точности и полноты, сценарные тесты. Показано, что применение андерсемплинга и комбинации YOLO с методами аугментации позволяет достичь средней точности (mAP) 0.570. Результаты визуализированы с помощью PR-кривых и матриц разброса.

**Ключевые слова:** компьютерное зрение, YOLO, распознавание дорожных знаков, дисбаланс классов, андерсемплинг.

**Введение.** Создание автономных транспортных систем требует решения задач восприятия окружающей среды, среди которых ключевой является распознавание дорожных знаков. Актуальность исследования обусловлена необходимостью повышения безопасности дорожного движения и интеграции систем автопилота в реальные условия.

Основная цель работы — разработка прототипа системы на базе модели YOLO, способной распознавать белорусские дорожные знаки в реальном времени. Для достижения цели решены следующие задачи:

1 Выполнены сбор и предобработка данных из открытых датасетов.

2 Проведен анализ дисбаланса классов и применение методов балансировки.

3 Обоснован выбор архитектуры YOLO.

4 Проведена оценка эффективности модели с использованием метрик качества.

**Используемые методы и алгоритмы:**

1 Сбор и предобработка данных

Использованы датасеты:

– «df-full-pic» (104 тыс. изображений) — для общего обучения;

– «Russian traffic sign images dataset» (180 тыс. изображений) — адаптирован под белорусские знаки.

Предобработка включала:

– нормализацию размеров изображений;

– андерсемплинг классов с числом экземпляров >2000 (снижение дисбаланса);

– аугментацию (повороты, изменение яркости) для редких классов.

2 Выбор модели.

Архитектура YOLO выбрана исходя из того, что обладает

– высокой скоростью (обработка 45 FPS на NVIDIA Jetson);

– необходимой точностью ( mAP 0.570, что выше аналогов (SSD: 0.520; Faster R-CNN: 0.550));

– поддержкой встраиваемых систем.

3 Используемые методы оценки:

– кросс-валидация (K=5): Оценка устойчивости модели;

– метрики: Precision (0.68), Recall (0.62), F1-score (0.65);

– сценарные тесты: Проверка работы при дожде, снеге, ночном освещении.

Проведен анализ дисбаланса классов. Исходный датасет содержал 198 классов с сильным дисбалансом (Рис. 1):

– максимальный класс: 24 264 экз. (5_19_1);

– минимальный класс: 1 экз. (3_33).

*Рисунок 1.* График распределения количества экземпляров от их классов.

В результате андерсемплинга получено:

– среднее количество экземпляров: 313.33 (ранее 673.28);

– стандартное отклонение снижено с 2259.58 до 525.03.

**Результаты обучения**

– PR-кривая (Рис. 2): Средняя точность (mAP) — 0.570;

– лучшие классы: «Стоп» (Precision=0.89), «Уступите дорогу» (Precision=0.85);

– Слабые классы: Редкие знаки (Precision <0.30);

– Матрица разброса: Выявлена корреляция между позицией знака (x, y) и его размерами (width, height).

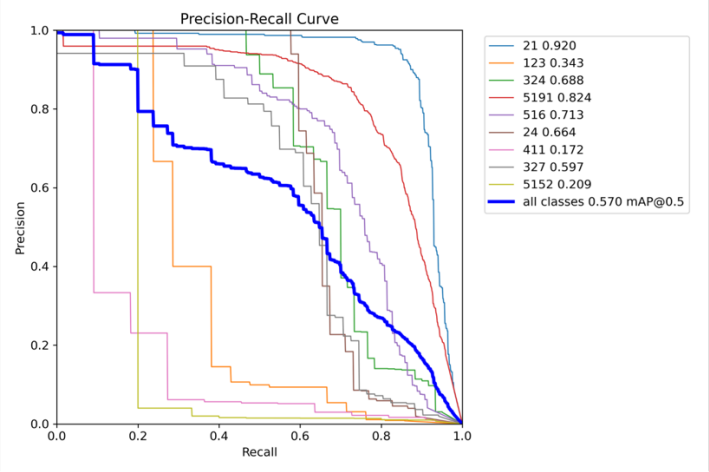



*Рисунок 2.* pr кривая для Yolov

**Принципы оптимизации**

– балансировка данных: Комбинация андерсемплинга и синтетической аугментации;

– использование взвешенных потерь: Учет весов классов в функции потерь;

– аппаратная оптимизация: Квантование модели для Raspberry Pi.

**Заключение.** Разработана система, которая имеет достаточную точность для распознавания основных дорожных знаков, однако требует доработки для редких классов. Дальнейшие исследования будут направлены на:

– сбор данных для малочисленных классов;

– тестирование на реальных дорогах Беларуси;

– интеграцию с системами управления автомобилем.

## Список литературы

[1] Redmon J. et al. YOLOv3: An Incremental Improvement // arXiv, 2018.

[2] Гудфеллоу Я. Глубокое обучение. М.: ДМК Пресс, 2022.

[3] Krizhevsky A. ImageNet Classification with Deep Convolutional Neural Networks // NIPS, 2012.

## Авторский вклад

**Шилин Леонид Юрьевич** – руководство исследованием по разработке алгоритмов глубокого обучения для систем распознавания.

**Сморщок Киррил Дмитриевич** – сбор данных, предобработка, обучение модели, анализ результатов.

**Хмыз Дарья Дмитриевна** – постановка задачи распознавания дорожных знаков.

# DEVELOPMENT OF A SYSTEM FOR RECOGNIZING BELARUSIAN ROAD SIGNS BASED ON YOLO MODEL

***V.F. Alekseev***
*Dean of the Faculty of Information Technologies and Control in BSUIR, Doctor of Technical Sciences, Professor*

***D.V. Likhachevsky***
*Student of department of Information Technologies in Automated Systems in BSUIR*

***G.A. Piskun***
*Assistant of department of Computational Methods and Programming in BSUIR*

**Abstract.** The paper presents the development of an automatic system for recognizing Belarusian road signs using the YOLO model. Data analysis, including collection, preprocessing, and class balancing, was conducted. Evaluation methods such as cross-validation, precision-recall metrics, and scenario tests were applied. The use of undersampling and YOLO-augmentation combination achieved a mean Average Precision (mAP) of 0.570. Results are visualized via PR-curves and scatter matrices

**Keywords**: computer vision, YOLO, road sign recognition, class imbalance, undersampling.

УДК 621.865

# AUTONOMOUS MOBILE ROBOT NAVIGATION AND MOTION CONTROL PROBLEMS AND ANALYSIS OF EXISTING METHODS

***Yang Shixin***
*PhD student of the Department of «Electric drive and automation of industrial installations and technological complexes» of BNTU*
yang_shi_xin@163.com

***S.A. Pauliukavets***
*Associate Professor of the Department of «Electric drive and automation of industrial installations and technological complexes» of BNTU, PhD of Technical Sciences, Associate Professor*
sap@bntu.by

***A.A. Velchenko***
*Associate Professor of the Department of «Electric drive and automation of industrial installations and technological complexes» of BNTU, PhD of Technical Sciences, Associate Professor*
anna.velchenko@gmail.com

***Yang Shixin***
*PhD student at the Belarusian National Technical University. The area of scientific interests is related to the study of control algorithms for wheeled mobile robots.*

***S.A. Pauliukavets***
*PhD in Technical Sciences, Associate Professor of the «Electric Drive and Automation of Industrial Installations and Technological Complexes» Department of the Belarusian National Technical University. The area of scientific interests is related to the development of methods and algorithms for controlling mobile robots.*

***A.A. Velchenko***
*PhD in Technical Sciences, Associate Professor of the «Electric Drive and Automation of Industrial Installations and Technological Complexes» Department of the Belarusian National Technical University. The area of scientific interests is related to the development and research of neural network controllers for electric drives.*

**Abstract.** This paper briefly introduces the characteristics of the current mobile robots , sorts out the current robot navigation problems, introduces the robot positioning and navigation algorithms, and then introduces the grid, topology and geometric feature maps that describe the environment model. Map-based further research for subsequent autonomous navigation.

**Keywords**: mobile robot, motion control, navigation, SLAM, LiDAR.

**Introduction.** In recent years, human beings are constantly pursuing to free their hands and replace human work with robots to improve production efficiency or quality of life. As the jobs being replaced become more and more complex, the degree of robot automation and intelligence is getting higher and higher. Both academia and industry have set off an upsurge in robot intelligence research. Intelligence requires robots to be able to complete established tasks with as little human intervention as possible. At present, intelligent robot technology is widely used in various fields of production and life, such as production robot arms for industrial production, sweeping robots for household cleaning, auxiliary robots for medical care, and storage robots for logistics. UAVs for aerial photography, underwater robots for deep-sea scientific research and exploration and other practical applications.

For intelligent mobile robots such as storage robots, drones, and underwater robots, they usually have functions such as environment perception, positioning and navigation, path planning, autonomous movement, and task operations. Among them, environmental perception is an important step in realizing intelligence. Obtaining important state information such as its own position, speed, attitude, and obstacle position through the monitoring of the environmental state by the sensor is the prerequisite for the mobile robot to independently perform complex tasks in an unknown environment.

The exploration of mobile robots originated in the late 1950s. In 1959, two scientists in the United States, Ingeberg and DeVore, developed the world's first mobile robot. This robot mainly identifies underground routes for simple exploration. Path planning [1] from 1966 to 1972, the Stanford Research Institute for International Studies successfully developed the Shakey mobile robot, which is mainly used indoors and can be controlled in real time in complex environments. With the continuous breakthrough of robot research technology and the continuous expansion of application fields, foreign defense researchers began to invest in robots and conduct research in the military field. Since the 1980s, the United States began to carry out the ALV program. At the same time, Japan Carry out robot research programs working in complex environments, and the EU launches underwater robot programs, etc. Since the 21st century, the development of mobile robots has advanced by leaps and bounds. In 2005, the American adept company researched the PatrolBot mobile robot, which is now used in shopping malls, factories and other occasions. It can accurately locate and navigate autonomously in complex environments, basically replacing manual services; Kiva systems The mobile robot developed by the company scans the QR code for autonomous navigation, helps warehouse personnel to process orders, and greatly improves the efficiency of inventory processing.

With the rapid development of science and technology brought about by informatization and intelligence, the demand for robots in society is also gradually increasing. replace. According to the Internet Data Center [2], by the end of 2020, the world's industrial and application expenses related to robots exceeded 188 billion US dollars, and the huge amount promoted the continuous expansion of the robot market. Since robots can help us solve many problems, the government has introduced policies to help the robot industry to flourish, making the society pay more and more attention to robots, which has broad development prospects, so its research is of great significance. At the same time, robots are a sunrise industry, and has also been favored and invested by the market and entrepreneurs.

Mobile robot is a comprehensive system, which integrates multiple functions such as environment perception, dynamic decision-making, behavior control and execution. In various fields, mobile robots have been widely used. For example, in the military, they can replace humans in dangerous areas to complete dangerous tasks such as reconnaissance and mine clearance; in industrial production, they can replace humans in harsh environments to complete equipment testing, Cargo handling and other tasks; in civilian use, as a blind guide vehicle, it can provide assistance in daily life for the blind; in scientific research, it can assist humans in mineral exploration or alien planet exploration and other activities. Mobile robots are closely related to many disciplines, such as computer science, communication, signal processing, pattern recognition, control theory, and artificial intelligence.

With the development of computer and robot technology, the development of robots has gone through three generations of evolution [3]: the first generation is a robot that works in a «teaching-reproduction» mode. This generation of robots has been widely used at present. It is a system that relies on people to give programs and repeatedly perform various operations. However, this generation of robots does not have sensor feedback information, so it cannot improve its own movement quality by continuously obtaining external information during the operation process, which greatly limits its application range. The second generation is an off-line programming robot, which has a certain ability of self-adaptation and perception of the external environment.

This kind of robot is equipped with simple internal and external sensors, which have sensing capabilities such as «vision», «tactile» and «auditory», and can detect the external environment, obtain information such as its actual position and direction, and then Under the algorithm and program instructions, the feedback information is used to adjust the action process. This generation of robots is closely related to research fields such as sensors, bionics, microcomputer technology, and control theory that have developed rapidly in recent years, and a few have been practically applied so far. The third-generation robot is an intelligent robot, which can perceive external objects and the environment, and has the ability to accurately process complex information and make autonomous decisions about its own behavior. This kind of robot is equipped with a variety of sensors, and can «fusion» the information detected by multiple sensors (multi-sensor information fusion) [4], and then make corresponding autonomous decisions to complete specific tasks. It has a strong self-adaptive ability, can effectively adapt to environmental changes, and is an intelligent robot with autonomy (autonomous decision-making) and learning functions. At present, robots are developing toward autonomy, complexity, and intelligence. In human life, robots are more and more closely connected with us.

**Navigation problems for mobile robots.** In the research of mobile robot related technologies, navigation technology is the core of its research. Common navigation methods for mobile robots include electromagnetic navigation, map model matching navigation based on environmental information, landmark navigation based on various navigation signals, taste navigation, sound navigation, visual navigation, etc. A commonly used navigation method at present is «following the path guidance», that is, the mobile robot navigates by responding to certain external continuous path reference information that can be perceived. The path-following guidance method differs according to the sensor used to identify the guidance line: embedded guidance based on the principle of electromagnetic induction, track in the mechanical method, magnetic strip in the magnetic guidance method, reflective tape in the optical guidance method, and Guidance methods based on computer vision. For example, magnetic guidance is to lay magnetic nails or metal wires on the moving path of the mobile robot, and determine the position and posture of the mobile robot by detecting the characteristic information generated by them. From the perspective of mobile robot navigation, the advantages of traditional navigation methods such as magnetic guidance are high reliability, but at the same time there are disadvantages such as single function, poor flexibility, weak adaptability to the environment, and high maintenance costs. Recognizing target features while moving, and also unable to detect obstacles, limits the application of these methods in mobile robot navigation. Visual navigation has the characteristics of high level of intelligence and rich information, and has been widely used in the autonomous navigation of mobile robots in recent years. Visual navigation is a navigation method based on local vision. The main function is to use the camera installed on the mobile robot to obtain video or images, simulate the visual function of the human eye through computer technology, and drive the mobile robot from the image or image sequence. The information of the road area is extracted, so that the mobile robot can observe and understand the world through vision like a human being, and has the ability to adapt to the environment autonomously. This is also the ultimate research goal of the visual navigation of the mobile robot.

For humans, more than 90% of the information comes from visual perception of the environment, and human vision provides humans with the most detailed and reliable information about the surrounding environment. Therefore, the use of computer vision to realize the navigation of mobile robots has some advantages that other navigation methods do not have [5]. First of all, for navigation, visual information not only provides distance information of objects in the field of view, but also provides characteristic information such as color and shape of objects in the field of view, which cannot be obtained when other active sensors are used for navigation. These unique information can make mobile robots have higher intelligence and become more intelligent. Secondly, for navigation, road direction and edge features can also be learned from visual information, which enables mobile robots to make high-level path planning that is more in line

with human traffic rules. In addition, the two advantages of computer vision technology in terms of information volume and acquisition speed determine that it plays a vital role in the development of mobile robots. In recent years, a series of research results in the field of visual navigation have fully reflected the great potential of visual navigation technology. The continuous progress of visual navigation technology is very positive for the realization of highly intelligent mobile robots and the improvement of the current level of technology. impetus. Road detection technology is one of the key technologies in the visual navigation of mobile robots. It mainly recognizes information such as color and edges from road images to understand the road environment and obtain location information for mobile robots to navigate autonomously. In this paper, the monocular vision system is used to realize the visual navigation of the mobile robot. By detecting the color and edge information in the road image, the position of the road guide line is determined, and then the navigation information of the mobile robot is obtained through the camera calibration model to realize its autonomous movement. Laying metal wires, magnetic nails and multi-eye visual navigation, etc. This method enhances the flexibility of the system and reduces maintenance costs. On the other hand, the quality of road detection and tracking directly affects the quality of mobile robot visual navigation. accuracy. Therefore, the research on road detection technology in mobile robot visual navigation has extensive application value and great theoretical significance.

**Methods for Perceiving Mobile Robots Moving in Urban Environments.** Simultaneous localization and mapping (SLAM) is a method that enables mobile robots to autonomously carry out environmental perception analysis and obtain the state of themselves and the environment in an unknown environment. In the field of robotics research [6], SLAM has always been the focus of research. Due to the complexity and unknownness of the environment, the complexity of the sensor hardware system, the diversity of measurement methods, and the inevitable errors in measurement results, how to obtain more accurate and robust state estimation and establish higher precision through sensor measurement The map can achieve precise control of the robot while reducing the hardware cost. These problems are still not well resolved and need further breakthroughs in sensor hardware technology, data acquisition technology, software and hardware integration technology, data processing technology, map construction and algorithm technology. and improvements. This paper conducts research on the simultaneous positioning and map construction related issues such as the acquisition method of the internal parameters of the lidar, the selection optimization algorithm of the occupancy grid map parameters, and the independent system design of the robot map construction. Methods for precision maps.

*SLAM research overview*, SLAM includes two parts of positioning and map construction, which means that the robot moves from one position to another in a completely unknown environment. In this process, the robot uses the information obtained by the sensor to estimate the change of its position And build the map incrementally according to the obtained innovation. In fact, the two are inextricably linked. On the one hand, in order to obtain the precise positioning of the robot in the environment, a correct map is required; on the other hand, in order to build a good map by adding the scanned elements to the map, the correct positioning of the robot is required. SLAM can be widely used in robot positioning and navigation, drones, unmanned driving, VR/AR and other fields.

The SLAM algorithm was first proposed at the ICRA (IEEE Robotics and Automation Conference) in 1986, and people began to use probabilistic methods to think about robot perception problems.

So far, the development of SLAM has roughly gone through three stages [7]. The first stage (1986-2004) is called the budding development period. During this period, some important probability methods were proposed, such as Extended Kalman Filter (EKF), Particle Filter (PF), Maximum Likelihood Estimation (MLE), and related issues of data association. The second stage (2004-2015) is called the algorithm analysis period. At this stage, a lot of research has been done on the characteristics of SLAM algorithms, including observability, convergence, and continuity. At the same time, a large number of SLAM algorithms were proposed during this period. The third

stage (2015-present), known as the robust perception period, people began to pursue the robustness of the SLAM algorithm, and expected the algorithm to perceive the surrounding environment more intelligently. From the type of sensor used, SLAM is mainly divided into lidar-based SLAM algorithm and computer vision-based visual SLAM algorithm.

At present, commonly used sensors for mobile robot SLAM include: single-line lidar, multi-line lidar, monocular camera, binocular camera, RGB-D camera, etc. They are the RplidarA3 laser radar from Silan Technology, the Pandar64-line mechanical laser radar from Hesai Technology, the DS-2CD7T47DWD-IZ camera from Hikvision, and the binocular camera from Xiaomi Smart.

*A review of lidar SLAM algorithms*. The research on lidar-based SLAM technology started relatively early, but due to the high cost of lidar, it is not widely used. However, in recent years, with the development of lidar technology, the cost has dropped rapidly, and unmanned driving has gradually become popular. Radar has the incomparable superiority of visual technology. It is believed that in the near future, radar SLAM will be irreplaceable in autonomous driving technology important part of Grisetti G. et al. proposed Gmapping, a particle filter-based SLAM algorithm, which maintains maps and pose estimates for each particle [8], and obtained considerable results by improving the proposed distribution and resampling methods, but this method relies on The odometer does not have a loop detection algorithm. Kohlbrecher S et al. proposed HectorSLAM, which is based on the occupancy grid map, overcomes the discontinuity of the occupancy grid map through linear interpolation, uses the Gauss-Newton method to optimize scan matching, and uses multi-resolution maps to avoid getting trapped when optimizing poses. local minimum. This method can fuse information from other sensors (such as inertial sensors) through the method of Extended Kalman Filter (EKF). KartoSLAM is a SLAM method based on graph optimization [9] , which uses window traversal to find the optimal scan match, and uses sparse point adjustment (Spare Pose Adjustment) or g2o for optimization estimation .

*A review of visual SLAM algorithms*, visual SLAM is the focus of current research. Compared with lidar, vision sensor camera is favored by the industry because of its low price. But at the same time, visual sensors usually introduce larger errors, so a good SLAM algorithm is very important. Andrew Davison brought the EKF-SLAM in the lidar into the monocular vision SLAM, and established the MonoSLAM algorithm, which is fast and can model the uncertainty of features [10], but the algorithm only extracts a very small number of features Points count, and the tracking is not very stable. Nister uses efficient corner point feature extraction and matching to implement visual odometry (VO, Visual Odometry), and improves accuracy through bundle adjustment, but cannot eliminate offsets and cannot track relocation after loss. Georg Klein and David Murray proposed the PTAM (Parallel Tracking and Mapping) algorithm, which is a Keyframe-based algorithm that pioneered the separation of tracking and mapping into two The independent thread can be optimized in the drawing thread to improve the real-time performance and accuracy of the algorithm. Richard Newcombe and others proposed DTAM (Dense Tracking and Mapping), which uses denser image information and image features, which greatly improves the robustness and accuracy of the algorithm. The ORB-SLAM proposed by Raul Mur-Artal has improved the map initialization. On the basis of the tracking and mapping threads, the loop detection thread is added, and the loop closure and relocation are performed through the bag of words method, which achieves a better algorithm effect. DSO (Direct Sparse Odometry) was proposed by Jakob Engel. It is a visual odometry calculation method with high precision, good robustness and small drift. The Hong Kong University of Science and Technology Shen Shaojie and others proposed the Vins-mono algorithm in 2018. This algorithm is an algorithm that combines inertial measurement units and monocular vision SLAM, and is applied to drones. The algorithm effect is relatively ideal.

*SLAM algorithm framework*. In terms of framework, the mainstream SLAM algorithm framework mainly includes five steps [11] . (1) Sensor data acquisition and processing, such as laser (sonar) SLAM, visual SLAM (monocular, binocular, depth); (2) matching algorithms, such as feature matching, ICP matching, correlation matching, etc.; (3) Back-end processing, such as

filter algorithm, graph optimization algorithm, etc.; (4) map construction algorithm, such as feature map, occupancy grid map, point cloud map, etc.; (5) loop detection, such as bag of words algorithm, etc. It can be seen that the current research on SLAM algorithms focuses on positioning algorithms and pose estimation, and the research on the accuracy of sensors and map construction is relatively good.

*An overview of robot map construction*. Maps are an important part of robotics. For a fully autonomous robot, a good map model plays a vital role in each step of perception positioning, path planning, and navigation. Up to now, map models mainly include scale map, feature map, topological map and semantic map.

*Scale maps are the most common and intuitive form of maps*. These maps are proportional to the real environment and are the most easily understood map models for humans. For example, park maps and world maps are all maps drawn according to the real environment through scaling. According to the dimension of scale information [12] , it can be divided into planar map (2D map) and stereo map (3D map).

*Grid map*. In robotics, the most common form of scale map is grid map. Because the creation and maintenance of grid maps are very simple, and the corresponding relationship with the real environment is simple and clear, it is often used in map construction based on sensors such as lidar and ultrasonic ranging. However, when the grid size of the grid map is constant, the number of grids will grow quadratically (planar map) or cubically (spatial map) as the area of the described environment grows, which causes the time complexity of updating the map to be the same as that of The space complexity of storing maps grows rapidly, so raster maps are often used in small environments or by approximating the global map with local submaps.

*Feature map*, the feature map does not store the information of the entire environment, but stores some iconic features. This is similar to the way humans remember maps. People tend to describe the environment and their location through distinctive landmarks in the environment (such as typical buildings, intersections, etc.). In robotics, there are various forms of features in feature maps. It can be features extracted from sensor data, such as point, line, surface and other corner features obtained from lidar, or operator features such as SIFT, SURF, ORB obtained from images, or artificial in the environment. Feature signpost added.

Feature maps are suitable for structured environments, where specific feature descriptions are more precise. For unstructured environments, it is difficult to describe with standard features. Another problem with using feature maps is the problem of data association. When maintaining the map, it is necessary to compare the obtained features with those of the current global map, and associate the corresponding feature data, so that a consistent map can be updated and maintained. How to efficiently match and correlate a large amount of feature data becomes another problem.

*Topological map*, topological map is usually used to describe the connection relationship of each part of the map, such as a city subway map. Unlike grid maps and feature maps that correspond to the real environment in scale, topological maps use nodes and connections between nodes to represent the environment. In the topological map, nodes are used to represent a certain location in the environment, and links are used to represent path information between nodes. The map model describes the topological relationship of various places in the environment, making path planning more efficient

*Semantic map*, semantic map not only contains the basic information of the environment, but also contains a high-level understanding of the environment, such as the location of the door, the purpose of the building, etc. Semantic maps play an extremely important role in realizing more intelligent robotic systems. The sequence of different types of maps are shown in Figure 1.

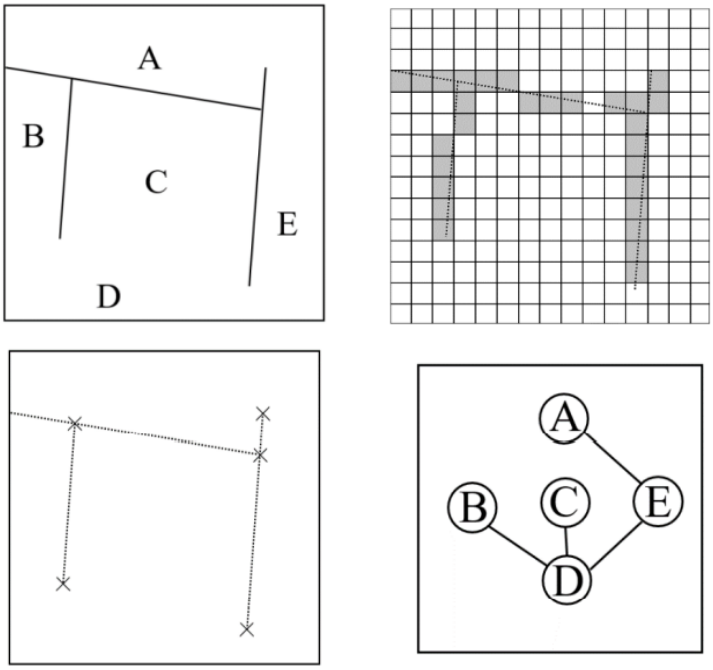



*Figure 1*. Different types of maps

*Lidar point cloud scan matching*, scan matching is a very important algorithm step in map construction. The purpose of scan matching is to match and correlate the data measured by the sensor with the existing map data, estimate the position and attitude of the current system, so as to update the position and attitude of the robot incrementally through the sensor data, and further realize the mapping of the map renew. There are three main types of scan matching methods based on lidar point cloud data.

*Point-based scan matching*. This type of method directly operates on the data points obtained by the sensor, and the most widely used and researched is the ICP algorithm (Iterative Closest Point). In the standard ICP algorithm, scan matching is achieved through continuous iterative optimization to minimize the Euclidean distance between two frames of point clouds. At present, there are various ICP improved algorithms, such as ICL (Iterative Closest Line) algorithm.

*Feature-based scan matching*. This kind of method considers the problem of noise and error in the sensor data measurement, so the point cloud data is processed first, the salient features are extracted, and the scan matching is realized through the correlation between the features. Therefore, good features become an important factor for the accuracy of the algorithm. Features include point features (such as corner points, intersections, etc.), line features (such as Split-and-Merge method), surface features, etc.

*Methods based on mathematical characteristics*, such as NDT method (Normal Distribution Transform) and correlation-based algorithms. The NDT method converts the discrete points in the scan into a probability density, and completes scan matching by optimizing the score of another frame of point cloud data on this probability density distribution. Correlation scan matching is based on the defined correlation description function, and optimizes the score of the point cloud of the second frame to achieve scan matching.

**Problem Analysis of Solving Navigation Problems Based on Laser Range Finder Data.** Navigation corresponds to several basic questions: «Where am I now?», «Where am I going?», «How will I get there?». The first problem corresponds to the positioning of the mobile robot. Based on the robot's own observations and historical poses, the current location is judged. The second and third problems are path planning problems for a given goal and planning a feasible route. Only when the positioning is completed can the starting point be established for subsequent path planning. Therefore, it can be said that the first question determines whether the follow-up

question can be answered more accurately. For a robot system with comprehensive exploration capabilities, positioning, mapping, and navigation are indispensable.

In order to solve the above three problems of the mobile robot, it is necessary to use the sensors carried by it for external perception to realize the feature recognition and path planning of the environment map. At present, the two most common types of sensors on the market are laser-based and vision-based. Among them, laser-based mapping and navigation solutions are widely used in industry and service due to their high precision and good work in dark environments. Industry and military, such as factory inspection robots, express logistics vehicles, etc.

For the positioning and mapping of robots, the current mainstream solutions include graph-based optimization and filtering-based solutions. Among them, the graph-based optimization can construct maps with high consistency, but due to the large resource consumption, it is more suitable for Applies to larger or more complex maps. In contrast, the filtering-based mapping scheme has higher real-time performance and performs better in small-scale indoor environments. As for the path planning algorithm, there are currently three schemes based on sampling, nodes, and models. For static scenarios, the node algorithm can more effectively solve the optimal path.

**Existing methods for solving the local navigation problem.** Navigation methods have been developed to the present, and there are many methods including a series of solutions such as sound, light, electricity, and machinery, but their essence is to use a series of historical moments in the past as the current unknown calculation standard for the robot. Common navigation methods include inertial navigation, magnetic navigation, visual navigation and lidar navigation.

*Inertial navigation*. The development of inertial navigation has gone through several generations, and it is one of the most basic navigation methods. Inertial navigation uses decoders, gyroscopes, etc. to predict the next trajectory of the mobile robot. This navigation method is similar to the Markov transfer process, that is, the current attitude depends only on the previous attitude and is not related to other historical moments. However, this characteristic lead to accumulation of errors. The ball bearing gyroscope in 1852 belongs to the first generation of inertial navigation sensors, because it cannot be associated with electronic signals, so it cannot complete the navigation task of the robot system; the second generation of floating liquid gyroscope was invented in 1942 and is widely used in In terms of military affairs, such as missiles, aircraft or early military robots, their advantages mainly lie in high navigation accuracy; third-generation laser gyroscopes, micro-electric accelerometers, etc. are widely used in the field of robots. At present, the relatively mature strapdown inertial navigation system uses a unified package of gyroscope and accelerometer.

*Magnetic navigation*. Magnetic navigation is mainly linked with the signal acquisition system of the robot. Using the landmark signal, collision rubber signal, infrared signal, etc. collected by the robot, after calculation and processing, the mechanical drive structure of the robot is comprehensively controlled to realize the real-time guidance of the mobile robot. Compared with other navigation schemes, this scheme has significant advantages: easy to control, cost saving, etc. Widely used in warehousing logistics, industrial production and other specific fields. However, the disadvantages exposed in the application of magnetic navigation are also very obvious: the magnetic strip embedded in the ground may be damaged, once the path is changed, the magnetic strip track can only be rebuilt, and dynamic obstacles cannot be avoided.

*Satellite navigation*, satellite navigation must first measure the distance between the current user receiver (mobile phone, radio, etc.) and the satellite, and then integrate the distance information of a series of satellites to calculate the current specific position of the receiver. In order to calculate the longitude and latitude altitude position of the user on the earth, the real-time position and the distance to the receiver of at least 4 satellites are required. The receiver clock can measure the approximate distance between the current user and the satellite. The satellite obtains the current position information through autonomous positioning, and uses the obtained four distance information and the current relative position of the satellite to construct an equation to solve the specific three-dimensional position of the user receiver. The advantage of this navigation

method lies in its universality and the practicability of global positioning, but the disadvantage is that indoor positioning is prone to signal interference, which leads to positioning failure.

*Visual navigation*, visual navigation is similar to humans using binoculars to identify and locate the surrounding environment. A common visual navigation solution generally uses a depth camera to scan the surrounding environment to obtain a series of point cloud information, and then constructs a two-dimensional or three-dimensional environmental map through feature extraction and distance calculation. Its application was first used in 1993. Atiya and Hager proposed a method of extracting static feature points in a dynamic environment to locate the robot in real time. At present, this method is still used in interstellar exploration. Foreign scholars and the National Space Administration of the United States have carried out a large number of experiments based on visual navigation; domestic visual navigation research focuses on extracting a series of images taken during the robot's movement to measure the robot's For example, Wang Yanli et al. have studied the autonomous operation mobile platform in the greenhouse environment. In addition, visual navigation technology has also participated in my country's space exploration plan. For example, Shao Wei et al. have studied the visual navigation algorithm of small celestial body landers.

*LiDAR navigation*. Lidar is the sensor with the highest accuracy and the fastest response in the current robot navigation ranging solution, but it is also the sensor with the highest price. Its navigation principle and ranging effect are shown in Figure 2.

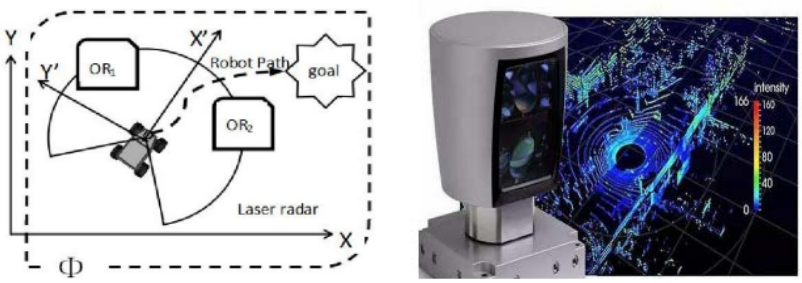



*Figure 2.* Navigation principle and ranging effect

According to the use scenarios of lidar, it can be divided into ground-based lidar, airborne lidar, and space-borne lidar. The accuracy of airborne and space-borne lidar is slightly worse, but its perception range can reach hundreds of meters. , in contrast, ground-based lidar has the highest accuracy, but has a shorter measurement range. Xia Jing et al. measured the measurement accuracy of the airborne lidar linearly with the increase of the height of the carrier. As for the spaceborne lidar, foreign research on it was earlier, and NASA applied it to the lunar probe when it landed. Aspects such as finding a safe landing spot and automatic navigation. At present, lidar is mainly used in the construction of 3D map reconstruction and detection in a closed environment. In specific environments such as rescue and underground exploration, lidar can take advantage of its high accuracy.

**Existing approaches to solving global navigation problems.** The method to solve the global navigation problem is satellite navigation . Satellite navigation first needs to measure the distance between the current user receiver (mobile phone, radio, etc.) Location. In order to calculate the longitude and latitude altitude position of the user on the earth, the real-time position and the distance to the receiver of at least 4 satellites are required. The receiver clock can measure the approximate distance between the current user and the satellite. The satellite obtains the current position information through autonomous positioning, and uses the obtained four distance information and the current relative position of the satellite to construct an equation to solve the specific three-dimensional position of the user receiver. The advantage of this navigation method lies in its universality and the practicability of global positioning, but the disadvantage is that indoor positioning is prone to signal interference, which leads to positioning failure.

**Conclusion.** To sum up, the path of magnetic navigation is relatively fixed and lacks flexibility; inertial navigation has a large drift after long-term use; satellite navigation cannot be

used directly under indoor conditions; visual navigation has too much calculation and is easily affected by light; and LiDAR navigation is less affected by environmental factors and has good measurement characteristics. Therefore, this paper uses LiDAR to realize the navigation system of mobile robots in various environments.

## References

[1] Davison A.J., Reid I.D., Molton N.D., et al. MonoSLAM: Real-Time Single Camera SLAM [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.

[2] Klein G., Murray D. Parallel Tracking and Mapping for Small AR Workspaces [C]. 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, Nara, Japan, 2007: 225-234.

[3] Li Guojun. Research on unmanned SLAM based on lidar and camera fusion [D]. Chengdu: Xihua University, 2021.

[4] Shen Sijie, Tian Xin, Wei Guoliang. A review of SLAM algorithm research based on 2D lidar [J]. Computer Technology and Development, 2022, 32(01): 13-18.

[5] Zhong Min. Research on autonomous navigation of mobile robot based on fusion of vision and lidar [D]. Harbin: Harbin Institute of Technology, 2021.

[6] Wang Shunlu, Li Dangjuan, Fan Yuan. Design of short distance laser ranging sensor system [J]. Foreign Electronic Measurement Technology, 2016, 35(11): 98-102.

[7] Zong Wenpeng, Li Guangyun, Li Minglei, etc. A review of laser scanning matching methods [J]. China Optics, 2018, 11(6): 914-930.

[8] Durrantwhyte H.F. Uncertain geometry in robotics [J]. IEEE Journal on Robotics & Automation, 1988, 4(1):23-31

[9] Yu Li, Tang Junhuai, Chen Juntong. An Improved EKF-SLAM Algorithm Based on Observation Range Constraints[J]. Journal of Zhejiang University of Technology, 2014, 42(5): 566-571.

[10] Guivant J., Nebot E. Optimization of the simultaneous localization and map-building algorithm for real-time implementation[C]. International Conference on Robotics and Automation, 2001, 17(3): 242- 257.

## Авторский вклад

Все авторы внесли равноценный вклад в написании статьи. Заявляем, что данный материал ранее не публиковался и не находится на рассмотрении в других издательствах.

## Благодарность

Работа выполнена за счет гранта Министерства образования Республики Беларусь (договор №10-48/260).

# ПРОБЛЕМЫ НАВИГАЦИИ И УПРАВЛЕНИЯ ДВИЖЕНИЕМ АВТОНОМНОГО МОБИЛЬНОГО РОБОТА: АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ

***Ян Шисинь***
*Аспирант кафедры «Электропривод и автоматизация промышленных установок и технологических комплексов» БНТУ*

***С.А. Павлюковец***
*Доцент кафедры «Электропривод и автоматизация промышленных установок и технологических комплексов» БНТУ, кандидат технических наук, доцент*

***А.А. Вельченко***
*Доцент кафедры «Электропривод и автоматизация промышленных установок и технологических комплексов» БНТУ, кандидат технических наук, доцент*

**Аннотация.** В данной статье кратко представлены характеристики современных мобильных роботов, разбираются текущие проблемы навигации роботов, представлены алгоритмы позиционирования и навигации роботов, а затем представлены карты сетки, топологии и геометрические характеристики, которые описывают модель окружающей среды. Дальнейшие исследования на основе карт необходимы для последующей автономной навигации.

**Ключевые слова:** мобильный робот, управление движением, навигация, SLAM, LiDAR.

УДК 004.021:004.75

# ОЦЕНКА КАЧЕСТВА ПЕРЕДАЧИ ИНФОРМАЦИИ В СИСТЕМЕ ДИСПЕТЧЕРИЗАЦИИ НА БАЗЕ MQTT-АРХИТЕКТУРЫ

***Е.И. Лещевич***
*аспирант кафедры электронной техники и технологии БГУИР*
e.leshchevich@bsuir.by

***С.И. Сидоров***
*студент кафедры электронной техники и технологии БГУИР*
stasevichsid@gmail.com

***И.И. Ревинская***
*ассистент кафедры электронной техники и технологии БГУИР*
irewinskaya.bsuir@gmail.com

***С.К. Дик***
*заведующий кафедрой инженерной и компьютерной графики БГУИР, кандидат физико-математических наук, доцент*

***Ситник Г.Д.***
*доцент кафедры общей врачебной практики с курсом гериатрии и паллиативной медицины БГМУ, кандидат медицинских наук, доцент, врач высшей категории по неврологии*

***Е.И. Лещевич***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Аспирант кафедры электронной техники и технологии. Область научного интереса – влияние инфразвука на биологические ткани.*

***С.И. Сидоров***
*Студент кафедры электронной техники и технологии Белорусского государственного университета информатики и радиоэлектроники. Область научного интереса – компьютерная инженерия, дизайн, программирование.*

***И.И. Ревинская***
*Окончила Белорусский государственный университет информатики и радиоэлектроники. Ассистент кафедры электронной техники и технологии. Область научного интереса – медицинская электроника и обработка медицинских сигналов.*

***С.К. Дик***
*Окончил Минский радиотехнический институт по специальности «Радиотехника», руководит научными исследованиями в области лазерной медицины.*

***Г.Д. Ситник***

*Кандидат медицинских наук, доцент, врач высшей категории по неврологии. Область научного интереса – лечение больных с неврологическими проявлениями поясничного остеохондроза.*

**Аннотация.** Имитационное моделирование играет важную роль в обучении и тестировании систем искусственного интеллекта (ИИ). Оно позволяет создавать виртуальные среды, в которых ИИ может безопасно развивать свои навыки без необходимости проведения дорогостоящих и потенциально опасных экспериментов в реальном мире. В данной работе рассматриваются основные принципы имитационного моделирования, его преимущества перед реальными средами, а также способы интеграции с системами ИИ. Особое внимание уделяется использованию имитационных моделей в генерации данных, обучении с подкреплением и внедрении обученных ИИ-моделей в виртуальные среды. Также рассматриваются современные решения, такие как Project Bonsai от Microsoft и H2O Driverless AI, позволяющие упростить процесс обучения и интеграции ИИ с моделями.

**Ключевые слова:** искусственный интеллект, имитационное моделирование, машинное обучение, обучение с подкреплением, генерация данных, виртуальная среда, Project Bonsai, H2O Driverless AI.

**Введение.** Одной из наиболее перспективных и актуальных сфер развития на данный момент является искусственный интеллект. Это объясняется высокой эффективностью его работы в различных направлениях: медицина, промышленность, сельское хозяйство, машиностроение и многие другие. Все большее развитие и совершенствование искусственного интеллекта позволяют ему все успешнее решать задачи распознавания лиц, создания уникальных текстов, модернизации изображений и так далее.

Для наиболее продуктивного обучения искусственный интеллект должен обучаться и изменять поведение в динамической среде. Для этого ему требуется виртуальная площадка, которую может обеспечить имитационное моделирование. Именно оно может позволить безопасно и эффективно обучать и тестировать искусственный интеллект.

**Терминология искусственного интеллекта.** Искусственный Интеллект — это комплекс технологических решений, позволяющий имитировать когнитивные функции человека и получать при выполнении конкретных практически значимых задач обработки данных результаты, сопоставимые, как минимум, с результатами интеллектуальной деятельности человека [1].

Понятие искусственного интеллекта тесно связано с машинным обучением — набором методов для разработки способных к обучению алгоритмов с целью повышения качества работы этих алгоритмов.

Машинное обучение делится на множество видов, но среди основных можно выделить основные [2].

Обучение с учителем, при котором модель обучается на примерах с известным ответом, чтобы произвести настройку параметров. Впоследствии она должна работать на примерах, на которых не обучалась, и давать точный ответ.

Обучение без учителя, при котором обучение производится на неразмеченных данных с целью обнаружения скрытых взаимосвязей между ними.

В обучении с подкреплением модель осуществляет действия в некоторой среде. Все действия несут в себе получение различной награды. Целью такого обучения является получение наибольшего вознаграждения. В результате модель должна выбирать действия, исходя из состояния среды.

**Имитационные модели.** Чтобы программа могла обучаться, а также работать с данными по итогу обучения, ей нужна среда.

Реальная среда является наиболее подходящей для обучения и тестирования, ввиду отсутствия каких-либо неточностей и упрощений, в отличие от модели. Однако их использование несет в себе некоторые недостатки. Одним из них является серьезные денежные затраты. Кроме этого, зачастую эксперименты над реальными объектами могут быть сложны в реализации или опасны для здоровья людей.

Имитационные модели обладают рядом преимуществ. Среди них экономия денежных средств, так как виртуальные эксперименты дешевле. В построенных моделях гораздо проще собрать статистику и провести анализ. Возможность управлять временем эксперимента также является важным достоинством. Благодаря этому можно ускорять или замедлять процесс, что позволяет серьезно экономить время или более тщательно изучать объект исследования.

**Имитационное моделирование в разработке искусственного интеллекта.** Имитационное моделирование может удачно применяться при разработке искусственного интеллекта. Стоит выделить три эффективных способа [3].

Имитационная модель может быть использована как генератор произвольного количества данных. Это достаточно актуальный способ использования имитационного моделирования, так как при обучении искусственного интеллекта часто возникает проблема нехватки данных.

Имитационная модель может стать средой для обучения и тестирования искусственного интеллекта. В этом варианте моделирование чередуется с циклом обучения. Агент анализирует состояние модели, изменяет её, и в зависимости от правильности действия агента происходит его поощрение или наказание, что в свою очередь меняет агента. Этот способ применения имитационного моделирования является весьма востребованным при различных разработках.

Одним из наиболее интересных в данной сфере является Project Bonsai от Microsoft [4]. Благодаря совместной работе компаний AnyLogic и Microsoft появилась возможность подключать созданные модели к платформе Bonsai. Это позволяет специалистам без знаний и опыта в области искусственного интеллекта использовать его для оптимизации и автоматизации своих задач.

Для простоты использования платформы в систему AnyLogic был добавлен новый вид эксперимента «RLExperiment», который используется для обучения с подкреплением. При его работе созданная пользователем модель становится обучающей средой для ИИ-агентов в Project Bonsai. Сам эксперимент уже содержит базовые элементы, позволяющие легко осуществить взаимосвязь имитационной модели и платформы Bonsai.

Ещё одним интересным и широко применимым является продукт сотрудничества компаний AnyLogic и H2O.ai [5], который также позволяет совмещать возможности имитационного моделирования и машинного обучения. H2O Driverless AI позволяет встраивать уже обученную ИИ-модель как автономный компонент модели пользователя.

Для работы с данным продуктом нужно скачать обученную модель (MOJO, Model Object Optimized) в виде файла, а затем использовать как функцию, которая по входным данным созданной модели будет выдавать выходные данные.

**Заключение.** В результате изучения совместного использования искусственного интеллекта и имитационного моделирования можно сделать вывод о том, что их совместное использование позволяет решать задачи эффективнее и расширять круг использования ИИ. Однако, существуют некоторые трудности.

Реальные задачи могут быть весьма трудоемкими и сложными вычислительно, что увеличивает время моделирования между циклами обучения искусственного интеллекта.

При условии использования моделирования результат действий обученного искусственного интеллекта может быть неточным, так как любая модель – это упрощенная в какой-либо степени реальность.

А также стоит отметить, что специалисты по искусственному интеллекту могут слабо разбираться в имитационном моделировании. Это усложняет их совместное использование. Но стоит отметить, что благодаря существующим современным решениям по уже готовым ИИ-агентам и средствам, позволяющим обычные модели превратить в тренажеры для обучения, применение имитационного моделирования и искусственного интеллекта становится все более эффективно.

**Список литературы**

[1] Системы искусственного интеллекта. Классификация систем искусственного интеллекта. — М.: Стандартинформ, 2021. — 16 с.

[2] Бурков А. Машинное обучение без лишних слов. — СПб.: Питер, 2020. — 192 с.

[3] Борщев, А. В. Имитационные модели как виртуальная среда для обучения и тестирования искусственного интеллекта для бизнес-приложений / А. В. Борщев, А. Махдави // Девятая всероссийская научно-практическая конференция по имитационному моделированию и его применению в науке и промышленности : Труды конференции, Екатеринбург, 16–18 октября 2019 года. – Екатеринбург: Издательство Уральского государственного педагогического университета, 2019. – С. 20-29.

[4] Обучение ИИ-агентов с помощью Progect Bonsai от Microsoft. https://www.anylogic.ru/features/artificial-intelligence/microsoft-bonsai/

[5] Точные прогнозы с имитационным моделированием и машинным обучением от H2O.ai. https://www.anylogic.ru/features/artificial-intelligence/h2o-ai/

**Авторский вклад**

Авторы внесли равный вклад в написание статьи.

# ASSESSMENT OF THE QUALITY OF INFORMATION TRANSMISSION IN A DISPATCHING SYSTEM BASED ON MQTT ARCHITECTURE

***Y.I. Liashchevich***
*Postgraduate student of the BSUIR, assistant of the Department of Electronic Technique and Technology of BSUIR*

***S.I. Sidarau***
*student of the Department of Electronic Technique and Technology of BSUIR*

***I.I. Revinskaya***
*assistent at the Department of Electronic Technique and Technology*

***S.K.Dzik***
*Chair of the department of Engineering and Computer Graphics, PhD, Associate Professor*

***G.D.Sitnik***
*PhD, Associate Professor at the Department of Belarusian State Medical University, Doctor of the highest category in neurology*

**Abstract.** Simulation modeling plays a crucial role in training and testing artificial intelligence (AI) systems. It enables the creation of virtual environments where AI can safely develop its skills without the need for costly and potentially dangerous real-world experiments. This paper explores the fundamental principles of simulation modeling, its advantages over real environments, and various integration methods with AI systems. Special attention is given to the use of simulation models for data generation, reinforcement learning, and the implementation of trained AI models in virtual environments. Additionally, modern solutions such as Microsoft's Project Bonsai and H2O Driverless AI are discussed, which facilitate the training and integration of AI with simulation models.

**Keywords**: artificial intelligence, simulation modeling, machine learning, reinforcement learning, data generation, virtual environment, Project Bonsai, H2O Driverless AI.

УДК 612.087.1

# АВТОМАТИЗИРОВАННАЯ СИСТЕМА ДЛЯ ПРОВЕДЕНИЯ ИССЛЕДОВАНИЙ МЕДИЦИНСКИХ ИМПЛАНТИРУЕМЫХ ПРИБОРОВ

***А.Н. Осипов***
*Заведующий лабораторией НИЧ БГУИР, кандидат технических наук, доцент*
osipov@bsuir.by

***С.П. Урбанович***
*Ведущий инженер Отраслевой лаборатории ЦМИ НИЧ БГУИР*
urbanovich@bsuir.by

***М.П. Батура***
*Заведующий лабораторией НИЧ БГУИР, доктор технических наук, профессор*
bmpbel@bsuir.by

***А.А. Ивановский***
*Студент БГУИР*
a.ivanovskiy03@mail.ru

***И.О. Хазановский***
*Аспирант БГУИР*
khaz_igor@bsuir.by

***А.В. Короткевич***
*Доцент кафедры микроэлектроники БГУИР, кандидат технических наук, доцент*
korotkevich@bsuir.by

***А.Н. Осипов***

*Окончил Минский радиотехнический институт. Область научных интересов – биомедицинская инженерия.*

***С.П. Урбанович***

*Окончил Белорусский государственный университет информатики и радиоэлектроники. Специалист в области синтеза и цифровой обработки сигналов*

***М.П. Батура***

*Окончил Минский радиотехнический институт. Область научных интересов - системный анализ, управление и обработка информации в технических системах*

***А.А. Ивановский***

*Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем.*

***И.О. Хазановский***

*Окончил Военную Академию Республики Беларусь. Область научных интересов связана с биомедицинской инженерией.*

***А.В. Короткевич***

*Окончил Минский радиотехнический институт. Область научных интересов - технология формирования изделий микро-,нано- и криоэлектроники с использованием анодных оксидов вентильных металлов.*

**Аннотация.** Рассмотрены подходы к созданию автоматизированной системы для исследований медицинских имплантируемых приборов. Описаны архитектура программного обеспечения, включающего взаимодействие с аппаратной платформой по протоколу *Modbus*, управление датчиками и исполнительными устройствами, а также способ хранения и обработки данных с использованием СУБД (Система управления базами данных). Показано, что модульная структура программно-аппаратного комплекса и использование современных инструментов позволяют адаптировать систему под различные сценарии испытаний, обеспечить хранение и обработку большого количества данных исследований.

**Ключевые слова:** автоматизированная система, медицинские имплантируемые приборы, ускоренные испытания, протокол *Modbus, Raspberry Pi CM4, PyQt5*, управление параметрами эксперимента, хранение и обработка данных.

**Введение.** Устройства, имплантируемые в тело человека, играют важную роль как в поддержании жизнедеятельности, так и в диагностике различных физиологических параметров, таких как уровень глюкозы, давление, температура, активность сердца и других. Беспроводные имплантируемые датчики – это сложные биомедицинские микроэлектромеханические системы (БиоМЭМС), включающие биомедицинские диагностические и лечебные устройства с интегрированными микроэлектронными и микромеханическими компонентами. К беспроводным имплантируемым приборам предъявляются жесткие требования в части минимизации габаритов, обеспечения длительной работоспособности, биосовместимости и т.д. Технологии создания таких датчиков – сложный междисциплинарный процесс, для решения которой необходимо комплексно решить ряд научных и инженерных задач в области микроэлектроники, радиотехнологий, материаловедения и медицины. Наиболее оптимальным для проведения исследований является использование специального технологического оборудования, имитирующего функционирование имплантируемых устройств в среде, которая по своим характеристикам (электродинамическим, механическим, и т.д.) аналогична реальным различным частям и/или тканям человеческого тела, позволяющего сократить сроки проведения лабораторных испытаний и доклинических исследований. В связи с этим в данной статье рассматриваются вопросы создания автоматизированной системы для исследований медицинских имплантируемых приборов, позволяющей адаптировать систему под различные сценарии испытаний и обеспечить хранение и обработку большого количества данных.

**Структурная схема системы и организация обмена информацией.** Структурная схема автоматизированной системы для проведения исследований медицинских имплантируемых приборов включает блок управления и пять тестовых ячеек, объединённых общей магистральной линией связи (рисунок 1). Блок управления выступает в роли «ведущего» (*master*-устройства), тогда как каждая тестовая ячейка функционирует как «ведомое» (*slave*-устройство), обмениваясь данными по стандартному протоколу *Modbus*. В качестве аппаратной основы для передачи данных выбран двухпроводный дифференциальный интерфейс *RS-485* (*TIA/EIA-485*), обеспечивающий полудуплексную связь и высокую помехоустойчивость в условиях лабораторных и производственных испытаний. Используемая топология позволяет минимизировать количество проводных линий и упростить монтаж.: для всех устройств используются две сигнальные жилы (*A* и *B*), к которым параллельно подключаются приёмопередатчики интерфейса *RS-485*. Поддержка стандартных скоростей и возможность применения внешних приёмопередатчиков, совместимых с популярными микроконтроллерами и микрокомпьютерами (*Arduino, ESP32, Raspberry Pi CM4* и другие), делают решение масштабируемым и легко адаптируемым к изменениям состава оборудования.

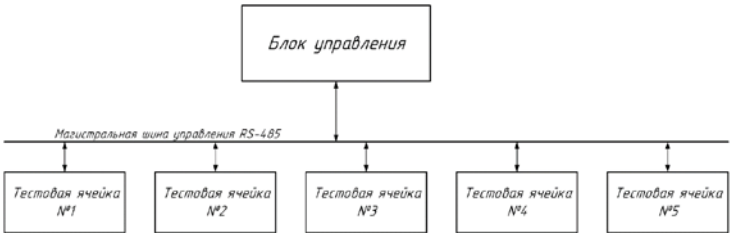



*Рисунок 1.* Структурная схема автоматизированной система для проведения исследований медицинских имплантируемых приборов

Обмен данными регулируется протоколом *Modbus* (*Modbus RTU*), в рамках которого блок управления последовательно обращается к каждой ячейке по её уникальному адресу (от 1 до 247). Формат посылки предусматривает передачу адреса, кода выполняемой функции и набора регистров с данными, а также контрольной суммы (*CRC*). Стандартные функции протокола *Modbus* (чтение и запись регистров, чтение дискретных входов/выходов, передача диагностической информации) обеспечивают универсальность системы.

Каждая тестовая ячейка содержит микроконтроллер, обрабатывающий запросы от блока управления и возвращающий ответные сообщения. Внутри тестовой ячейки размещаются исполнительные и измерительные устройства, среди которых: датчики температуры, насосы и клапаны для перекачки и регулирования состава буферного раствора, УФ-модули для облучения исследуемых материалов, нагреватели с системами *PID*-регулирования, дополнительные датчики. Все устройства внутри ячейки управляются локальным микроконтроллером, который, в свою очередь, обменивается данными с блоком управления через регистры *Modbus*.

Блок управления, представленный в виде микрокомпьютера (*Raspberry Pi CM4*) с установленным программным обеспечением, обеспечивает визуальный интерфейс пользователя, хранение данных в базе (*MariaDB*) и взаимодействие по сети *Ethernet/Wi-Fi* при необходимости. Программный модуль, реализующий протокол Modbus, опрашивает каждую ячейку по очереди, формирует запросы на чтение или запись регистров и обрабатывает ответы. Вся собранная информация сохраняется в базу данных и отображается оператору в реальном времени.

**Программное обеспечение блока управления.** Программное обеспечение (ПО) блока управления разработано в соответствии с требованиями ГОСТ IEC 62304-2022 «Изделия медицинские. Программное обеспечение. Процессы жизненного цикла» и обеспечивает централизованное управление пятью тестовыми ячейками через протокол *Modbus*, используя микрокомпьютер *Raspberry Pi CM4* в качестве аппаратной платформы. Применение *Raspberry Pi* даёт возможность функционировать под управлением операционной системы семейства Linux, а также использовать язык программирования *Python* с фреймворком *PyQt5*. Хранение данных организовано в базе *MariaDB*, что способствует надёжной и удобной интеграции с компонентами программного обеспечения. ПО выполняет функции контроля и диагностики, выявляя потенциальные отказы, неисправности и аварийные ситуации в блоке управления и тестовых ячейках. Способность системы обнаруживать сбои и формировать сигналы оповещения минимизирует риск нештатных ситуаций и обеспечивает безопасность работы оператора со стендом. ПО организовано по модульному принципу, что способствует расширяемости и прозрачному разделению функционала. Основные модули: *Core* (ядро программы), *Database* (работа с базой данных), *GUI* (графический интерфейс пользователя) и *Settings* (настройки). Диаграмма, отображающая модули ПО блока управления и их взаимодействие отображена на рисунке 2.

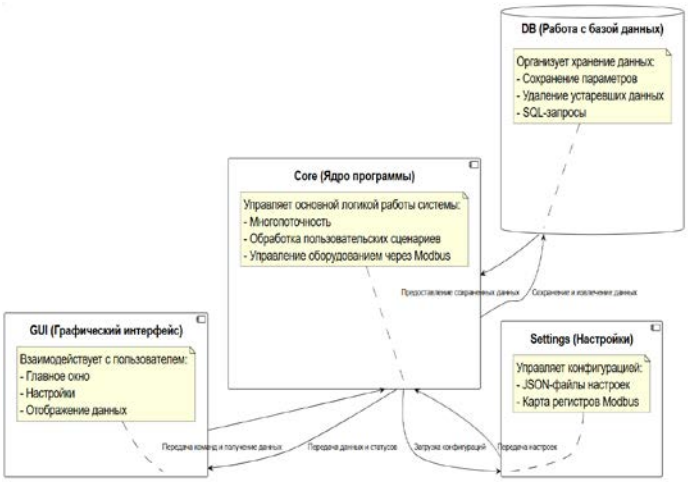



*Рисунок 2.* Диаграмма, отображающая модули ПО блока управления и их взаимодействие

Модуль *Core* обеспечивает многопоточную архитектуру и выполняет главную логику управления стендом. Он обрабатывает запросы от пользовательского интерфейса и взаимодействует с оборудованием посредством протокола *Modbus*. Данный модуль осуществляет координацию потоков данных, поддерживает синхронизацию процессов и распределяет задания между различными подсистемами. Параллельная обработка входящей/выходящей информации и результатов опроса датчиков даёт возможность оперативно реагировать на изменения параметров исследований, повышая общую производительность системы и уменьшая задержки.

Модуль *Database* организует хранение и доступ к данным, используя СУБД *MariaDB,* совместимую с архитектурой *ARM*. Структура базы ориентирована на высокую надёжность и удобство масштабирования. Модуль изолирован от прочих компонентов ПО, что позволяет вносить изменения в логику хранения или мигрировать на иную СУБД без нарушения функционирования других частей приложения. *Database* ведёт журнал изменений конфигурации, сохраняет результаты данных исследований, управляет удалением устаревших записей и даёт возможность оперативного доступа к собранным данным для анализа.

Модуль графического интерфейса пользователя (*GUI*) отвечает за взаимодействие с оператором, формируя удобное рабочее пространство. Он содержит набор окон и виджетов, позволяющих:

– управлять параметрами системой для проведения исследований;

– визуализировать данные в реальном времени в виде таблицы или графика;

– задавать сценарии автоматизированных исследований (последовательность команд и временные интервалы их исполнения).

Модуль *GUI* разработан на базе *PyQt5* и работает в многопоточном режиме, получая информацию от *Core* и отображая её в понятном для пользователя виде. Интерфейс выполнен с учётом требований ГОСТ ISO 10993-1–2021 относительно корректной настройки параметров УФ-воздействия на объекты, помещаемые в фантомы.

Модуль *Settings* отвечает за конфигурационные файлы в формате *JSON (JavaScript Object Notation)*, где зафиксированы основные параметры работы ячеек: значения регистров *Modbus*, интервалы опроса, настройки нагревателей, конфигурация насосов и УФ-модулей. Данный подход с использованием *JSON* обеспечивает удобочитаемость настроек и гибкость при обновлениях: внесение изменений не требует правки исходного кода ПО. Модуль загружает необходимые конфигурации при запуске системы, а также предоставляет возможность их оперативного обновления на этапе исследовательских работ без перезапуска ПО. Информация из *Settings* передаётся в *Core*, а при необходимости сохраняется в *Database*.

Все модули объединены в единую систему обмена данными, где *Core* выступает ключевым связующим звеном. *Core* взаимодействует с *Database* для записи результатов и считывания архива, а также получает настройки через *Settings*. Данные, необходимые пользователю, отображаются в *GUI*, а команды, заданные оператором или сценариями исследований, передаются обратно в *Core*, который транслирует их оборудованию по *Modbus*. Данная схема исключает дублирование логики и упрощает обслуживание.

ПО предназначено для работы в круглосуточном режиме, что предусматривает отработку исключительных ситуаций: внезапные сбои сетевого взаимодействия, непредвиденные перепады напряжения, отключение отдельных тестовых ячеек. Блок управления формирует аварийные сообщения при обнаружении нештатных событий, регистрирует их в базе и выводит на экран оператора. Модульная структура и соответствие современным стандартам делают систему масштабируемой, позволяя расширять функционал или подключать новые модули без изменения базовых принципов работы.

**Сценарии проведения исследований.** Сценарии проведения исследований представляют собой детально прописанную последовательность действий, связанных с изменением параметров в одном или нескольких стендах, чтобы обеспечить максимально точное воспроизведение условий испытаний для медицинских имплантируемых приборов. Его подготовка начинается во внешнем приложении, в котором оператор указывает целевые значения для различных параметров (скорость вращения рабочего насоса, температура нагревательного модуля, длительность или частота импульсов ультрафиолетового облучения), а также время, через которое каждое действие должно быть выполнено относительно начала эксперимента.

После завершения подготовки сценарий сохраняется в специальный файл, который загружается в основное программное обеспечение блока управления. Программа анализирует структуру сценария, определяет его общую продолжительность и отображает основные сведения на экране оператора. В частности, пользователь может видеть, сколько времени займёт выполнение всех действий, на какие стенды направлены соответствующие команды и как именно изменятся параметры. При необходимости оператор запускает сценарий, и система пошагово отрабатывает запланированные изменения. Если для некоторых экспериментов нужно вводить разные условия в разных стендах, сценарий допускает включение нескольких параллельных ветвей. Например, в одной ячейке автоматически поднимается температура, в другой поддерживается стабильный нагрев и регулируется интенсивность ультрафиолетового облучения, а в третьей меняется скорость подачи буферной жидкости. В процессе выполнения каждый шаг сценария сопровождается записью фактических значений в базу данных: в ней фиксируются временные метки, состояние насосов и клапанов, текущая температура или мощность УФ-излучения. Когда очередной шаг срабатывает в момент, заданный пользователем, программа проверяет корректность передачи команд, отслеживает их выполнение, а также реагирует на возможные сбои. При возникновении нештатных обстоятельств (выхода за пределы заданной температуры или выявления неисправности в одном из узлов) оператор имеет возможность приостановить ход сценария, устранить проблемы и продолжить работу с момента прерывания.

При многоячейковой конфигурации сценарии могут варьировать условия испытаний для каждого стенда отдельно, что особенно востребовано при сравнительных тестах различных типов имплантируемых приборов. Система способна параллельно обрабатывать несколько потоков команд, не ограничиваясь единым набором параметров, что обеспечивает высокую гибкость и масштабируемость исследований. Пользователь, запустив сценарий, видит ход выполнения в режиме реального времени и может проанализировать, какие действия уже завершены и какие ещё предстоит выполнить. В случае необходимости изменения плана существует возможность откорректировать значения или интервалы до завершения экспериментального цикла.

Сценарии проведения исследований предоставляют операторам возможность комплексно и гибко управлять параметрами испытаний, формировать многоуровневые планы автоматического изменения среды в каждой ячейке и добиваться высокой точности в воспроизведении экспериментальных условий. Параллельное выполнение команд по нескольким стендам сокращает время для интегральных исследований, а система мониторинга и логирования даёт возможность приостанавливать и возобновлять сценарий при сохранении всех ранее собранных данных. Подобная организация экспериментов открывает широкие перспективы для исследователей, разрабатывающих и тестирующих медицинские имплантируемые приборы различной сложности.

**Заключение.** Представлена автоматизированная система для проведения исследований медицинских имплантируемых приборов. Система обеспечивает централизованное управление пяти и более тестовыми ячейками в соответствии с заданным оператором сценарием проведения исследований. Описаны архитектура программного обеспечения, включающего взаимодействие с аппаратной платформой по протоколу *Modbus*, управление датчиками и исполнительными устройствами, а также способ хранения и обработки данных с использованием СУБД. Модульная структура программного обеспечения и использование современных инструментов позволяют адаптировать систему под различные сценарии испытаний имплантируемых датчиков различных типов, обеспечить хранение и обработку большого количества данных исследований. Использование автоматизированной системы на практике позволит сократить время проведения исследований.

## Список литературы

[1] A miniature implantable RF-wireless active glaucoma intraocular pressure monitor : научная статья / E. Y. Chow [и др.] – IEEE Trans. Biomed : Circuits Syst, 2010. – 350 с.

[2] Программно-аппаратный комплекс ускоренных испытаний имплантированных медицинских датчиков : научная статья / А.Н. Осипов [и др.] – Минск : ResearchGate, 2025. – 5 с.

[3] WaveShare [Электронный ресурс]. – Режим доступа : https://www.waveshare.com/wiki/CM4-DISPBASE-7A-BOX.

## Авторский вклад

Авторы внесли равный вклад в написание статьи

# AUTOMATED SYSTEM FOR CONDUCTING RESEARCH ON MEDICAL IMPLANTABLE DEVICES

***A.N. Osipov***
*Head of the Research Laboratory, BSUIR, PhD of Technical Sciences, Associate Professor*

***S.P. Urbanovich***
*The leading engineer of the Research Laboratory, BSUIR*

***M.P. Batura***
*Head of the Research Laboratory, BSUIR, Doctor of Technical Sciences, Professor*

***A.A. Ivanovsky***
*Student, BSUIR*

***I.O. Khazanovsky***
*Postgraduate student, BSUIR*

***A.V. Korotkevich***
*Associate Professor of the Department of Micronanoelectronics, BSUIR, PhD of Technical Sciences, Associate Professor*

**Abstract.** Approaches to the creation of an automated system for the research of medical implantable devices are considered. The architecture of the software is described, including interaction with the hardware platform via the Modbus protocol, control of sensors and actuators, as well as a method for storing and processing data using a DBMS (Database Management System). It is shown that the modular structure of the software and hardware complex and the use of modern tools make it possible to adapt the system to various test scenarios, to ensure the storage and processing of a large amount of research data.

**Keywords**: automated system, medical implantable devices, accelerated testing, Modbus protocol, Raspberry Pi CM4, PyQt5, electromagnetic compatibility, control of experimental parameters.

УДК 658.8:004.9

# ОПРЕДЕЛЕНИЕ ЦЕЛЕВЫХ ФУНКЦИЙ И ИХ ОГРАНИЧЕНИЙ В ЗАДАЧАХ ОПТИМИЗАЦИИ ПРИ ГЕНЕРАТИВНОМ ПРОЕКТИРОВАНИИ

***В.Ф. Алексеев***
*Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент*
alexvikt.minsk@gmail.com

***Д.В. Лихачевский***
*Декан факультета компьютерного проектирования БГУИР, кандидат технических наук, доцент*
likhachevskyd@bsuir.by

***Г.А. Пискун***
*Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент*
piskunbsuir@gmail.com

***В.Ф. Алексеев***
*Окончил Минский радиотехнический институт. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.*

***Д.В. Лихаческий***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием проблем радиочастотной идентификации объектов, организацией учебного и научно-исследовательского процессов в техническом университете.*

***Г.А. Пискун***
*Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.*

**Аннотация.** Рассматриваются ключевые аспекты определения целевых функций и их ограничений в контексте задач оптимизации, связанных с генеративным проектированием. В рамках статьи анализируются методы формулирования целевых функций, их влияние на результаты проектирования, а также способы определения ограничений, которые учитывают различные технические, экономические и экологические факторы. Особое внимание уделяется примерам из практики, иллюстрирующим влияние корректного определения целевых функций и ограничений на эффективность проектирования.

Статья будет представлять интерес для специалистов в области инженерии, дизайна и оптимизации, а также на исследователей, интересующихся современными методами проектирования.

**Ключевые слова:** генеративное проектирование, целевая функция, ограничения, оптимизация проектных решений

**Введение.** Одним из наиболее перспективных путей в области разработки электронных систем является генеративное проектирование, которое позволяет использовать алгоритмические подходы для автоматизации процесса создания проектных решений. Основная идея генеративного проектирования заключается в использовании алгоритмов, которые способны генерировать множество вариантов проектирования на основе заданных параметров и ограничений. Этот процесс позволяет значительно ускорить разработку проектов, обеспечить высокую степень инновации. Особое внимание следует уделить определению целевых функций и их ограничений, так как они играют ключевую

роль в процессе проектирования. Само решение целевой функции ориентировано на принятие решений с учётом ограничений [1].

В общем случае целевая функция зависит от целого ряда аргументов, которые в той или иной степени влияют на процессе оптимизации. Целевые функции, как правило, представляют собой количественные показатели, которые необходимо максимизировать или минимизировать. Это могут быть такие параметры, как масса конструкции, прочность, устойчивость к нагрузкам и др. Принимая во внимание множество факторов, влияющих на конечный продукт, необходимо обеспечить баланс между техническими, технологическими, производственными требованиями и экономическими показателями [2].

Ограничения в задачах оптимизации выполняют важную роль, так как они определяют допустимые пределы для проектируемых решений. Они могут быть связаны с физическими характеристиками материалов, геометрическими ограничениями, соблюдением нормативных требований и другими параметрами. Правильная формулировка и учёт этих ограничений помогают избежать нецелесообразных конструктивных решений и гарантируют выполнение критических требований к проекту [3].

Актуальность исследований подчеркивается работами [4–7], которые поднимают вопросы о том, как генеративные методы могут быть интегрированы в существующие системы управления проектами. В свою очередь, зарубежные работы, например, исследования [4], показывают значимость применения машинного обучения и искусственного интеллекта в контексте генеративной политики, что открывает новые горизонты для оптимизации и инноваций.

Цель настоящей статьи – рассмотреть существующие методы и подходы по определению целевых функций и ограничений в задачах оптимизации.

**Определение целевых функций и ограничений в задачах оптимизации.** Целевая функция в задаче оптимизации представляет собой математическое выражение, которое необходимо максимизировать или минимизировать. Например, в области генеративного проектирования целевой функцией может быть масса конструкции, которую следует минимизировать, или прочность материала, которую необходимо максимизировать. Определение целевой функции является ключевым моментом, так как именно от нее зависит, насколько эффективно будет найдено оптимальное решение [1–3].

Рассмотрим задачу на примере проектирования печатной платы. В задачах оптимизации, используемых в генеративном проектировании электронных средств, ключевым этапом является четкое определение целевых функций и ограничений. Целевая функция (objective function) представляет собой математическое выражение, которое необходимо минимизировать или максимизировать для достижения оптимального решения. Ограничения (constraints) – это условия, которые должны быть соблюдены в процессе оптимизации и определяют допустимую область решений.

Выбор целевых функций и ограничений напрямую влияет на конечный результат генеративного проектирования. Некорректно сформулированные целевые функции могут привести к получению непрактичных или неэффективных решений, а игнорирование важных ограничений – к созданию нереализуемых проектов.

Целевые функции в генеративном проектировании электронных средств могут включать:

– минимизацию занимаемой площади печатной платы (PCB) – стремление к уменьшению габаритов устройства, что особенно важно для портативных и мобильных устройств. Математически это может быть выражено как:

$$\min f(x) = \text{Area}(PCB(x))$$

где Area(PCB(x)) – функция, вычисляющая площадь печатной платы в зависимости от параметров проектирования x (например, расположение компонентов, трассировка проводников).

Пример: при размещении компонентов на печатной плате целевая функция может заключаться в минимизации площади платы при заданных ограничениях на расстояние между компонентами и требованиях к трассировке.

– минимизация длины проводников (trace length) – уменьшение длины проводников снижает паразитные эффекты (индуктивность, емкость) и улучшает целостность сигнала. Целевая функция может быть определена как:

$$\min f(x) = \Sigma \text{ Length}(\text{trace_i}(x)) \quad \text{for } i = 1 \text{ to } N$$

где Length(trace_i(x)) – функция, вычисляющая длину i-го проводника в зависимости от параметров трассировки x, N – общее количество проводников.

Пример: оптимизация трассировки печатной платы с целью минимизации общей длины проводников, особенно для высокоскоростных сигналов.

– максимизация рассеиваемой мощности (heat dissipation) – обеспечение эффективного отвода тепла от компонентов, особенно для мощных устройств. Целевая функция в этом случае выглядит как:

$$\max f(x) = \text{HeatDissipation}(x)$$

где HeatDissipation(x) – функция, вычисляющая рассеиваемую мощность в зависимости от параметров конструкции x (например, расположение радиаторов, вентиляционных отверстий).

Пример: оптимизация расположения компонентов и радиаторов для обеспечения максимального отвода тепла от микропроцессора.

Минимизация стоимости (cost) – снижение затрат на материалы, производство и сборку устройства. В этом случае целевая функция выглядит так:

$$\min f(x) = \Sigma \text{ Cost}(\text{component_i}(x)) \quad \text{for } i = 1 \text{ to } M$$

где Cost(component_i(x)) – функция, вычисляющая стоимость i-го компонента в зависимости от выбора компонентов x, M – общее количество компонентов.

Пример: выбор оптимальных компонентов и технологий производства для минимизации общей стоимости печатной платы.

Ограничения в генеративном проектировании электронных средств могут быть представлены в виде:

– геометрические ограничения – ограничения на размеры компонентов, расстояние между ними, минимальную ширину проводников, зазоры.

Пример: минимальное расстояние между двумя BGA компонентами должно быть не менее 2 мм.

$$\text{Distance}(\text{Component_A}, \text{Component_B}) >= \text{MinDistance}$$

– электрические ограничения – ограничения на импеданс проводников, максимальный ток, допустимое падение напряжения, электромагнитную совместимость (EMC).

Пример: импеданс проводника должен быть в пределах 50 ± 10 Ом.

$$Z_\text{trace}(x) \in [Z_\text{min}, Z_\text{max}]$$

– тепловые ограничения – ограничения на максимальную температуру компонентов, градиент температуры.

Пример: температура микропроцессора не должна превышать 85 °C.

Temperature(Microprocessor) <= MaxTemperature

– производственные ограничения – ограничения, связанные с технологическими возможностями производства (например, минимальная ширина дорожек, максимальное количество слоев печатной платы).

Пример: минимальная ширина проводника на печатной плате должна быть не менее 0,1 мм.

Width(trace_i) >= MinWidth

– ограничения по надежности – ограничения, направленные на обеспечение требуемого уровня надежности и долговечности устройства.

Пример: обеспечение заданного среднего времени наработки на отказ (MTBF).

MTBF(system) >= Required_MTBF

Формулирование целевых функций и ограничений является итеративным процессом, требующим тесного взаимодействия между разработчиками и экспертами в предметной области. Важно учитывать все существенные факторы, влияющие на производительность, надежность и стоимость электронного устройства.

Пример: рассмотрим задачу оптимизации размещения компонентов на печатной плате для питания беспроводного зарядного устройства.

Целевая функция: Минимизация площади печатной платы.

Ограничения: минимальное расстояние между индукционной катушкой и другими компонентами: 5 мм% максимальная температура компонентов: 70 °C; соответствие требованиям электромагнитной совместимости (EMC); ограничения на габариты печатной платы (например, она должна помещаться в заданный корпус); соответствие требованиям по импедансу проводников питания.

В этом случае, алгоритм оптимизации будет искать оптимальное размещение компонентов, удовлетворяющее всем заданным ограничениям и минимизирующее площадь печатной платы. Результатом будет компактная и эффективная схема питания беспроводного зарядного устройства, удовлетворяющая требованиям по надежности и электромагнитной совместимости.

**Примеры оптимизации электронных средств.** Авторы имеют большой практический опыт проектирования электронных средств с учетом критериев оптимизации.

Статья [4] рассматривает ключевые аспекты прототипирования и концептуального проектирования средств медицинской электроники, акцентируя внимание на важности интерактивного взаимодействия врача и медицинского оборудования. Рассматриваются факторы, способствующие эффективной коммуникации между врачом и медицинским оборудованием, такие как интуитивный интерфейс и адаптация технологий к требованиям клинических задач. Описываются методы прототипирования интерфейсов, позволяющие учитывать потребности пользователей на ранних стадиях разработки. Рассматриваются инструменты и технологии, которые помогают создать прототипы, способствующие тестированию концепций и сбору обратной связи от врачей, что, в свою очередь, повышает качество и удобство использования медицинских устройств. Обсуждается процесс концептуального проектирования, который включает в себя этапы исследования

потребностей, генерации идей и создания дизайн-концепций. Подчеркивается значимость прототипирования и концептуального проектирования в развитии медицинских технологий, направленных на улучшение качества медицинских услуг и опыт взаимодействия врачей с оборудованием.

В работе [5] приводится описание общих принципов проектирования технических систем. Рассмотрены основные этапы и методология моделирования и проектирования технических систем. Приведены практические примеры моделирования технических систем. Предназначено для студентов второй ступени высшего образования технических университетов, может быть использовано аспирантами и инженерами, занимающимися вопросами моделирования и оптимального проектирования технических систем.

В статье [6] рассмотрены ключевые аспекты тепловых процессов, происходящих в радиоэлектронных устройствах, и предложены алгоритмы, позволяющие эффективно анализировать и прогнозировать температурные режимы в таких конструкциях. Авторы подчеркивают важность термического анализа для обеспечения надежности и долговечности элементов радиоэлектронной аппаратуры, а также необходимость учета различных факторов, влияющих на тепловые характеристики. В результате проведенных исследований предложены алгоритмы, позволяющие оптимизировать тепловые процессы, повышая эффективность радиоэлектронных устройств.

Статья [7] показывает, что современные технические (вычислительные, электронные) средства в большинстве случаев строятся на базе радиоэлектронных компонентов, например процессоры, графические чипы и т. д., которые при работе выделяют десятки ватт тепловой энергии. В связи с этим одной из приоритетных задач проектирования таких устройств становится эффективное отведение посредством введения пассивных или активных систем охлаждения избыточной тепловой энергии как от полупроводниковых приборов, так и от электронных средств в целом. Цель работы. Исследование влияния конфигурации и формы внешних ребер герметичных корпусов технических средств (ГК ТС), особенностью которых является использование внутри корпусов только пассивных систем охлаждения, а также сравнительный анализ эффективности отведения тепла от процессора для каждой рассматриваемой конструкции корпуса. Материалы и методы. Эксперименты проводились на разработанных в программной среде SolidWorks Flow Simulation трехмерных параметрических моделях различных типов ГК ТС, особенностью которых являлась различная конфигурация тепловых каналов, сформированных внешним оребрением крышки корпуса. Результаты. Реализация моделей позволила исследовать процесс охлаждения процессора, установленного в современных ГК ТС, а также проанализировать влияние конфигурации и формы ребер на отведение избыточной тепловой энергии от процессора в режиме пассивного охлаждения и при обдуве ГК ТС воздухом, движущимся сверху (перпендикулярно крышке) или сбоку (параллельно крышке) при постепенном увеличении мощности процессора с 10 до 25 Вт. Показано, что оребренный корпус при пассивном охлаждении обеспечивает отвод тепла от процессора мощностью 10 Вт больше, чем неоребренный (понижение температуры составляет 4.1 °С); при 25 Вт – на 11.01 °С. Установлено, что направление (перпендикулярное или параллельное) движения воздуха при обдуве ГК ТС значительно влияет на эффективность охлаждения нагретой поверхности корпуса (при мощности процессора 45 Вт разность составляет более 10 °С). Заключение. Разработанные трехмерные модели позволили наиболее эффективно реализовать систему охлаждения теплонагруженных высокомощных радиоэлектронных компонентов, расположенных в герметичных корпусах, за счет реализации внешнего оребрения корпусов.

В материалах [8] Представлены результаты исследования по отведению тепловой энергии от процессоров при помощи кулеров с воздушным охлаждением (КВО). Проведены четыре эксперимента, в которых при помощи трехмерного твердотельного параметрического моделирования выявлены наиболее эффективные варианты исполнения

КВО. В первом эксперименте был разработан башенный радиатор с несущей конструкцией в виде медных прутков и установленным вертикально вентилятором. Это дало возможность определить эффективность передачи тепловой энергии по радиатору. В следующем эксперименте в разработанной ранее модели заменили медные прутки на тепловые трубки и убрали вентилятор, что позволило обосновать необходимость наличия или отсутствия обдува при постепенном увеличении мощности процессора во время его эксплуатации. В третьем случае изменяли скорость вращения вентилятора, благодаря чему установили предельную скорость воздушного потока для эффективного охлаждения радиатора. В последнем эксперименте определили оптимальную конструкцию КВО, позволившую эффективно реализовать рассеивание тепловой энергии от процессора при различной его мощности. Экспериментально доказана необходимость использования КВО в таких современных технических средствах, как персональные компьютеры, сервера, вычислительные комплексы и т. д., в которых нет ограничений по обеспечению массогабаритных параметров.

В [9] представлены результаты изучения влияния длины и количества тепловых трубок, входящих в состав радиаторной конструкции, на эффективность отведения избыточной тепловой энергии от современных процессоров. Проведены исследования для радиаторных конструкций, состоящих из теплосъемной пластины, тепловой трубки и теплообменника (ребристого радиатора), установленных на процессор и находящихся в открытой среде (движение воздуха происходит без перемешивания, что характерно для свободной конвекции) или в замкнутой сре де (происходит циркуляция потоков воздуха в замкнутом контуре, что свойственно для естественной конвекции в ограниченном пространстве). Численное моделирование осуществлялось посредством модуля Flow Simulation программного комплекса SolidWorks. Показано, что от естественного движения потоков воздуха в открытой или замкнутой среде значительно зависит значение разности температур, сформированных на концах тепловых трубок (ТТ). Установлено, что с увеличением длины ТТ от 100 мм до 500 мм происходит увеличение разности температур как в случае движения потоков воздуха в открытой среде, так и в замкнутой. В частности, увеличение разности температур на концах одной ТТ диаметром 6 мм при мощности процессора 50 Вт составило 29,54 °С (открытая среда) и 47,14 °С (замкнутая среда); для трех ТТ – 9,13 °С (открытая среда) и 16,28 °С (замкнутая среда); для пяти ТТ – 5,24 °С (открытая среда) и 10,11 °С (замкнутая среда). Установлено, что увеличение количества ТТ диаметром 6 мм и длиной 500 мм от 1 до 5 шт. приводит к снижению разности температур, в частности, при мощности процессора 50 Вт разность температур составила 36,17 °С (одна ТТ в открытой среде) и 55,59 °С (одна ТТ в замкнутой среде); 11,04 °С (три ТТ в открытой среде) и 19,06 °С (три ТТ в замкнутой среде); а также 6,3 °С (пять ТТ в открытой среде) и 11,56 °С (пять ТТ в замкнутой среде). Полученные результаты могут быть использованы для модернизации систем охлаждения различных технических средств, построенных на базе процессоров, а также проектировании новой высокопроизводительной аппаратуры с учетом использования тепловых трубок.

**Заключение.** В современных условиях, когда требования к проектированию становятся все более сложными и многообразными, задача оптимизации в генеративном проектировании приобретает особую актуальность. Определение целевых функций и их ограничений является ключевым этапом в этом процессе, так как от правильного выбора этих параметров зависит эффективность и успешность проектных решений.

В ходе исследования проанализированы различные подходы к формулированию целевых функций. Основное внимание было уделено тому, как правильно установить ограничения, которые могут быть связаны как с физическими свойствами материалов, так и с технологическими процессами.

Рассмотрены примеры применения методов оптимизации в генеративном проектировании. Гибкость и адаптивность в подходах к оптимизации открывают новые

возможности для инновационных решений, способствуя созданию более эффективных и устойчивых продуктов.

Дальнейшие исследования в области определения целевых функций и их ограничений будут способствовать развитию генеративного проектирования и повышению конкурентоспособности предприятий в различных отраслях. Интеграция передовых методов оптимизации и современных технологий, таких как машинное обучение и искусственный интеллект, может значительно расширить возможности проектирования, открывая новые пути для реализации креативных и инженерных идей.

## Список литературы

[1] Розенберг, Р. К. Введение в динамику физических систем / Р. К. Розенберг, Д. К. Карнопп. – New York : McGraw-Hill, 1983. – 420 с.

[2] Арора, Я. С. Введение в оптимальное проектирование / Я. С. Арора. – 4-е изд. – Oxford : Butterworth-heinemann, 2016. – 928 с.

[3] Деб, К. Многокритериальная оптимизация с использованием эволюционных алгоритмов / К. Деб. – Chichester : John Wiley & Sons, 2001. – 508 с.

[4] Алексеев, В. Ф. Прототипирование и концептуальное проектирование средств медицинской электроники = Prototyping and conceptual design of medical electronics / В. Ф. Алексеев, Д. В. Лихачевский, Г. А. Пискун // Медэлектроника –2024. Средства медицинской электроники и новые медицинские технологии : сборник научных статей XIV Международной научно-технической конференции, Минск, 5–6 декабря 2024 г. / Белорусский государственный университет информатики и радиоэлектроники [и др.]. – Минск, 2024. – С. 202–206.

[5] Моделирование и оптимальное проектирование технических систем : учебно-методическое пособие / В. Ф. Алексеев, Д. В. Лихачевский, Г. А. Пискун, В. В. Шаталова. − Минск : БГУИР, 2024. – 99 с.

[6] Алексеев, В. Ф. Алгоритмы иерархического анализа тепловых процессов радиоэлектронных средств / В. Ф. Алексеев, Д. В. Лихачевский // Цифровая среда: технологии и перспективы : сборник материалов II Международной научно-практической конференции, Брест, 31 октября–1 ноября 2024 г. / Брестский государственный технический университет ; редкол.: Н. Н. Шалобыта (гл. ред.) [и др.]. – Брест, 2024. – С. 221–226.

[7] Пискун, Г. А. Влияние конфигурации и формы внешних ребер герметичных корпусов технических средств на эффективность отведения тепла от процессора = Effect of the Configuration and Shape of External Ribs of Sealed Enclosures of Electronic Devices on Heat Removal Efficiency / Г. А. Пискун [и др.] // Известия вузов России. Радиоэлектроника. – 2023. – Т. 26, № 5. – С. 63–75.

[8] Моделирование отведения тепловой энергии от процессоров при помощи кулеров воздушного охлаждения=Simulation of Thermal Energy Removal from Processors Using Air Coolers / Г. А. Пискун [и др.] // Доклады БГУИР. – 2023. – Т. 21, № 4. – С. 54–62.

[9] Оценка влияния длины и количества тепловых трубок на эффективность отведения избыточной тепловой энергии от процессора = Assessment of the influence of the length and number of heat pipes on the efficiency of the removal of excess thermal energy from the processor / Г. А. Пискун [и др.] // Весці Нацыянальнай акадэміі навук Беларусі. Серыя фізіка-тэхнічных навук. – 2024. – Т. 69, № 2. – С. 139–150.

## Авторский вклад

**Алексеев Виктор Федорович** – постановка задачи исследования, определение целевых функций и ограничений в задачах оптимизации, заключение.

**Лихаческий Дмитрий Викторович** – постановка задачи исследования, примеры оптимизации электронных средств.

**Пискун Геннадий Адамович** – введение, определение ограничений в задачах оптимизации, примеры оптимизации электронных средств.

# DETERMINATION OF OBJECTIVE FUNCTIONS AND THEIR CONSTRAINTS IN OPTIMIZATION PROBLEMS IN GENERATIVE DESIGN

***V.F. Alekseev***
*Associate Professor, Department of Information Computer Systems Design, PhD of Technical sciences, Associate Professor*

***D.V. Likhachevsky***
*Dean of the Faculty of Computer Design of BSUIR, PhD of Technical Sciences, Associate Professor*

***G.A. Piskun***
*Associate Professor of the Department of Design of Information and Computer Systems of BSUIR, PhD of Technical Sciences, Associate Professor*

**Abstract.** The article considers key aspects of defining objective functions and their constraints in the context of optimization problems related to generative design. The article analyzes methods for formulating objective functions, their impact on design results, and methods for defining constraints that take into account various technical, economic, and environmental factors. Particular attention is paid to practical examples illustrating the impact of correctly defining objective functions and constraints on design efficiency.

The article will be of interest to specialists in the field of engineering, design, and optimization, as well as to researchers interested in modern design methods.

**Keywords**: generative design, objective function, constraints, optimization of design solutions.

УДК 04.056.53

# МЕТОДИКА ПОВЫШЕНИЯ БЕЗОПАСНОСТИ ДИСТРИБУТИВА ОС LINUX С БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИЕЙ

***Г.П. Колбанов***
*учащийся УО «Национальный детский технопарк»*
grigoriy.kolbanov@gmail.com

***Д.А. Романов***
*учащийся УО «Национальный детский технопарк»*
dimromgomell@gmail.com

***Е.С. Белоусова***
*Доцент кафедры защиты инфомации БГУИР, кандидат технических наук, доцент*
belousova@bsuir.by

***Г.П. Колбанов***
*Обучается в Гимназии №1 г. Витебска. Область научных интересов связана с информационной безопасностью, криптографией, разработкой программ и защищённых операционных систем.*

***Д.А. Романов***
*Обучается в Гомельском государственном областном лицее. Область научных интересов связана с искусственным интеллектом, машинным обучением, нейронными сетями и созданием программным продуктов.*

***Е.С. Белоусова***
*Окончила Белорусский государственный университет транспорта. Область научных интересов связана с областью информационной безопасность, разработкой и внедрением средств защиты инфомации в корпоративные сети организаций.*

**Аннотация.** В статье рассматриваются современные подходы к повышению безопасности операционных систем (ОС), так как в условиях роста киберугроз необходимо применять комплексные меры защиты Linux-дистрибутивов, применяемых в критически важных объектах информатизации, обрабатывающих большие объемы данных. В процессе разработки нового дистрибутива ОС Secux Linux авторами статьи была разработана методика повышения безопасности подобный операционных систем на основе внедрения доверенной загрузки с верификацией initramfs, использования Secure Boot с собственными ключами, интеграции TPM и PCR sign policy.

**Ключевые слова:** Linux; Secure Boot, TPM, шифрование диска, изоляция приложений.

**Введение.** В условиях активного использования ОС Linux в системах обработки и хранения больших данных, в том числе и в критически важных объектах информатизации, необходимо применять комплексные меры по защите таких систем. В современных дистрибутивах Linux механизм загрузки системы имеет ряд уязвимостей, которые могут быть использованы нарушителями для обхода систем защиты ОС, включая компрометацию процесса расшифровки диска. Стандартная загрузка Linux в большинстве случаев основана на Shim, подписанном ключами Microsoft. Grand Unified Bootloader (GRUB), являясь основным загрузчиком, не выполняет проверку целостности загружаемых компонентов, включая initramfs – критический этап, отвечающий за расшифровку корневой файловой системы. В результате нарушитель может подменить initramfs, внедрив вредоносный код, способный перехватить пароль от зашифрованного диска.

**Выбор методов повышения безопасности ОС Linux.** К основным методам повышения относятся:

1 Обеспечение безопасности процесса загрузки. Рекомендуется использовать собственные ключи Secure Boot, отказаться от загрузчика GRUB в пользу systemd-boot.

2 Контроль целостности и TPM. Рекомендуется использовать аппаратную проверку целостности системы с использованием TPM [1].

3 Использование ядра linux-hardened – ядро Linux, ориентированное на безопасность, применяющее набор патчей для усиления защиты, чтобы уменьшить вероятность эксплуатации ядра.

4 Реализация полного шифрования диска с использованием LUKS (Linux Unified Key Setup).

5 Политики мандатного контроля доступа.

6 Использование Wayland в качестве графического сервера.

7 Изоляция приложений с помощью Flatpak.

8 Использование межсетевого экранирования (Uncomplicated Firewall, UFW).

9 Использование биометрической аутентификации.

**Методика повышения безопасности дистрибутива ОС Linux.** В процессе создания дистрибутива ОС Secux Linux были использованы все описанные выше методы повышения безопасности.

ОС Secux Linux – это современный дистрибутив на базе Arch Linux, разработанный для обеспечения максимального уровня безопасности на всех этапах работы системы. В Secux Linux есть возможность установки KIRTapp – программы, обеспечивающей биометрическую аутентификацию с распознаванием лица пользователя.

На основе опыта полученного при разработке дистрибутива ОС Secux Linux была сформирована методика повышения безопасности ОС Linux, которая включает следующие этапы:

1 Шифрование диска с помощью LUKS. Многие дистрибутивы позволяют это сделать при установке. Используйте LUKS 2 и функцию формирования ключа Argon2id [2].

2 Установка системы, ядра linux-hardened, apparmor, ufw. AppArmor должен запускаться при каждом включении устройства по этому добавим его в параметры запуска cmdline. Для применения профилей apparmor необходимо включить автоматический запуск сервиса apparmor [3].

3 Настройка генерации UKI. Для этого дистрибутив Secux Linux настраивает mkinitcpio, редактирует cmdline и создает политику подписи PCR.

4 Использование собственных ключей Secure Boot. Для этого Secux Linux использует утилиту sbctl. Настройте автоматическую подпись загрузочных программ (sbctl sign --save). Если ваше устройство не поддерживает замену ключей Secure Boot, то используйте SHIM и собственные ключи Machine Owner Key (MOK).

5 Настройка разблокировки диска с помощью TPM и программы Security Manager (интегрирован в Secux Linux).

6 Использование Wayland.

**Заключение.** Разработанная методика повышения безопасности дистрибутива ОС Linux была реализована в созданном дистрибутиве Secux Linux. Тестирование Secux Linux показало, что применяемые комплексные меры защиты такие как проверка загрузки, использование TPM, усиленное ядро, шифрование, мандатный контроль доступа и изоляция приложений, позволяет минимизировать риски компрометации системы. Описанная методика повышения безопасности может использоваться для любых дистрибутивов ОС Linux.

Необходимо отметить, что дистрибутив Secux Linux включает в себя все описанные методы повышения безопасности, которые полностью активируются при установке ОС. Использование ОС Secux Linux с системой биометрической аутентификации KIRTapp в

критических важных объектах информатизации позволяет минимизировать риск утечки конфиденциальных данных.

### Список литературы

[1] Arthur W., Challener D., Goldman K. A Practical Guide to TPM 2.0. USA: Apress, 2015. – 392 pp.

[2] Dodis Y., Shrimpton E. T. Advances in Cryptology. USA: Springer, 2022. – 828 pp.

[3] Donald A. Tevault Mastering Linux Security and Hardening. United Kingdom, Packt Publishing, 2023. – 652 pp.

### Авторский вклад

**Колбанов Григорий Павлович** – выбор методов повышения безопасности ОС Secux Linux, реализация и тестирование методов повышения безопасности ОС Secux Linux, составление методики повышения безопасности ОС Linux.

**Романов Дмитрий Алексеевич** – выбор методов биометрической аутентификации, реализация и внедрение системы биометрической аутентификации в ОС Secux Linux.

**Белоусова Елена Сергеевна** – постановка цели, формулировка задач исследования, координирование процесса разработки дистрибутива ОС Secux Linux, функциональное тестирование дистрибутива ОС Secux Linux.

## METHOD OF INCREASING THE SECURITY OF THE LINUX OS DISTRIBUTION WITH BIOMETRIC AUTHENTICATION

***G.P. Kolbanov***
*Student of the Educational Institution «National Children's Technopark»*

***D.A. Romanov***
*Student of the Educational Institution «National Children's Technopark»*

***E.S. Belousova***
*Associate Professor of the Informational Security Department of BSUIR, PhD of Technical Sciences, Associate Professor*

**Abstract.** The article discusses modern approaches to improving the security of operating systems (OS), since in the context of growing cyber threats it is necessary to apply comprehensive measures to protect Linux distributions used in critical IT facilities that process large volumes of data. In the process of developing a new OS distribution Secux Linux, the authors of the article developed a technique for improving the security of such operating systems based on the implementation of trusted boot with initramfs verification, using Secure Boot with its own keys, integrating TPM and PCR sign policy.

**Keywords**: Linux, Secure Boot, TPM, Disk Encryption, Isolation.

*Научное издание*

# BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ

# BIG DATA AND ADVANCED ANALYTICS

СБОРНИК НАУЧНЫХ СТАТЕЙ
XI МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

(Республика Беларусь, Минск, 23–24 апреля 2025 года)

В авторской редакции
Ответственный за выпуск *Д. В. Лихачевский*
Компьютерная верстка *Е. М. Косарева*

Подписано в печать 17.04.2025. Формат 60×84 1/8. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 58,36. Уч.-изд. л. 56,26. Тираж 50 экз. Заказ 48.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,
№2/113 от 07.04.2014, №3/615 от 07.04.2014.
Ул. П. Бровки, 6, 220013, г. Минск