

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 обуславливает 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.

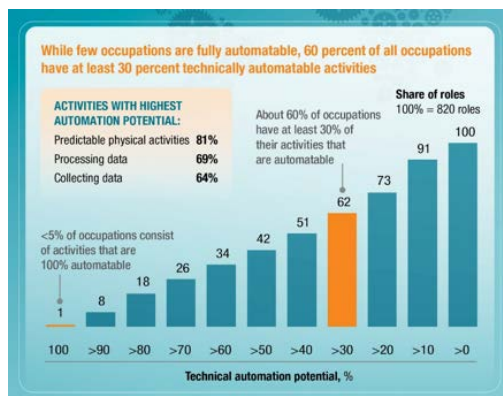


Рисунок 2. Степень автоматизируемости различных профессий

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-procness-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.