

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.6:52

Береснев
Дмитрий Владимирович

Система обработки больших массивов астрономических данных

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники
по специальности 1-40 81 02 «Технологии виртуализации и облачных
вычислений»

Научный руководитель
Самаль Дмитрий Иванович
кандидат техн. наук, доцент

МИНСК 2019

ОГЛАВЛЕНИЕ

ОГЛАВЛЕНИЕ	2
ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ.....	3
ВВЕДЕНИЕ	7
СОДЕРЖАНИЕ РАБОТЫ	11
ЗАКЛЮЧЕНИЕ	12
СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ	15

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы магистерской диссертации: Крайне высокие темпы роста объемов информации, повышение качественных характеристик собранных данных открывают перед астрономами и астрофизиками новые направления для исследований. Однако очень большие объемы собранных данных требуют создания и применения новых подходов к их обработке и новых методов анализа, которые реализуются с помощью технологий распределенной обработки больших данных, облачных вычислений, а также с помощью последних достижений физико-математических наук, которые в свою очередь позволяют реализовать эффективные алгоритмы, составляющие основу машинного обучения и интеллектуального анализа данных. Таким образом, в виду быстрого роста объемов экспериментальных данных в современной астрономии и астрофизике, а также появлением более эффективных методов их анализа, вопросы, связанные с проектированием и разработкой систем распределенной обработки данных и созданием алгоритмов машинного обучения и анализа данных, реализующих более эффективные методики обработки и анализа, останутся актуальными еще достаточно длительное время.

Аргументация выбора темы. Тема диссертационного исследования «Система обработки больших массивов астрономических данных» полностью соответствует паспорту специальности 1-40 81 02 «Технологии виртуализации и облачных вычислений» [1], а также данная работа соответствует основным направлениям научных исследований университета и факультета [2, 3].

Противоречие. Наличие прогрессивных методов машинного обучения и алгоритмов интеллектуального анализа данных и недостаточная проработанность способов реализации данных методов в виде завершенной системы, эффективно решающей задачи астрономии и астрофизики.

Проблематика. Комплекс вопросов, связанных с поиском эффективных способов реализации методов машинного обучения и алгоритмов анализа данных в виде завершенной системы.

Цель работы: состоит в исследовании процесса автоматизации обработки больших объемов астрономических данных с применением методов машинного обучения и интеллектуального анализа данных и разработке соответствующей системы.

Задачи исследования: В соответствии с поставленной целью, можно выделить следующие задачи:

1. Выполнить анализ предметной области.
2. Спроектировать программное средство для автоматизации процесса анализа астрономических данных.

3. Разработать систему обработки астрономических данных.

Объект исследования – разнородные астрономические данные, собранные из разных источников.

Предмет исследования – методы машинного обучения и алгоритмы анализа данных, которые позволяют производить обработку астрономической информации.

Общая характеристика проблематики научных вопросов. В последнее время практически во всех сферах науки наблюдается очень быстрый рост объемов данных, полученных в ходе научных наблюдений и экспериментов. И такие науки как астрономия и астрофизика не являются исключением. Стремительный рост экспериментальных данных в астрономии и астрофизике обусловлен совершенствованием оптических телескопов и развитием сферы информационно-коммуникационных технологий. Прогресс в области высоких технологий, микро- и наноэлектроники, привел к созданию исследователями экспериментальных систем, генерирующих большие объемы данных, достигающие сотен терабайт и петабайт в самых различных сферах науки и техники, в том числе и в астрономических наблюдениях. Большие массивы данных, получаемые в ходе проведения научных исследований и экспериментов необходимо хранить, обрабатывать, передавать и анализировать. По этой причине современные исследования в научных направлениях, находящиеся на стыке наук, требуют новых математических методов, современных инженерных подходов и технических решений, которые позволили бы получать из больших массивов данных новые знания. Разнородность, распределенность, колоссальные объёмы и скорость прироста этих данных, усложняют решение задач анализа, хранения, обработки и передачи информации. Таким образом, в современных научных исследованиях приходится сталкиваемся с необходимостью решения проблемы резкого увеличения передаваемых объемов информации и роста информационных потоков. Особо актуально решение проблемы работы с большими массивами данных в современной астрономии. Системы для астрономических наблюдений позволяют получать данные с более высоким разрешением и с больших участков неба, чем это было ранее. Сегодня, астрономические наблюдения ведутся не только в видимом спектре, а во всем диапазоне электромагнитного спектра, при этом единственное наблюдение, которое длится от нескольких секунд до нескольких минут, дает от нескольких мегабайт до нескольких гигабайт информации. Таким образом, построение систем обработки астрономических данных, а также разработка эффективных методов анализа являются крайне актуальными задачами в современных исследованиях.

Методологический базис. Методологическую основу данной диссертационной работы составляют различные методы познания. При

подготовке работы использовались следующие общенаучные подходы и специальные методы [4]:

- системный подход;
- структурно-функциональный метод;
- моделирование;
- методы эмпирического исследования (эксперимент, сравнение, описание и измерение);
- методы теоретического познания (формализация, восхождение от абстрактного к конкретному);
- логические методы.

А также использовались методологии архитектурного проектирования информационных систем из сферы ИКТ.

Степень достоверности и апробация результатов. Результаты исследований, сформулированные в данной работе, получены на основе корректного использования взаимно дополняющих друг друга теоретических и экспериментальных методов исследований. О результатах работы можно судить по фактическому функционированию программного обеспечения, которое было разработано в рамках темы диссертации, которое в целом подтверждает некоторые из выдвинутых гипотез. Результаты исследований опубликованы в виде тезисов на следующих научных конференциях:

1. 54-я научно-техническая конференция аспирантов, магистрантов и студентов БГУИР, которая проходила в мае 2019 в Минске. Тема тезисов: «Морфологическая классификация галактик» [1 – А.].

2. Пятая международная научно-практическая конференция «Big Data and Advanced Analytics», которая проходила 13.03.2019 – 14.03.2019 в Минске. Тема тезисов: «Классификация галактик с использованием методов глубокого обучения» [2 – А.].

3. Третья международная научно-техническая конференция «COMPUTER AND INFORMATIONAL SYSTEMS AND TECHNOLOGIES 2019», которая проходила 23.04.2019 – 24.04.2019 в городе Харьков, Украина. Тема тезисов: «Анализ сверточных и капсульных нейронных сетей» [3 – А.].

4. Третья международная научно-техническая конференция «COMPUTER AND INFORMATIONAL SYSTEMS AND TECHNOLOGIES 2019», которая проходила 23.04.2019 – 24.04.2019 в городе Харьков, Украина. Тема тезисов: «Анализ методов понижения размерности пространства» [4 – А.].

Степень разработанности темы. Популярность исследований в такой области как искусственный интеллект сейчас достигает высокого уровня. Об этом может свидетельствовать целый ряд книг и количество опубликованных научных статей. Однако только сейчас приходит осознание философских концепций и математических методов, лежащий в основе построения

искусственного интеллекта, а также сфер их применения и накладываемых ограничений.

Научная новизна. Проблематика, которая была обозначена несколько выше по тексту данного раздела, рассмотрена с использованием другого подхода к разработке. Данный подход характеризуется объединением научно-технических методологий [5,6] и нестандартной архитектурой программного обеспечения, особенности которой будут рассмотрены более подробно в следующих главах.

Теоретическая и практическая значимость. Материалы данной работы с теоретической точки зрения будут способствовать развитию и дальнейшему осмыслению применения методов машинного обучения и алгоритмов интеллектуального анализа данных для реализации интеллектуальных систем, решающих задачи астрономии и астрофизики, а с практической даст толчок к созданию производительных систем для автоматизации процессов обработки больших объемов астрофизических данных.

Полученные результаты. Создана система для обработки больших массивов астрономических данных, которая решает целый ряд астрофизических задач.

Перспективы. Данная работа открывает следующие направления исследований, по которым можно производить дальнейшее проектирование и разработку, основываясь на текущих результатах:

1 Разработка более совершенных парадигм для архитектурного проектирования инфраструктуры.

2 Разработка новых концептуальных решений при построении интеллектуальных систем для автоматизации процессов обработки астрофизических данных.

3 Создание алгоритмов и математических моделей, адаптированных для обработки больших объемов астрофизических данных при ограниченных вычислительных ресурсах.

4 Новые математические методы для решения астрономических и астрофизических задач.

5 Адаптация новейших методов интеллектуального анализа данных и машинного обучения для построения прогностических моделей в условиях ограниченности вычислительных ресурсов.

6 Создание системы аналитической обработки информации на базе высокопроизводительного сервера для агрегации данных, собранных из различных астрономических баз данных и хранилищ.

ВВЕДЕНИЕ

Современная астрономия, астрофизика и астроинформатика решают целый ряд достаточно сложных, но крайне интересных с научной точки зрения задач. Таким образом, следует дать оценку современного положения и состояния решаемых задач, осветить основания диссертационного исследования и исходные данные для начала разработки темы диссертации, дать характеристику степени достоверности и апробации результатов, степени разработанности диссертационной темы и назначению диссертационной работы.

Астрономия, астрофизика и астроинформатика решают следующий перечень задач:

1 *Классификация при помощи исследования спектра космических объектов (например, является ли объект галактикой, квазаром или звездой).* В современной астрономии и астрофизике задача описания, составления и фиксации характеристик космических объектов является крайне важной задачей. Очень часто исследователям крайне сложно по изображению определённого участка неба определить – это очень далёкая галактика, звезда, пульсар или квазар. Данные объекты на снимках могут выглядеть абсолютно одинаково. Эти космические объекты можно отличить, если исследовать их спектр. Различные атомы поглощают только определенные длины волн, поэтому, если есть спектр некоторого космического объекта, то есть возможность определить химический состав. Солнце является очень ярким космическим объектом, и его свет можно разложить в спектр с помощью даже простейших приборов, но чем меньше излучение от объекта, тем сложнее получить его спектр. Для объектов, которые обладают очень низким уровнем излучения, получение спектра вообще практически невозможно, так как требуется слишком много времени. Однако, для того, чтобы получить снимки некоторого объекта с помощью разных фильтров потребуется относительно немного времени. Фильтры пропускают только волну определённой длины. Таким образом, можно все объекты снимать с помощью широкополосных фильтров, и для тех космических объектов, для которых это возможно, снять спектр. Далее, по спектральным данным анализируемых объектов можно выполнить классификацию на группы – звезда, галактика или квазар. После выполнения классификации может быть составлен достаточно качественный набор данных, используя который можно строить Data Science конвейер для обработки данных – одни сочетания спектральных характеристик – квазар, а другие – галактика или звезда. Затем, «предобученной» Data Science модели на вход подать новые экспериментальные данные, для которых меток с правильными ответами – и получить некоторый прогноз. Таким образом, с помощью технологий машинного обучения и анализа данных

получиться определить классификационную группу объекта, не снимая его спектр.

2 *Исследование красного смещения.* Чем дальше расположен квазар, галактика или звезда, тем выше будет значение скорости удаления (отдаления), соответственно и выше будет величина показателя красного смещения. Зная спектр объекта, можно вычислить красное смещение: для этого достаточно найти линии самого распространенного химического элемента, например, водорода. Однако, получить спектр не всегда возможно. Таким образом, следует получить изображение объекта в разных фильтрах и применить методы машинного обучения и анализа данных. Располагая информацией о красном смещении для множества объектов, можно значительно улучшить понимание эволюции Вселенной и более детально изучить некоторые аспекты такого понятия как тёмная энергия, которая, по современным представлениям в астрофизике, является причиной того, что наша Вселенная расширяется.

3 *Классификация галактик.* Морфологические характеристики галактики могут содержать информацию о внутренней динамике движения звёзд, ядерной активности и процессах формирования. Классификация галактик, базирующаяся на визуальных признаках сильно связана и с астрофизическими характеристиками. Например, наличие выраженных приливных сил и нескольких ядер означают то, что главным механизмом звездного формирования в этой галактике было слияние. А также, отсутствие ранее перечисленных признаков подсказывает, что эволюция данной галактики обусловлена протеканием более медленного процесса.

4 *Исследование сверхновых (англоязычное название проекта *The Dark Energy Survey*).* В 2013 году стартовала исследовательская программа под названием *The Dark Energy Survey*, которая была рассчитана на 5 лет. В рамках данной программы исследователи получают изображения неба Южного полушария с помощью инфракрасных и оптических камер. Одной из главных целей данной программы является проведение научных исследований динамики Вселенной, а также ее крупномасштабной структуры, при этом используя в качестве одного из методов сверхновые типа Ia. В рамках данной программы ученые получают ежедневно множество изображений, и каждое из этих изображений нуждается в обработке. Исследователи разработали алгоритм под названием *autoScan*, который способен отличать артефакт на изображении от реального объекта. Такой же подход будет использоваться для большого телескопа LSST, который будет запущен в 2019 году для обзора неба. Благодаря алгоритму *autoScan* исследователям представляется возможность обнаруживать с максимальной эффективностью новые космические объекты (например, сверхновые), а новейшие системы и технологии позволяют отличить их от дефектов на изображении.

5 Распознавание состава атмосфер экзопланет. В феврале 2016 года была опубликована научная статья, в тексте которой рассказывалось про Data Science модель на базе искусственной нейронной сети RobERt [7], точность которой составила 99,7%. RobERt (Robotic Exoplanet Recognition) – это модель на базе нейронной сети, которая предназначена для классификации спектров экзопланет. Спектральный поиск экзопланетных атмосфер часто требует отбора молекулярных или атомарных зетемнений, которые определяются человеком. Однако, в современную эпоху следует избегать ручной обработки данных. Входные данные, зависящие от человеческого фактора, могут в худшем случае привести к построению неполной модели и искажениям в поиске. Алгоритм RobERt основан на глубоких нейронных сетях доверия (DBNN, Deep Belief Neural Network), обученных достаточно точно распознавать молекулярные сигнатуры для широкого спектра планет, атмосферных тепловых профилей и композиций. Реконструкции изученных признаков, также называемые «снами» сети, указывают на хорошую конвергенцию и точное представление молекулярных признаков в DBN. Используя эти глубокие нейронные сети, можно работать над интеллектуальными поисковыми алгоритмами, которые сами понимают природу наблюдаемых спектров, могут учиться на основе текущих и прошлых данных и делать качественные прогнозы на основе атмосферных признаков, используемых на количественном этапе процесса поиска. Нейронная сеть RobERt оказалась способна распознавать характеристики атмосфер планет по их спектру. Для обучения модели на базе данной нейросети были смоделированы 85750 спектров экзопланет, которые были ранее обнаружены на пяти типах экзопланет. В спектре каждой из планет содержался какой-то один газ и нейросеть должна была определить данный газ. В итоге Data Science модель была способна определять качественный состав атмосферы, соотношение химических элементов и температурные характеристики. RobERt предполагается использовать для обработки данных разнообразных телескопов в будущем, которые смогут получать массово спектры экзопланет.

В январе 2018 была опубликована другая научная статья по идентификации экзопланет [8]. Космический телескоп Kepler был разработан в NASA для определения частоты вращения планет размером с Землю, вокруг солнцеподобных звезд, но эти планеты находятся на самой границе чувствительности обнаружения. Точное определение частоты для этих планет потребует автоматизированной и точной оценки вероятности того, что индивидуальные кандидаты действительно являются планетами, даже при низких соотношениях сигнал-шум. Таким образом, метод классификации сигналов потенциальных планет с использованием глубокого обучения, как класса алгоритмов машинного обучения и анализа данных, которые недавно

стали наиболее эффективны в самых разных задачах. Тренировка модели на базе глубокой сверточной нейронной сети необходима для того, чтобы выполнить прогноз, является ли данный сигнал от экзопланеты или это ложное срабатывание, вызванное астрофизическими или инструментальными явлениями. Данная модель оказалась очень высокоэффективной при ранжировании отдельных кандидатов по вероятности того, что они действительно являются планетами: 98,8% сигнал от экзопланет выше ложноположительных сигналов в тестовом наборе. Также данная модель была использована при анализе нового набора данных, который был идентифицирован при поиске известных многопланетных систем Кеплера. Статистически были проверены две новые планеты, которые были найдены моделью с высокой степенью уверенности. Одна из этих планет является частью пятипланетной резонансной цепи вокруг Кеплера-80. Другая планета вращается вокруг Kepler-90, звезды, которая ранее была известна как система система содержащая семь планет.

6 *Воспроизводимость результатов.* Если эксперимент содержит только одно наблюдение, то с научной точки зрения сомнительно считать результат данного эксперимента достоверным. Каждое наблюдение в астрономии или астрофизике сопровождается некоторыми метаданными, например, параметры съемки, время наблюдения, местоположение, погода, фаза Луны, инструменты наблюдения, средства обработки и т.д. Сегодня получают и обрабатывают результаты, которые нельзя воспроизвести, используя другой инструмент. Таким образом, становится крайне актуальным построение модели машинного обучения и анализа данных для обработки всех наблюдений с учетом всех метаданных, которая была бы способна учесть все эти факторы и снизить вероятность ошибки, а также свести все результаты к стандартному «идеальному наблюдателю в вакууме». Таким образом, это способствовало бы увеличению точности всех научных исследований в области астрономии и астрофизики.

7 *Поиск радиосигналов искусственного происхождения.* Breakthrough Listen — проект Стивена Хокинга и Юрия Мильнера, целью которого является проведение поиска разумной внеземной жизни во Вселенной. Данный проект имеет проектный бюджет в 100 миллионов долларов и по времени будет иметь продолжительность 10 лет и является важной частью комплекса проектов Мильнера, которые имеют общее название Breakthrough Initiatives. Идея проекта была признана научным сообществом как наиболее всесторонняя и одна из масштабных попыток поиска внеземной разумной жизни. В настоящее время несколько команд смогли зафиксировать десятки радиосигналов, имеющих искусственное происхождение.

СОДЕРЖАНИЕ РАБОТЫ

Крайне высокие темпы роста объемов информации, повышение качественных характеристик собранных данных открывают перед астрономами и астрофизиками новые направления для исследований. Однако очень большие объемы собранных данных требуют создания и применения новых подходов к их обработке и новых методов анализа, которые стали доступны благодаря высокотехнологичному астрофизическому оборудованию, и которые реализуются с помощью последних достижений физико-математических наук, которые в свою очередь позволяют реализовать эффективные алгоритмы, составляющие основу машинного обучения и интеллектуального анализа данных.

Работа состоит из следующих структурных частей: Общая характеристика работы, Введение, 3 глав и Заключение.

В общей характеристике раскрыто описание следующих пунктов: отмечена актуальность темы магистерской диссертации, дана аргументация выбора темы и связь ее со специальностью, выделено противоречие и как следствие отмечено проблематика, вытекающая из противоречия, определена цель работы, сформулированы основные задачи диссертационного исследования, которые далее в главах конкретизируются, отмечен объект и предмет исследования, представлена общая характеристика научных вопросов в выбранной области, выделен пункт с апробацией результатов на конференциях, описана степень разработанности темы, сформулирована научная новизна, отмечена теоретическая и практическая значимость, сформулированы результаты и перспективы работы.

Во Введении дано обоснование актуальности диссертационного исследования, а также представлен обзор научной области и основных направлений в астрономии.

В первой главе «Характеристика предметной области и аналогов» дано описание современных направлений в астрономии и астрофизике, а также основным объектам – галактикам и квазарам. Представлена характеристика близких аналогов.

Во второй главе «Анализ астрофизических данных» представлен полный ход решения двух основных задач: классификации галактик на эллиптические, спиральные и иррегулярные, классификации объектов на звезды, галактики и квазары.

В третьей главе «Корреляционный и регрессионный анализ данных SDSS» выделены цели и задачи корреляционно-регрессионного анализа, приведены графические и численные результаты анализа, сделаны выводы по результатам.

ЗАКЛЮЧЕНИЕ

Очень быстрые темпы прироста объемов информации, улучшение характеристик собранных данных позволяют осваивать новые направления исследований в астрономии и астрофизике. Однако задачи хранения и структуризации очень больших массивов собранных данных требуют внедрения совершенно новых подходов в обработке и разработки новых методов анализа. Таким образом, возникает острая необходимость применения новых методов к последним собранным данным, и астрономические снимки галактик не являются исключением. Поэтому решение задачи классификации галактик крайне актуально сегодня в астрономии и астрофизике.

Методы и подходы машинного обучения и анализа данных позволяют осуществлять прогнозирование красного смещение галактик и производить построение объемных карт распределения космических объектов во Вселенной, которые сегодня крайне часто используются в исследовательских задачах астрономии и космологии. Однако, стоит отметить, что практическое применение построенных моделей к доступным на сегодня астрономическим каталогам, которые содержат фотометрические данные для космических объектов, требует внушительных вычислительных ресурсов.

Галактики – одно из самых ярких космических явлений, и с помощью таких инструментов как «Sloan Digital Sky Survey» и «Hubble» научному сообществу теперь доступны тысячи изображений галактик, которые можно анализировать и исследовать. Однако одна из современных проблем астрономов и астрофизиков заключается в способах и методах классификации этих изображений галактик. Целью настоящей работы являлась разработка программного средства на базе методов глубокого обучения для решения задачи морфологической классификации галактик. В качестве основной модели использовалась свёрточная нейронная сеть. Свёрточные нейронные сети (CNN) уже успешно использовались и применялись в разнообразных исследованиях в этой области, прежде всего благодаря их эффективности именно в распознавании изображений. Как и другие нейронные сети, CNN состоят из слоев нейронов (которые обрабатывают входные данные), связей и их соответствующих весов.

При использовании свёрточной нейронной сети возникает проблема – большое количество выделенных признаков. Использование огромного количества признаков для решения задач распознавания изображений оказалось неэффективно. Для уменьшения размера пространства признаков проводится субдискретизация (англ. pooling), путём разделения карты признаков, полученных от свёрточной нейронной сети, на фиксированное количество

частей на каждой части вычисляется её максимальное значение (англ. max pooling) или среднее значение (англ. mean pooling).

Можно выделить следующие преимущества сверточной нейронной сети:

- уменьшение количества обучаемых параметров и повышение скорости обучения по сравнению с полносвязной нейронной сетью;
- возможность распараллеливания вычислений и реализации алгоритмов обучения сети на графических процессорах (GPU).
- устойчивость к сдвигу позиции объекта во входных данных. При обучении свёрточная нейронная сеть сдвигается по частям объекта. Поэтому обучаемые признаки не зависят от позиции «важных частей», т.е. свёрточная нейронная сеть выделяет одинаковые признаки для двух картинок, хотя позиции на изображениях могут быть разные. Таким образом, это свойство свёрточной нейронной сети помогает повышать качество при решении разнообразных задач распознавания изображений.

Метод, на базе которого производилось построение программного средства для классификации, подробно описан в [13]. Данный подход дополняет и адаптирует последовательность Хаббла для решения практических задач на реальных астрономических данных (см. рис. 2.1). При проектировании программного обеспечения для морфологической классификации галактик использовались отдельные теоретические и прикладные аспекты из [14]. В ходе реализации программного средства была построена нейронная сеть, архитектура которой представлена на рис. 2.3.

Можно выделить следующие перспективы для дальнейших исследований:

- использование других архитектур нейронных сетей;
- подготовка и использов. свежих данных из цифрового обзора SDSS;
- использование аугментации данных;
- реализация дополнительных методов для предобработки;
- построение n-компонентного вектора признаков для последующей кластеризации данных и т.д.

Таким образом, разработанное программное средство позволяет производить классификацию галактик на три основных класса: эллиптические, спиральные и иррегулярные в соответствии с упомянутым выше практико-ориентированным подходом с достаточно высоким уровнем точности. Построение систем обработки астрономических данных, а также разработка эффективных методов анализа являются крайне актуальными задачами в современных исследованиях. Разработка интеллектуальных систем позволит автоматизировать многие процессы, что в свою очередь способно ускорить прогресс в астрономии и астрофизике.

В ходе выполнения данного диссертационного исследования были рассмотрены *ключевые понятия и концепции различных технологий машинного*

обучения и анализа данных, приобретены навыки по сборке и установке основных библиотек глубокого обучения, а также были рассмотрены ключевые алгоритмы и методы Data Science, которые позволяют построить достаточно эффективные конвейеры обработки для решения различного рода интеллектуальных задач.

По результатам обучения и работы различных моделей был проведен их анализ, дано описание каждого из этапов конвейера обработки.

Основные результаты диссертационного исследования:

– исследование в рамках задачи морфологической классификации галактик позволило реализовать соответствующую систему, которая производит классификацию с достаточно высоким уровнем качества;

– исследование в рамках задачи классификации объектов на звезды, галактики и квазары позволило разработать систему, способную выполнять классификацию с достаточно высоким уровнем качества, а также выбрать лучшую модель;

– исследование в рамках задач корреляционно-регрессионного анализа данных позволило выявить связь признаков из датасета SDSS с красным смещением квазаров.

Разработан конвейер обработки данных для решения задачи морфологической классификации галактик. Разработан конвейер обработки данных для решения задачи классификации объектов на звезды, галактики и квазары. Проведено исследование в рамках задач корреляционно-регрессионного анализа данных. Результаты исследований опубликованы в виде тезисов в научных конференциях. Также стоит отметить, что все поставленные цели достигнуты, а задачи решены в полном объеме.

Подводя итог, также стоит отметить, что в настоящее время неслучайно вызывают повышенный интерес при построении интеллектуальных систем различного масштаба и назначения. Алгоритмы и методы машинного обучения и анализа данных позволяют решить огромнейших пласт интеллектуальных задач, таких как распознавание изображений и видео, распознавание и синтез речи, распознавание и перевод текстов, прогнозирование событий и рядов и т.д. Поэтому интеллектуальное ПО позволяет повысить удобство использования компьютерных систем на новый качественный уровень и делает дисциплину Data Science одним из ключевых направлений в сфере исследований искусственного интеллекта.

СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

[1 – А.] Береснев, Д. В. Классификация галактик / Д. В. Береснев // Компьютерные системы и сети : материалы 54 науч. конф. аспирантов, магистрантов и студентов, Минск, 23–27 апреля 2018 г. / Белорус. гос. ун-т информатики и радиоэлектроники ; редкол.: В. А. Прытков (гл. ред.) [и др.]. – Минск, 2018. – С. 17.

[2 – А.] Береснев, Д. В. Классификация галактик с использованием методов глубокого обучения / Д. В. Береснев, Д. И. Самаль // BIG DATA Advanced Analytics: collection of materials of the five international scientific and practical conference, Minsk, Belarus, March 13 – 14, 2019 / editorial board: V. Bogush [etc.]. – Minsk, BSUIR, 2019. – P. 355-359.

[3 – А.] Береснев, Д. В. Анализ сверточных и капсульных нейронных сетей / Д. В. Береснев // Компьютерные и информационные системы и технологии : материалы Третьей международной научно-технической конференции, Харьков, 23–24 апреля 2019 г. / Харьковский нац. ун-т радиоэлектроники ; редкол.: И. В. Рубан (гл. ред.) [и др.]. – Харьков, 2019. – С. 18.

[4 – А.] Береснев, Д. В. Анализ методов понижения размерности пространства / Д. В. Береснев, Е. В. Шараев // Компьютерные и информационные системы и технологии : материалы Третьей международной научно-технической конференции, Харьков, 23–24 апреля 2019 г. / Харьковский нац. ун-т радиоэлектроники ; редкол.: И. В. Рубан (гл. ред.) [и др.]. – Харьков, 2019. – С. 27.