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

Информационные технологии и системы 2025 (ИТС 2025)

Материалы международной научной конференции
(Республика Беларусь, Минск, 19 ноября 2025)

Information Technologies and Systems 2025 (ITS 2025)

Proceeding of the International Conference
(Republic of Belarus, Minsk, 19th November 2025)

Минск БГУИР 2025

УДК 004
ББК 32.81
И74

Редакционная коллегия:

Л. Ю. Шилин (главный редактор),
Т. Н. Беляцкая, В. В. Голенков, В. И. Журавлев, А. А. Иванюк, С. В. Колосов,
Д. П. Кукин, А. В. Марков, А. А. Навроцкий, В. А. Рыбак, Д. В. Шункевич,
А. Ф. Трофимович, Н. В. Гракова,
М. В. Ковалёв (ответственный секретарь)

Информационные технологии и системы 2025 (ИТС 2025 =
И74 Information Technologies and Systems 2025 (ITS 2025) : материалы междунар. науч. конф., Минск, 19 нояб. / редкол. : Л. Ю. Шилин [и др.]. – Минск : БГУИР, 2025. – 306 с.
ISBN 978-985-543-858-9

Материалы международной научной конференции включают прошедшие рецензирование статьи международной научной конференции «Информационные технологии и системы 2025» (ИТС 2025).

Материалы конференции предназначены для преподавателей высших учебных заведений, научных сотрудников, студентов, аспирантов, магистрантов, а также для специалистов предприятий в сфере IT-технологий.

Материалы конференции одобрены организационным комитетом и публикуются в авторской редакции.

УДК 004
ББК 32.973.202

ISBN 978-985-543-858-9

©«Белорусский государственный университет информатики и радиоэлектроники», 2025

ПРЕДИСЛОВИЕ

19 ноября 2025 года в городе Минске прошла международная научная конференция «Информационные технологии и системы» (ИТС 2025, организованная факультетом информационных технологий и управления учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», базовой организации государств-участников СНГ по высшему образованию в области информатики и радиоэлектроники.

В течение дня на пленарном и шести секционных заседаниях было представлено 138 докладов по следующим направлениям:

Секция «Системы управления»:
- моделирование систем управления;
- автоматизированные исследования систем управления;
- проектирование систем управления;
- микропроцессорные системы управления;
- информационные технологии в системах управления;
- системы передачи и защиты дискретной информации;
- системы телемеханики;
- элементы и устройства систем управления..

Секция «Интеллектуальные информационные технологии»:
- проектирование баз знаний;
- интеллектуальный поиск информации;
- интеллектуальные решатели задач;
- пользовательские интерфейсы интеллектуальных систем.

Секция «Автоматизированные системы обработки информации»:
- математические модели и методы обработки информации: статистические, численные, логические, графовые, нейросетевые и другие;
- анализ многомерных данных;
- методологические проблемы и технологии подготовки специалистов в области систем обработки информации;
- технологии и стандарты программирования и проектирования систем обработки информации;
- системы и средства совместной работы пользователей;
- защита информации в автоматизированных системах обработки информации.

Секция «Проектирование встраиваемых систем»:
- проектирование и быстрое прототипирование встраиваемых систем на ПЛИС;
- методологии проектирования IP-компонент встраиваемых систем;
- программная и аппаратная защита IP-компонент;
- контролепригодное проектирование и тестирование встраиваемых систем;
- реконфигурируемые встраиваемые системы.

Секция «Математическое моделирование и компьютерная графика»:

- методы решения уравнений математической физики;
- методы нелинейной оптимизации;
- модели и методы тепло и массопереноса;
- математические модели и методы компьютерной графики;
- аппаратное обеспечение компьютерной графики;
- алгоритмы и методы интерактивных программных продуктов.

Секция «Технологии цифровой экономики»:

- технологии цифрового маркетинга;
- бизнес-аналитика и экономика данных;
- информационные системы принятия решений, планирования, учёта и контроля.

В работе конференции приняли участие учёные и специалисты Республики Беларусь, Российской Федерации, Китая, Мьянмы и Туркменистана.

Докладчики представляли такие высшие учебные заведения как: Белорусский государственный университет информатики и радиоэлектороники, Белорусский государственный университет, Белорусско-Российский университет, Гомельский государственный технически университет имени П. О. Сухого, Евразийский национальный университет им. Л.Н. Гумилёва, Инженерно-технологический университет Туркменистана имени Огуз хана, Московский региональный социально-экономический институт, Национальный исследовательский технологический университет МИСИС, Национальный исследовательский университет Высшая школа экономики (Пермский филиал), Полоцкий государственный университет имени Евфросинии Полоцкой, Российский университет дружбы народов, Санкт-Петербургский политехнический университет Петра Великого, Учреждение Гомельский областной клинический онкологический диспансер, Учреждение образования «Национальный детский технопарк».

Кроме того, в работе конференции приняли участие представители таких организаций как Объединённый институт проблем информатики Национальной академии наук Беларуси, Государственное предприятие «Центр радиотехники Национальной академии наук Беларуси», Kaspi Bank JSC, АТЕК, ООО «АльпИндустрия».

Оргкомитет благодарит всех, кто счел возможным поделиться результатами своей деятельности и принял участие в работе конференции.

ORGANISING COMMITEE

Prof. Dr.	**Vadim Bogush**	Rector of The Belarusian State University of Informatics and Radioelectronics
PhD	**Maksim Davydov**	First Vice-Rector of BSUIR
PhD	**Viktor Stempitsky**	Vice-Rector for Research of BSUIR
PhD	**Evgeni Shneiderov**	Vice-Rector for Education of BSUIR
	Dmitry Kuznetsov	Vice-rector for Personal Development of BSUIR
PhD	**Anatoly Navrotsky**	Dean of The Faculty of Information Technologies and Control of BSUIR

PROGRAMME COMMITEE

Prof. Dr.	**Leonid Shilin**	Professor of Information Technology automated systems Department of BSUIR
Prof. Dr.	**Vladimir Golenkov**	Professor of Intelligent Information Technologies Department of BSUIR
Prof. Dr.	**Vladimir Golovko**	Head of Intelligent Information Technologies Department of BrSTU
Prof. Dr.	**Alexander Ivaniuk**	Professor of Computer Science Department of BSUIR
Prof. Dr.	**Tatiana Belyatskaya**	Head of Management Department of BSUIR
PhD	**Dmitry Kukin**	Head of Computational Methods and Programming Department of BSUIR
PhD	**Alexander Markov**	Head of Control Systems Department of BSUIR
PhD	**Victor Rybak**	Head of Information Technologies in Automated Systems Department of BSUIR
PhD	**Anatoly Navrotsky**	Dean of The Faculty of Information Technologies and Control of BSUIR
PhD	**Daniil Shunkevich**	Head of Intelligent Information Technologies Department of BSUIR
Prof. Dr.	**Stanislav Kolosov**	Professor of Computational Methods and Programming Department of BSUIR
PhD	**Vadim Zhuravliov**	Head of Fundamental Electrical Engineering Department of BSUIR

Секция «Системы управления»

Section «Control Systems»

РАДИОУПРАВЛЯЕМЫЙ КАТЕР ДЛЯ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ НА ВОДЕ

Борис Д. А., Гармашук В. Р.
Факультет компьютерных систем и сетей, факультет компьютерного проектирования,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: samokaterscie438@gmail.com

Данный проект направлен на снижение численности погибших людей на воде с помощью радиоуправляемого устройства – катера. За счет своей высокой скорости, которую он достигает благодаря обтекаемому корпусу и максимально сниженным весом, спасатели смогут оперативно доставить тонущему спасательный круг, что гораздо увеличит шансы на его спасение.

Введение

По оперативным данным ОСВОД за первую половину 2025 года зарегистрировано 213 происшествий на воде, 120 (56%) из которых закончились летальным исходом, среди них 6 несовершеннолетних (5%) [1].

Из данных, приведенных выше, видно, что проблема гибели людей на воде очень актуальна. Ее не решили, и не решат никогда, но люди разработали методы и приспособления, которые снижают вероятность трагедии, и увеличивают шанс на спасение уже тонущего человека.

I. Принципы и тактика обеспечения безопасности на воде

Обеспечение безопасности на воде делится на три этапа:

- Предостережение от несчастного случая;
- Спасение утопающего;
- Оказание первой помощи.

Людей предупреждают запрещающие таблички, буйки, правила безопасностного поведения на воде, но не всех это останавливает. Люди нарушают правила, плавают в состоянии алкогольного опьянения, не следят за детьми, заплывают за буйки, балуются и играются. Даже с человеком, соблюдающим все правила и предосторожности может произойти несчастный случай. Достаточно внезапно ушудшиться самочувствию или возникнуть судорогам, и человек может не вернуться на сушу живым [2].

Если человек потерял контроль, спасателям необходимо как можно быстрее вытащить его из воды. Для этого чаще всего используют спасательные круги (рис. 1) [3] и конец Александрова (рис. 2) [4], либо спасатели сами подплывают на лодке или катере. После этого ему оказывают первую помощь.

Рис. 1 – Спасательны круг – КС-Планета4

Рис. 2 – Конец Александрова

Данные методы используются не один век, они проверены и надежны, но только на небольшом расстоянии. Спасательный круг и конец Александрова не получится забросить дальше 30 метров, а на лодке, даже с мотором, можно не успеть. Проблема гибели людей на водоемах является одной из социально значимых, поэтому мероприятиям предупреждения гибели и спасения людей и вопросам повышения эффективности спасательных работ уделяется постоянное внимание. В связи с этим, разработка новых средств для обеспечения спасательных работ на воде весьма актуальна. Поэтому следующим шагом развития стало оснащение спасательными кругами дронов (рис. 3) [5], и создание наводных радиоуправляемых спасательных кругов (рис. 4) [6]. Они оперативны, что сильно увеличивает шансы на спасение утопающего.

Рис. 3 – Спасательный дрон, разработанный совместно с военным технополисом «Эра» и Российской академией наук

Рис. 4 – Дистанционно управляемый спасательный круг Havospark

II. Идея создания

Данные устройства кардинально совершенствуют технологии спасения тонущих, они на голову выше по скорости и дальности действия стандартных методов, но нужно продолжать совершенствовать их. Сравнения принципов работы обоих устройств показывают, что дрон быстрее доставит спасательный круг к утопающему, но после сброса второй попытки нет, а радиоуправляемый круг может корректировать свою траекторию. Предлагается совместить положительные качества и получить альтернативное устройство с высокой скоростью перемещения и возможностью корректировать траекторию. Таким образом возникла идея создания радиоуправляемого катера для спасения жизни людей на воде.

III. Устройство катера

Радиоуправляемый катер имея хорошую обтекаемость и вытянутую форму развивает очень высокие скорости за короткий период, что дает ему преимущество перед кругом.

Немаловажным фактором для увеличения скорости является вес. В дистанционно-управляемом кругу корпус сам выполняет функцию непотопляемого элемента, что добавляет ему немало веса. В катере предлагается установить баллон сжатого воздуха и надувной круг. В баллон можно будет закачать воздух с помощью компрессора или обычного ручного насоса, это делает устройство пригодным для "полевых"условий, в нашем случае для "пляжных". Воздух из баллона в надувной круг будет запускаться через пульт, с помощью соленоидного, или по другому электромагнитного клапана. Данная технология в разы компактнее радиоуправляемого круга, что позволит разместить ее в обтекаемом корпусе катера.

В основе всей электронной системы предлагается использовать платформу Arduino. Для приведения катера в движение будет использоваться бесколлекторный мотор, а рулевой механизм будет двигать сервопривод.

В результате устройство обеспечит высокую скорость доставки спасательного круга утопающему, что увеличит шансы на его спасение.

IV. Алгоритм работы с катером

- Включить источник питания;
- опустить катер на воду;
- пультом дистанционного управления осуществлять управление, по радиоканалу, катером;
- по прибытию к утопающему, нажать на кнопку, отвечающую за надув спасательного средства;
- если утопающий не схватился за спасательное средство, путем перемещения катера дать возможность утопающему схватиться за спасательное средство;
- дожидаться спасателей.

1. СПРАВКА ОСВОД о несчастных случаях с людьми на водах в первом полугодии 2025 года (по оперативным данным ОСВОД) [Электронный ресурс]/Республиканское государственно-общественное объединение «Белорусское республиканское общество спасания на водах» – Режим доступа: https://osvod.by. – Дата доступа: 19.10.2025.
2. О правилах безопасного поведения на воде ("Минская правда") [Электронный ресурс]/Республиканский центр управления и реагирования на ЧС – Режим доступа: https://erc.mchs.gov.by/novosti/453913/. – Дата доступа: 19.10.2025.
3. Круг спасательный морской (сертиф. РМРС, 4кг);КС-Планета4 [Электронный ресурс]/Альбатрос (Моторов.нет) – Режим доступа: https://motorov.net/product/krug-spasatelnyy-morskoy-sertif-rmrs-4kg. – Дата доступа: 19.10.2025.
4. Конец Александрова 15 метров [Электронный ресурс]/Швартов – Режим доступа: https://www.shvartov.ru/. – Дата доступа: 19.10.2025.
5. В Анапе введут доставку спасательных кругов дронами терпящим бедствие на воде [Электронный ресурс]/Информационное агентство ТАСС. – Режим доступа: https://tass.ru/obschestvo/11364021. – Дата доступа: 19.10.2025.
6. Дистанционно управляемый спасательный круг Havospark [Электронный ресурс] /Havospark. – Режим доступа: https://www.hoverstar.com/ru/. – Дата доступа: 19.10.2025.

МОДЕЛЬ КОМПАКТНОГО ПРЕДСТАВЛЕНИЯ ХАРАКРЕИСТИК ГОЛОСА ДИКТОРА

Захарьев В. А.
Кафедра систем управления
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: zahariev@bsuir.by

В статье представлена архитектура нейросетевой модели для компактного параметрического представления характеристик голоса диктора (CAEM). Данна модель, например, может быть использована для решения задач обнаружения речевой активности и классификации пола диктора, оптимизированная для мобильных устройств и встреиваемых систем.

Введение

Современные мобильные устройства требуют локальной обработки речи без передачи данных в облако, что гарантирует конфиденциальность и низкую задержку. Обнаружение речевой активности (VAD) и определение пола диктора являются базовыми функциями для персонализированных голосовых ассистентов, фильтрации звонков и систем «умного дома» [1]. Традиционные подходы (WebRTC VAD, GMM на MFCC) теряют эффективность в шуме, а облачные решения увеличивают энергопотребление и нарушают приватность [2].

Физиологическая основа разделения по полу – различия в длине голосовых связок и резонансных полостях: у женщин выше основная частота F_0 (180–250 Гц) и смещены форманты, у мужчин – 80–180 Гц [3]. VAD опирается на временную непрерывность речи и спектральную энергию в гармониках. Гибридные модели, объединяющие нейронные сети и DSP, позволяют одновременно решать обе задачи, используя общие признаки. В статье предлагается архитектура с новым блоком компактного представления харакреистик голоса диктора (Compact Audio Embeddings Model – CAEM). Она извлекает 64-мерные эмбеддинги из сырого аудио (0,5 млн параметров). Модель достигает F1-score VAD 95,2% и точности пола 97,5%.

I. Архитектура модели

Архитектура представляет собой глубокую нейронную сеть сквозного (end-to-end) типа с 0,5 млн параметров, который напрямую из сырого сигнала 16 кГц формирует 64-мерные эмбеддинги, несущие информацию о высоте основного тона, формантной структуре, гармоническом составе и временной динамике голоса.

Слои архитектуры нейросетевой модели CAEM включают следующие основные блоки:

- Блок SincConv (16 фильтров, ядро 101) – параметризуемая фильтрация по частоте;
- Блок 1D-CNN (2 слоя: 32→64 фильтра, ядро 5, stride 2, max-pooling);
- Блок Transformer-encoder (1 слой, 4 головы, скрытая размерность 128, FFN 256, dropout 0,1);
- Блок Projection head (128→64, LayerNorm).

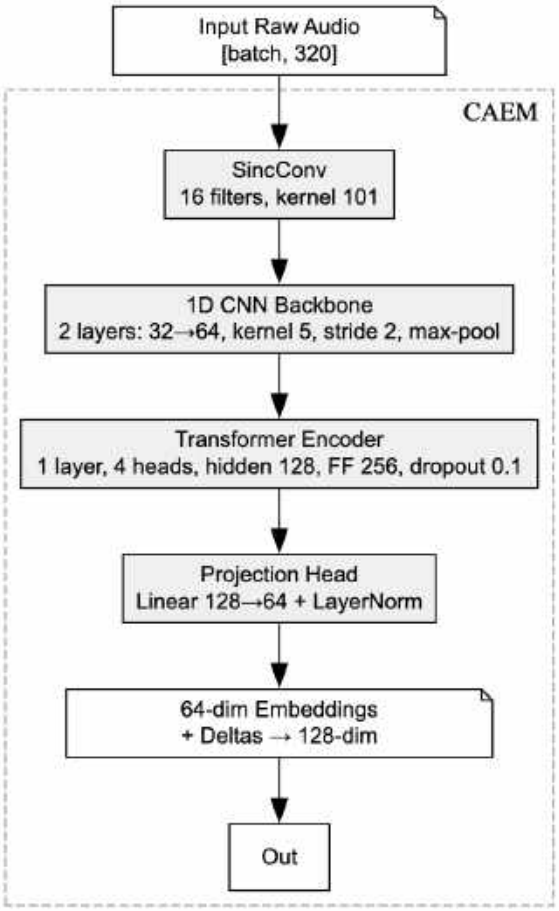


Рис. 1 – Архитектура модели компактного представления харакреистик голоса диктора (CAEM)

Архитектура состоит из четырёх последовательно соединённых блоков, каждый из которых решает строго определённую акустическую задачу и дополняет предыдущий:

Сырой аудиосигнал (320 отсчётов) сразу поступает на блок SincConv – банк из 16 обучаемых полосовых sinc-фильтров. Они автоматически выделяют речезначимые диапазоны (300–800 Гц для мужских голосов, 400–1200 Гц для женских) и

подавляют внеполосный шум без STFT и мел-спектрограмм.

Далее следует двухслойная 1D-CNN (16→32→64 канала, ядро 5, stride 2, max-pooling). Она захватывает локальные временные паттерны – микроструктуру возбуждения, атаку/спад энергии, быстрые формантные переходы – и уменьшает временную размерность в четыре раза, обеспечивая устойчивость к микросдвигам и фазовым искажениям.

Затем признаки подаются в однослойный трансформер-энкодер с четырьмя головами внимания. Самовнимание позволяет каждому шагу учитывать весь контекст окна: одна голова отслеживает основной тон, две другие – формантные области и их динамику, четвёртая подавляет шум. Это значительно точнее кодирует пол диктора и наличие речи по сравнению с чисто свёрточными или рекуррентными моделями той же размерности.

После глобального усреднения по времени 128-мерное представление сжимается линейным слоем до 64 измерений и нормализуется LayerNorm. При необходимости добавляются дельты, получая финальный 128-мерный эмбеддинг, готовый к INT8-квантованию и работе на мобильных устройствах в реальном времени.

II. Результаты экспериментов

Предобучение модели осуществлялось на LibriSpeech (960 ч, masked reconstruction, 15% маскированных кадров, AdamW, lr=10^{-4}, 20 эпох). Дообучение на VoxCeleb1 (определение пола) + NOISEX-92 (VAD) с контрастивной потерей и аугментацией (SNR 5–20 дБ). После INT8-квантования падение точности составило менее 1,0%. Вход: сырой сигнал 16 кГц, окна 20 мс (320 отсчётов), шаг 10 мс. Выход: 64-мерный эмбеддинг на кадр.

Разработана модель в рамках архитектуры верхнего уровняя для решения задач комплексной задачи определения наличия голосовой активности и расспознавания пола диктора по голосу показала следующие результаты на основе тестовой выборки на наборе данных VoxCeleb1:

- F1-score VAD 95,2%, точность пола 97,5% при SNR 5–20 дБ;
- Задержку <15 мс, энергопотребление –15% по сравнению с постоянным соединением к модели размещённой в облаке;
- INT8-квантование с падением точности <1,0%;
- Шумоподавление +13% при отсутсвии блока CAEM.

Таблица 1 – Метрики производительности при различном SNR

SNR (дБ)	5	10	15	20	60
VAD F1	0.942	0.951	0.953	0.957	0.962
Gen Acc	0.968	0.973	0.976	0.979	0.982

Тесты на реальных устройствах (Redmi Note 13, ExecuTorch): задержка 14,8 мс, энергопотребление 0,9 Вт в режиме в режиме потоковой обработки.

III. Выводы

В докладе предложена архитектура нейросетевой модели для компактного параметрического представления характеристик голоса диктора (CAEM). К достоинствам такой комбинации слоёв можно отнести следующие важные её особенности:

1. SincConv + CNN полностью заменяют классическую цепочку STFT → мел-фильтры → лог → DCT, но остаются обучаемыми под конкретные задачи, такие как VAD и классификация пола диктора по голосу.
2. Один слой Transformer обеспечивает глобальный контекст, недоступный чистым свёрточным или рекуррентным сетям того же размера.
3. Общий объём вычислений составляет всего 50 MFLOPs на 20-мс кадр, что позволяет запускать CAEM даже на CPU среднего смартфона в режиме реального времени.
4. Обучение по методу "без учителя"с использованием техники реконструирования маски на 960 часах LibriSpeech заставляет модель выучивать непосредственно инвариантные акустические характеристики речи, а не артефакты конкретной предобработки.

Благодаря предлагаемой последовательности слоёв получаемые 64-мерные эмбеддинги оказываются значительно более информативными и шумоустойчивыми, чем традиционные MFCC, log-mel или даже эмбеддинги более сложных моделей классической архитектуры, что было подтверждено экспериментально.

IV. Список литературы

1. Zhang, A. et al. On-device neural networks for speech processing // Proc. ICASSP. – 2020.
2. Fitch, W. T., Giedd, J. Morphology and development of the human vocal tract // J. Acoust. Soc. Am. – 1999. – Vol. 106, № 1. – P. 632-643.
3. Sadjadi, S. O. et al. Unsupervised speech activity detection using voicing measures // IEEE Trans. Audio Speech Lang. Process. – 2015. – Vol. 23, № 2. – P. 338–348.
4. Ravanelli, M., Bengio, Y. Speaker Recognition from Raw Waveform with SincNet // Proc. SLT. – 2018.
5. Baevski, A. et al. wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations // Proc. NeurIPS. – 2020.
6. Nagrani, A. et al. VoxCeleb: A large-scale speaker identification dataset // Proc. INTERSPEECH. – 2017.

ДИНАМИЧЕСКОЕ МОДЕЛИРОВАНИЕ ОРГАНИЗАЦИОННЫХ СИСТЕМ

Киселевский О. С.
Кафедра «Основы проектирования машин», Белорусско-Российский университет
Могилёв, Республика Беларусь
E-mail: kiselevskioleg@gmail.com

В статье описан метод динамического моделирования инновационных процессов в социально-экономической системе. Динамическое моделирование выполнено на основании математической формализации теории эмоционального труда и нематериального капитала. Обосновывается использование модели для построения систем поддержки принятия решений на производстве.

Введение

Среди объявленных в 2024 году приоритетных направлений научно- технического развития выделено использование искусственного интеллекта в управлении социальными и экономически значимыми системами. Такое управление опирается на разработку импортозамещающих интеллектуальных производственных решений, роботизированных и высокопроизводительных вычислительных систем. В связи с этим Российская академия образования консолидировала усилия на развитии исследований функций человеческого сознания и его математического и технического моделирования.

Количественно формализованные, рациональные категории в оценке взаимодействия человека с производственной средой заложили базу для развития научной школы техногенной рационализации и моделирования человеческого фактора [1]. Роль человеческого фактора в деятельности предприятия пока недооценивается. Его принято ассоциировать с источником ошибок, но никак не с источником инновационного знания. При этом человекоориентированность отечественных научно-технических решений в сфере когнитивного управления организационными системами способна не только в ближайшем времени вывести экономику стран ЕАЭС на опережающие позиции устойчивого развития, но и стать конкурентоспособным информационным товаром [2].

I. Описание динамической модели

Динамическая модель управления инновационной деятельностью предприятий в условиях информационной неопределённости представлена системой дифференциальных уравнений. Уравнения формализуют влияние материальных и нематериальных факторов инновационной деятельности: материального поощрения P, взаимного доверия внутри коллектива R, эмоциональной вовлечённости работников A и их личного неявного знания C.

Эта система решена (Рис. 1) численным интегрированием по методу Рунге-Кутты [3].

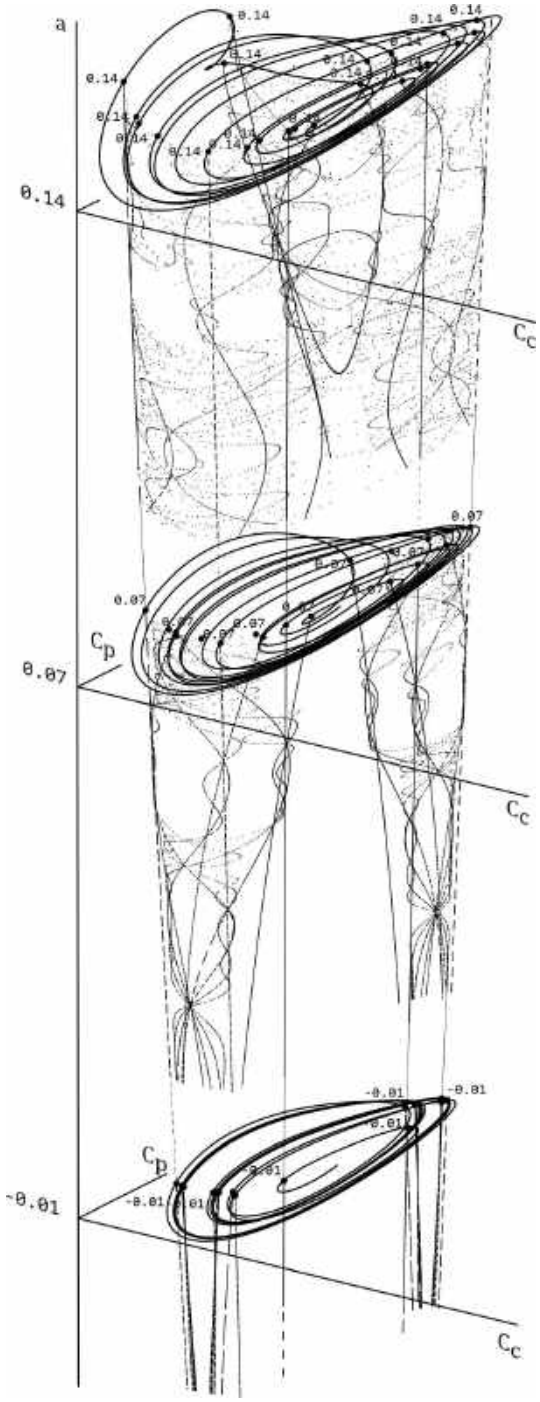


Рис. 1 – Фазовые портреты и бифуркационная диаграмма динамической системы

На рисунке фазовые портреты динамической системы совмещены с бифуркационной диаграммой. Бифуркационная диаграмма визуализирует переходы от стационарного режима автоколебаний в системе к квазистационарному странному аттрактору.

В основе построения модели лежат:

- концепции нематериального культурного [4], социального [4] и эмоционального капитала [5];
- гипотеза о взаимной конвертации и законе сохранения капитала;
- теории эмоционального труда [6] и удовлетворённости им [7];
- многофакторная производственная функция Кобба-Дугласа [8].

Опуская эмпирическое обоснование коэффициентов баланса между параметрами уравнений, а также стандартные математические преобразования, динамическую модель можно описать системой из четырёх уравнений:

$$dC_c/dt = 0.15C_p - 0.35C_c$$

$$dC_p/dt = 0.25A_p - 0.15C_p$$

$$dP_p/dt = 0.25C_c - 0.1C_c dC_c$$

$$dA_p/dt = 0.1C_p + 0.1C_c dC_c - a(\cdot)P_p$$

где C_p – неявное знание сотрудников; C_c – корпоративная культура предприятия; A_p – эмоциональная вовлечённость; P_p – материальное поощрение; a – варьируемый коэффициент.

Бифуркационная диаграмма, составленная из сечений 4-мерного годографа функции гиперплоскостью P_p=0.1 показывает то, как изменяется характер фазового портрета системы при варьировании коэффициента a. На практике этот коэффициент отвечает за соразмерность материального вознаграждения выполненному эмоциональному труду. На фазовой диаграмме видно, что в районе значений a=0.02...0.03 наблюдается бифуркация и динамическая система переходит из режима автоколебаний со стабильным показателем Ляпунова близким к нулю в режим странного аттрактора с гармонически изменяющимся показателем Ляпунова. Такой режим не стационарен. Его принято характеризовать как «детерминированный хаос». Однако именно такой режим отличается самопроизвольными всплесками вовлечённости и инновационной активности персонала, порождающей уникальный инновационный продукт.

II. Заключение

Предложенная динамическая модель не только позволяет анализировать и диагностировать процессы трансформации материальных и нематериальных ресурсов в организационной системе. Более важной её функцией является прогностическая и рекомендательная. Математическое моделирование социально-экономических процессов на предприятии и сопоставление результатов с реальными данными эконометрики и психометрики является инструментом для автоматизации принятия решений (СППР). А сопровождающие математическую модель фрактальные годографы и бифуркационные диаграммы, кроме прочего, производят гипнотический эффект на экономистов-гуманитариев, что также немаловажно в маркетинговом продвижении наукоёмкой идеи.

В целом методика динамического моделирования процессов трансформации социально-эмоциональных ресурсов персонала предприятия в производственную информацию, а затем - в полезный материальный продукт является ключевым звеном инновационной технологии когнитивного менеджмента и сопутствующей ему цифровой технологии ERP-III.

1. Дергачева Е. А., Кузьменко А. А. Диалектика дизайна и эргодизайна в рационализации среды жизнедеятельности техногенного мира // Гуманитарий Юга России. – 2025. – Т. 14, № 2. – С. 131-144.
2. Спасенников В. В., Кузьменко А. А., Киричек А. В. Эргономическая парадигма устойчивости биосферной жизни в контексте эволюции и синергии междисциплинарных исследований // Глобалистика-2023 : Сб. матер. междунар. науч. конгр. – М. : Межрегиональная общественная организация содействия изучению, пропаганде научного наследия Н. Д. Кондратьева, 2024. – С. 363-371.
3. Кнауб Л. В., Лаевский Ю. М., Новиков Е. А. Алгоритм интегрирования переменного порядка и шага на основе явного двухстадийного метода Рунге–Кутты // Сибирский журнал вычислительной математики. – 2007. – Т. 10. – № 2. – С. 177-185.
4. Bourdieu P. Ökonomisches Kapital, kulturelles Kapital, soziales Kapital, Kreckel, Reinhard (Hrsg.) : Soziale Ungleichheiten. Soziale Welt, Sonderheft 2. – 1983. – P. 183-198
5. Gendron B. Mindful management & capital émotionnel: L'humain au c ur d'une performance et d'une économie bienveillantes. Bruxelles, Paris. De Boeck Supérieur. – 2015. – 130 p.
6. Hochschild A. R. The managed heart : commercialization of human feeling / Berkeley : University of California Press, 1983. – 307 p.
7. Locke E. A. Latham G. P. A theory of goal setting and task performance // Academy of Management Review. Prentice Hall. – 1991. – 16(2). – 62 p.
8. Киселевский О. С., Косякова Е. В. Эмоциональный труд как фактор производственной функции Кобба-Дугласа // Фундаментальные и прикладные исследования в области управления, экономики и торговли : Сб. тр. всерос. науч.-практ. конф. СПб : Политехпресс. – 2024. – р. 113-121.

МНОГОФУНКЦИОНАЛЬНАЯ РОБОТОТЕХНИЧЕСКАЯ ПЛАТФОРМА С АДАПТИВНЫМИ РЕЖИМАМИ УПРАВЛЕНИЯ

Климашевич З. Д., Клепацкий Е. Э., Линкевич А. О., Болабсов Р. А.
Факультет компьютерного проектирования,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: zahareshnik@gmail.com

Проект посвящен созданию многофункциональной робототехнической платформы с адаптивными режимами управления. Основная особенность платформы – наличие модуля выжигания сорных растений, управляемого с помощью микроконтроллера Arduino Uno и Raspberry Pi. Система сочетает автономную и дистанционную работу, применяя датчики расстояния, света и пламени, а также модуль камеры для анализа окружающей среды. Реализованы алгоритмы движения, позиционирования и автоматического наведения лазера. Результаты проекта могут применяться в образовательной робототехнике и сельскохозяйственных разработках.

Введение

Лазерное уничтожение сорных растений в полевых условиях имеет значительную актуальность по нескольким причинам. Прежде всего, сорные растения являются серьезной проблемой для сельского хозяйства, так как они конкурируют с посевами за ресурсы и могут значительно снизить урожайность. Использование химических гербицидов негативно сказывается на окружающей среде и здоровье людей. Лазерное уничтожение представляет собой экологически чистую и эффективную альтернативу для борьбы с этой проблемой. Оно не загрязняет почву, воду и воздух и сокращает использование химических средств, что положительно сказывается на здоровье людей и окружающей среды. Кроме того, лазеры могут быть точно настроены на определенную длину волны, что позволяет уничтожать сорные растения без повреждения полезных культур. Они также могут быть использованы для борьбы с сорными растениями в труднодоступных местах и можно программировать для автоматической работы, что экономит время и силы работников сферы сельского хозяйства. В данном проекте разработана робототехническая платформа, предназначенная для автономного выжигания сорных растений с помощью лазерного модуля.

I. Аппаратная часть робототехнической платформы

Аппаратная часть робототехнической платформы построена на микроконтроллере ATmega328 [3] и представлена на рисунке 1. Она обеспечивает вычислительную мощность и контроль за внешними устройствами. Плата имеет 14 цифровых выводов, из которых 6 могут работать как ШИМ-выходы. Есть 6 аналоговых входов для считывания данных с датчиков. В проекте Arduino Uno используется для движения и позиционирования платформы.

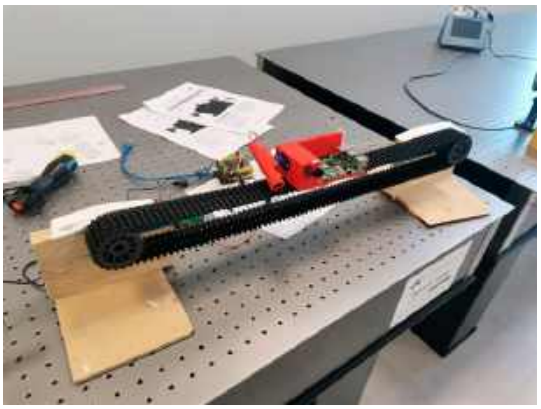


Рис. 1 – Внешний вид прототипа платформы

Для управления моторами используется драйвер Keyestudio L298P DC Motor Driver Shield [4]. Моторы подключаются к пинам 3, 11, 12 и 13, где пины 12 и 13 задают направление, а пины 3 и 11 регулируют скорость посредством ШИМ-сигнала. Кнопка управления подключена к пинам GND и 4, а сервопривод SG90 – к пинам 5V, GND и GPIO6. Питание осуществляется от батареи 9V, обеспечивающей достаточный ток для всех компонентов. Общая схема подключения представлена на рисунке 2.

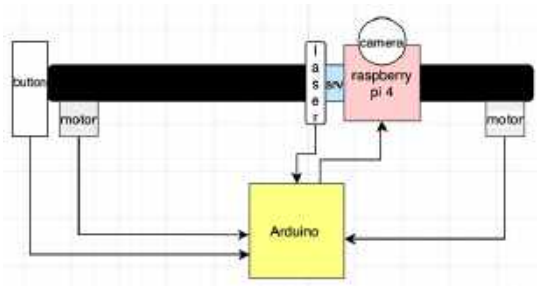


Рис. 2 – Общая схема подключения

На платформе установлены ультразвуковые и инфракрасные датчики, датчики света и пламени, модуль Raspberry Pi 4 с камерой и лазерный модуль выжигания сорных растений [1]. 3D-

модель модуля выжигания представлена на рисунке 3. Raspberry Pi выполняет анализ изображений, определяет местоположение сорняка и передаёт координаты Arduino, которая управляет сервоприводом и лазером. Таким образом, система работает в автоматическом режиме без участия оператора.

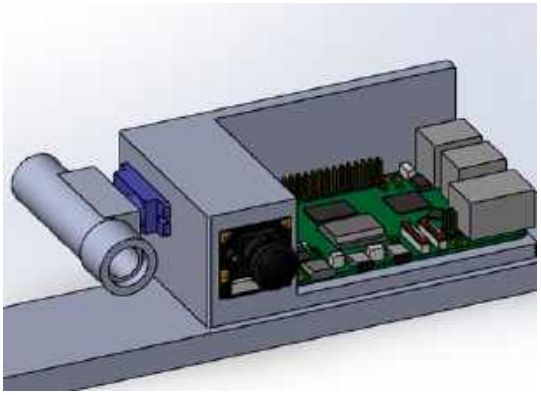


Рис. 3 – 3D-модель модуля выжигания сорных растений

II. Программная архитерктура робототехнического комплекса

Raspberry Pi 4 работает как любой другой компьютер, основанный на ARM-процессоре. Он использует операционную систему (например, Raspbian), которая управляет всеми компонентами устройства, включая процессор, память, ввод-вывод и т. д. Пользователь может подключать монитор, клавиатуру и мышь, чтобы взаимодействовать с Raspberry Pi 4, а также использовать различные порты для подключения устройств, таких как флэш-накопители, камеры, дисплеи и другие периферийные устройства. К контролирующей плате Raspberry Pi 4 подключили модуль камеры и лазерный модуль в соответствующие пины. Программное обеспечение реализовано на языке C++ [1] и Python. Raspberry Pi анализирует изображения, используя алгоритмы компьютерного зрения, а Arduino управляет двигателями, сервоприводами и лазером. Взаимодействие между микроконтроллерами осуществляется через последовательный порт. Основная логика программы включает следующие этапы:

1. Считывание данных с датчиков и камеры;
2. Анализ изображений для поиска сорняков;
3. Наведение лазера и включение модуля выжигания;
4. Коррекция траектории движения платформы при обнаружении препятствий;
5. Отправка отчётов о выполненных действиях.

III. Практическая значимость и перспективы

Лазерное уничтожение сорных растений в полевых условиях является эффективным и перспективным методом борьбы с нежелательными растениями. Эта технология позволяет точно и безопасно уничтожать сорные растения, не нанося вреда полезным культурам. Кроме того, данный метод экологически безопасен и не имеет негативного воздействия на почву и окружающую среду. В результате применения лазерного уничтожения сорных растений, улучшается качество урожая, снижается нужда в применении химических гербицидов, а также сокращается затрата времени и ресурсов на удаление сорняков. В целом, лазерное уничтожение сорных растений представляет собой инновационный метод, который может быть эффективно применен в сельском хозяйстве и садоводстве для повышения урожайности и сокращения вреда от сорных растений. Предлагаемый прототип системы сочетает образовательный и прикладной потенциал. Он может использоваться как учебный стенд по робототехнике и системам технического зрения, а также как прототип автоматизированного устройства для уничтожения сорных растений. В дальнейшем планируется внедрение алгоритмов машинного обучения для распознавания различных типов растений и улучшения точности наведения.

IV. Заключение

Предложена многофункциональная робототехническая платформа с адаптивными режимами управления. Система демонстрирует высокую точность позиционирования и устойчивость к внешним воздействиям. В перспективе планируется расширение функционала за счёт интеграции интеллектуальных алгоритмов и оптимизации энергопотребления.

V. Список литературы

1. Шилдт Б. Программирование на C++. – М. : Вильямс, 2015. – 864 с.
2. Dewdney A. K. The Armchair Universe: An Exploration of Computer Worlds. – New York: W. H. Freeman, 1990. – 240 p.
3. Arduino. Arduino UNO Rev3 [Электронный ресурс]. – Режим доступа: `https://www.arduino.cc/`, свободный.
4. Keyestudio. L298P DC Motor Driver Shield [Электронный ресурс]. – Режим доступа: `https://wiki.keyestudio.com/`.

РОБАСТНАЯ ВЕРИФИКАЦИЯ ДИКТОРА В ЗАШУМЛЁННЫХ УСЛОВИЯХ НА ОСНОВЕ X-VECTORS

Крищенович В. А., Захарьев В. А.
Кафедра систем управления, Кафедра интеллектуальных информационных технологий
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {krish, zahariev}@bsuir.by

В работе предложен метод повышения робастности системы верификации диктора путём интеграции модели голоса на основе x-vectors с шумоподавлением в модифицированной архитектуре LSTM с блоком самовнимания. Это обеспечивает устойчивость к фоновому шуму, достигая F1-score выше 0,94 при SNR от 0 до 20 дБ и точность классификации 95% после 8-битной квантования весов с падением менее 1%.

Введение

Современные системы биометрической верификации диктора активно применяются в задачах управления доступом, банковской аутентификации, криминалистической экспертизе и мобильных устройствах. Однако в реальных акустических условиях, таких как улицы, транспорт, офисы или производственные помещения, речевой сигнал неизбежно искажается аддитивным и конволютивным шумом, реверберацией и перекрывающимися голосами. Это приводит к резкому падению точности систем верификации.

В работе представлен архитектурный подход, объединяющий преимущества x-vectors и блоков "внимания"нейронной сети для задачи шумоподавления в рамках модифицированной рекуррентной архитектуры с самовниманием [1]. Предлагаемый метод позволяет не только эффективно подавлять шум, но и адаптивно усиливать дикторски-информативные компоненты сигнала, сохраняя устойчивость при квантовании модели для внедрения на встраиваемые устройства [2, 3].

I. Предлагаемая архитектура

Извлечение x-vectors осуществляется на основе TDNN-архитектуры, предобученной на датасете VoxCeleb2 [4]. Входной сигнал преобразуется в 80-мерные лог-мел-спектрограммы с окном 25 мс и шагом 10 мс. TDNN-сеть включает 5 временных контекстных слоёв с расширяющимися дилатациями (1, 2, 3, 4, 1), статистический pooling-слой (mean + std) и два полносвязных слоя с активацией ReLU, формирующих 512-мерный x-vector. Финальная классификация выполняется с помощью PLDA, обученного на чистых эмбеддингах.

Ключевое нововведение – модуль шумоподваления на основе механизма самовнимания, встроенный перед TDNN. Зашумлённые мел-спектрограммы подаются в модифицированную двунаправленную LSTM (BiLSTM) с интегрированными слоями самовнимания. Каждый рекуррентный блок содержит 512 скрытых юнитов в прямом и обратном направлениях, за которым следует блок многоголовое самовнимания (multi-head self-attention) с 8 головами и размерностью ключа/запроса 64. Механизм self-attention вычисляет матрицу внимания по формуле:

$$\alpha_{ij} = \frac{\exp\left(\frac{Q_i K_j^\top}{\sqrt{d_k}}\right)}{\sum_{k=1}^{T} \exp\left(\frac{Q_i K_k^\top}{\sqrt{d_k}}\right)},$$

где Q_i, K_j, V_j – проекции скрытых состояний через линейные преобразования. Взвешенная сумма значений V_j формирует контекстно-усиленный вектор, передаваемый на следующий слой. Такая структура позволяет модели динамически фокусироваться на речевых гармониках, формантных переходах и подавлять шумовые компоненты в частотно-временной области.

Для стабилизации обучения используется multi-head self-attention с residual connections, layer normalization и dropout ($p = 0.1$). После каждого attention-блока добавляется feed-forward сеть с GELU-активацией. Всего в модуле шумоподавления – 3 последовательных BiLSTM + self-attention блока. Выход модуля шумоподавления – очищенная спектрограмма – нормализуется (дисперсии глобальной средней) и подаётся в TDNN.

Общая функция потерь:

$$L = L_{\text{CE}}(y, \hat{y}) + \lambda_1 L_{\text{denoise}} + \lambda_2 L_{\text{reg}} + \lambda_3 L_{\text{contrast}},$$

где L_{CE} – кросс-энтропия верификации, L_{denoise} – MSE между очищенным и эталонным чистым сигналом, L_{reg} – L2-регуляризация, L_{contrast} – контрастивная потеря для разделения дикторов в пространстве эмбеддингов. Коэффициенты $\lambda_1 = 0.5$, $\lambda_2 = 10^{-4}$, $\lambda_3 = 0.1$ подобраны эмпирически.

Квантование весов до 8 бит выполняется методом квантования после обучения с калибровочным набором из 1000 зашумлённых записей (SNR 0–20 дБ). Используется симметричное квантование с по тензенрным масштабированием. Архитектура реализована в PyTorch с библиотеками torch-quant, SpeechBrain и Kaldi. Общее количество параметров – 24,6 млн (denoising – 18,2 млн, TDNN – 6,4 млн).

II. Результаты экспериментов

Эксперименты проведены на тестовых наборах VoxCeleb1-O и VoxCeleb1-E (37 720 пар) с добавлением шумов из корпуса MUSAN (babble, music, noise) и реверберации из RIRs при SNR от 0 до 20 дБ. Базовая система x-vectors (без шумоподавления) показывает EER 12,4% при SNR 5 дБ и 8,1% в чистых условиях. Предлагаемый метод снижает EER до 4,7% при SNR 5 дБ и 2,9% в чистых условиях – улучшение на 62% и 64% соответственно.

F1-score верификации в зашумлённых условиях представлен далее. После квантования: точность классификации – 95,2%, падение 0,8% по сравнению с FP32. Средний F1-score: 0.945. Дополнительно оценена робастность к перекрывающимся голосам (two-speaker interference): при SNR=0 дБ с интерферирующим диктором EER = 7,3% против 19,8% у базовой модели. При добавлении музыкального шума (поп, рок музыка) – EER = 5,1% при SNR=10 дБ.

Сравнение архитектурных подходов

- WavLM + PLDA – EER 6,2% (SNR=5 дБ);
- ECAPA-TDNN + SE-blocks – EER 5,9%;
- RawNet3 + denoising – EER 5,4%.

Предлагаемый метод превосходит их при низком SNR и смешанных шумах.

Проведены дополнительные тесты на мобильных устройствах (Android-смартфон с процессором Snapdragon 8 Gen 1): время инференса 210 мс, энергопотребление 1,2 Вт, точность сохраняется на уровне 94,8% (падение 0,4%).

III. Выводы

Разработанный методов интеграции x-vectors с attention-based шумоподавлением в модифицированной BiLSTM с блоками самовнимания обеспечивает высокую робастность верификации диктора в широком диапазоне зашумлённых условий. Достигнута точность классификации 95,2% после 8-битного квантования с падением менее 1%, EER 4,7% при SNR=5 дБ и F1-score 0,945 в среднем – лучшие показатели среди аналогов при сопоставимой вычислительной сложности.

Предложен подход на основе совместной оптимизации шумоподваления и верификации с блоками самовнимания в рекуррентной архитектуре, устойчивой к квантованию. Метод пригоден для внедрения на мобильных устройствах таких как смартфоны и планшеты.

Дальнейшее развитие включает:

- адаптацию блока самовнимания для разных видов шумов и многопользовательских сценариев;
- интеграцию со сквозными моделями (ECAPA-TDNN, RawNet);
- федеративное обучение на децентрализованных данных;
- исследование 4-битного квантования с сохранением точности.

Планируется публикация модели в открытом доступе и тестирование в реальных системах безопасности.

IV. Список литературы

1. Chen, S. et al. WavLM: Large-Scale Self-Supervised Pre-Training for Full Stack Speech Processing // IEEE JSEL. – 2022. – Vol. 16. – P. 123-135.
2. Vaswani, A. et al. Attention Is All You Need // NeurIPS. – 2017. – P. 5998-6008.
3. Snyder, D. et al. Deep Neural Network Embeddings for Text-Independent Speaker Verification // Interspeech. – 2017. – P. 999-1003.
4. Nagrani, A. et al. VoxCeleb: Large-Scale Speaker Verification Dataset // Interspeech. – 2017. – P. 123-135.

Таблица 1 – Результаты F1-score верификации при различном SNR после квантования

SNR (дБ)	0	5	10	15	20	чистый	среднее	падение
F1-score	0.94	0.95	0.93	0.96	0.94	0.96	0.945	0.8%

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ КОМАНДНОЙ РАБОТЫ В РАСПРЕДЕЛЁННЫХ ПРОЕКТАХ ЗА СЧЁТ ИНТЕГРАЦИИ АДАПТИВНОЙ KANBAN-ДОСКИ С МЕХАНИЗМАМИ РЕАЛЬНОГО ВРЕМЕНИ И ГИБКОЙ НАСТРОЙКИ РАБОЧИХ ПРОЦЕССОВ

Нестеренков С. Н., Шмаргун А. Э.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by, al1402004@gmail.com

Рассматривается подход к повышению эффективности командной работы в распределённых проектах через внедрение адаптивной Kanban-доски с поддержкой обновлений в реальном времени и гибкой настройкой workflow. Анализируются ключевые преимущества, архитектурные особенности и практическое применение решения для современных распределённых команд.

Введение

Распределённые команды сталкиваются с проблемами синхронизации, прозрачности и гибкости управления задачами. Традиционные инструменты (Trello, Jira, Asana) обеспечивают лишь базовую визуализацию и не отвечают требованиям динамичных проектов, где процессы постоянно меняются. Особенно критичны задержки в обновлении статусов и отсутствие кастомизации workflow.

Предлагаемая адаптивная Kanban-доска решает эти проблемы за счёт трёх компонентов: (1) настраиваемого workflow под специфику проекта, (2) синхронизации в реальном времени и (3) глубокой интеграции с экосистемой разработки. Это позволяет не просто отслеживать, но и активно управлять потоком работы, снижая издержки и повышая производительность [2].

I. Ключевые аспекты применения

Прозрачность и синхронизация в реальном времени

Одной из главных проблем распределённых команд является отсутствие единой «картины дня». Участники часто работают с устаревшими данными, что приводит к дублированию усилий, конфликтам и снижению доверия внутри коллектива. Реализация механизма обновления через WebSocket-соединения позволяет мгновенно отражать любые изменения на доске у всех участников, независимо от их местоположения. Это устраняет необходимость в ежедневных статус-митингах, которые в глобальных командах зачастую проводятся в неудобное время для части участников. Более того, система поддерживает историю изменений с возможностью отката, что повышает надёжность и предсказуемость рабочего процесса.

Гибкость и адаптивность workflow

Каждый проект уникален: одни команды используют строгий Scrum, другие — гибридный Kanban-Scrum, третьи — собственные методологии. Адаптивная Kanban-доска позволяет настраивать не только количество и названия колонок, но и логику переходов между ними. Например, можно задать правило: «если задача находится в статусе «На проверке» более 24 часов, автоматически отправить уведомление ответственному тестировщику». Также поддерживаются WIP-лимиты (Work In Progress), которые помогают избежать перегрузки отдельных участников и выявлять узкие места в процессе. Возможность сохранять и применять шаблоны workflow для разных типов проектов (например, «баг-фикс», «новая фича», «маркетинговая кампания») значительно ускоряет старт новых инициатив [1].

Снижение когнитивной нагрузки

Информационная перегрузка — частая проблема в современных системах управления проектами. Адаптивная доска решает её за счёт персонализации интерфейса. Каждый участник может настроить фильтры по тегам, приоритетам, исполнителям или типам задач. Поддерживается система подписок: пользователь получает уведомления только по тем задачам, которые его непосредственно касаются.

II. Архитектура и технологический стек

Архитектура системы построена по принципам микросервисов и событийно-ориентированного взаимодействия [3, 4]. Клиентская часть реализована с использованием современных фреймворков – React или Vue.js – с применением библиотек для управления состоянием (Redux или Pinia). Для обеспечения реального времени используется WebSocket-протокол в связке с сервером на базе Node.js или Python (FastAPI).

Серверная логика разделена на модули: управление пользователями, workflow-движок, система уведомлений, аналитика и интеграции. Данные хранятся в гибридной модели: структурированная информация (пользователи, проекты) – в PostgreSQL, а гибкие настройки workflow и история изменений – в MongoDB. Все операции логируются, а доступ к данным регулируется на основе ролевой модели (RBAC).

Особое внимание уделено работе в условиях нестабильного соединения: клиентская часть поддерживает офлайн-режим с локальным кэшированием изменений и последующей бесконфликтной синхронизацией при восстановлении связи.

III. Практическое применение

Адаптивная Kanban-доска успешно применяется в различных отраслях:

- В «IT-разработке» она интегрируется с Git-платформами (GitHub, GitLab), автоматически создавая задачи при открытии pull request'ов и обновляя статусы при мерже. Это устраняет ручной ввод и снижает риск рассогласованности между кодом и трекером задач [1].
- В «маркетинговых агентствах» доска используется для планирования контент-календарей: каждая публикация — отдельная карточка с этапами «идея», «контент», «редактура», «публикация», «аналитика».
- В «образовательных проектах» преподаватели распределяют задания, отслеживают прогресс студентов и дают обратную связь прямо в карточках.
- В «продуктовых командах» доска помогает визуализировать roadmap: эпик-задачи разбиваются на пользовательские истории, которые проходят полный цикл от идеи до релиза.

По данным пилотного внедрения в трёх компаниях (IT-стартап, маркетинговое агентство, университетский исследовательский центр), среднее время выполнения задач сократилось на 27%, а количество встреч по согласованию снизилось на 40%.

IV. Преимущества и ограничения

Преимущества адаптивной Kanban-доски перед классическими решениями очевидны. Во-первых, скорость реакции системы: изменения отображаются менее чем за 100 мс, тогда как в Jira или Trello задержка может достигать нескольких секунд. Во-вторых, глубина кастомизации: пользователи не ограничены шаблонами, а могут проектировать workflow «с нуля». В-третьих, система поддерживает сложные сценарии автоматизации, недоступные в стандартных инструментах без дорогостоящих плагинов [7].

Однако существуют и ограничения. Для стабильной работы в реальном времени требуется минимальная пропускная способность канала (не менее 1 Мбит/с на пользователя). Кроме того, внедрение требует определённого уровня цифровой зрелости команды: участники должны понимать основы Kanban и быть готовы к переходу от пассивного отслеживания к активному управлению потоком [2].

Заключение

Адаптивная Kanban-доска с поддержкой реального времени переосмысливает взаимодействие в распределённых командах – от статичного трекера к динамической среде управления потоком работы. Она повышает прозрачность, снижает коммуникационные издержки и позволяет быстрее реагировать на изменения.

Будущие направления включают интеграцию ИИ для прогнозирования сроков и автоматической оптимизации workflow, а также режим совместного редактирования карточек в реальном времени – особенно полезный для удалённых мозговых штурмов и планирования. Перспективным представляется использование архитектур, основанных на Retrieval-Augmented Generation (RAG), которые позволяют ИИ-ассистентам опираться на корпоративные базы знаний, обеспечивая достоверность рекомендаций и снижая риск генерации ложной информации [7].

1. Такунов, П. А. Big Data в веб-приложении по контролю выполненной работы в команде / П. А. Такунов, С. Н. Нестеренков, А. Н. Марков // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. работ IX Междунар. науч.-практ. конф. В 2 ч. Ч. 2 (Республика Беларусь, Минск, 17–18 мая 2023 года) / редкол. : В. А. Богуш [и др.]. – Минск : БГУИР, 2023. – С. 67–71.
2. Андерсон, Д. Дж. Kanban: Успешные эволюционные изменения в вашем технологическом бизнесе. – М. : ДМК Пресс, 2023. – 272 с.
3. Ньюман, С. От монолита к микросервисам: эволюционные шаблоны трансформации. – М. : Питер, 2024. – 260 с.
4. Вернон, В. Практическая архитектура событийно-ориентированных микросервисов. – М. : ДМК Пресс, 2024. – 520 с.
5. Рубин, К. С. Essential Scrum: Практическое руководство по самому популярному гибкому процессу. – М. : Вильямс, 2023. – 528 с.
6. Швабер, К., Сазерленд, Дж. Scrum Guide: Официальное руководство по Scrum (обновление 2023). – М. : Изд-во Scrum.org, 2023. – 16 с.
7. Баяк, Е. И. Искусственный интеллект в управлении проектами для создания адаптивных стратегий / Е. И. Баяк, С. Н. Нестеренков, Д. А. Жалейко // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. работ X Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13 марта 2024 года). В 2 ч. Ч. 1 / редкол. : В. А. Богуш [и др.]. – Минск : БГУИР, 2024. – С. 316–320.

ЭЛЕМЕНТ СИСТЕМЫ МОНИТОРИНГА ПРОМЫШЛЕННЫХ ПРОИЗВОДСТВ

Русакович Е. С., Деменковец Д. В.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: smegor55@gmail.com, Demenkovets@bsuir.by

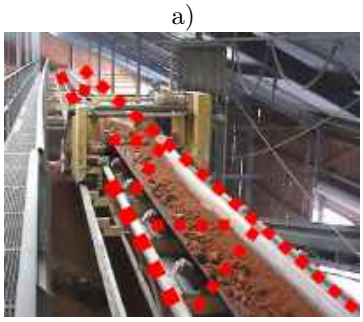
Рассматривается возможная реализация элемента составной части индустриальной системы мониторинга, контролирующего работу технологического оборудования или перемещение людей в запрещённых зонах с помощью видеокамеры низкого разрешения. Обоснован выбор вычислительного модуля элемента, предложен эффективный алгоритм удаления фона изображения, полученного из видеопотока.

Введение

Промышленные производства требуют внедрения роботизированных технологий, компьютерного зрения и систем безопасности персонала.

Разрабатываются индустриальные системы мониторинга для контроля технологических параметров, детекции сбоев и нарушений работы оборудования. Эти системы в частности применяются для соблюдения правил техники безопасности, охраны труда и контроля движения в специальных зонах.

а)



б)

Рис. 1 – Примеры контролируемых производственных пространств

Например, на некоторых производствах рабочие выполняют операции на специально оборудованных постах, выход на линию движения конвейера не допускается; на загрузочных питателях в горнодобывающих и перерабатывающих производствах, не должно образовываться заторов материала в связи с остановкой конвейерного оборудования. Отслеживается контроль и безопасность рабочего персонала в производственных ангарах и цехах. Эти события могут детектироваться автоматически с передачей информации по коммуникационным каналам в общую систему управления или охраны предприятия.

I. Аппаратная платформа

Ранние системы мониторинга и безопасности включали в себя различные датчики положения, световые барьеры, ограждения, емкостные датчики присутствия. В современных системах контроля перемещений источником данных является видеокамера. Полученные изображения должны быть должным образом обработаны с целью классификации событий.

Возникает необходимость контролировать различные технологические производства, будь то цеха переработки, выработки полезных ископаемых, роботизированные сборочные линии и другие. Выполнение обработки видеопотока в этих условиях на персональных компьютерах или выделенных серверах требует значительных вычислительных мощностей оборудования. Целесообразным является реализация алгоритмов обработки и распознавания непосредственно в встроенном программном обеспечении видеокамеры, которая будет передавать сигнал о произошедшем событии по выделенному каналу связи (проводному либо беспроводному).

В качестве камеры совмещённой с вычислительным модулем предлагается использовать плату ESP32-CAM [1], включающую:

- 2 микропроцессора Xtensa LX6 (производительность до 240 МГц);
- модули Bluetooth и Wi-Fi;
- интерфейсы SPI, UART, I2C;
- 4 МБ встроенной флэш-памяти;
- 512 КБ оперативной памяти;
- интерфейс SD-карты памяти;
- сенсор камеры в UXGA разрешении.

Плата включает множество различный интерфейсов, обладает достаточной вычислительной мощностью и может быть интегрирована в компактное устройство.

II. Программная часть

Программная реализация системы контроля включает следующую последовательность модулей:

- взаимодействие с видеокамерой;
- обработка кадров;
- детектирование событий;
- отправка сообщений по каналам связи.

Работа этих модулей в системе, логически ассинхронная, реализуется параллельно для эффективного использования доступных вычислительных ресурсов. Для этого встроенная исполняемая программа работает в рамках операционной системы реального времени, которая распределяет процессорное время между модулями. Используется FreeRTOS – операционная система реального времени с поддержкой многозадачности и механизмами синхронизации [2].

III. Обработка кадров

Наиболее трудоёмкая задача – обработка кадров видеопотока. Заключается в выделении объектов расположенных в кадре. В данном случае задача сводится к подзадаче отделения статичного фона от других объектов. Фон изображения принимается за неизменный, так как система предназначена для использования в цехах, рабочих помещениях, где фон значительно не изменяется за все время контроля. В задаче отделения фона выделяют следующие этапы:

1. Получение кадра.
2. Предобработка кадра.
3. Удаление фона.
4. Постобработка кадра.

Алгоритмы удаления фона отличаются по следующим параметрам: затраты памяти, время исполнения, результативность в усложнённых условиях. К усложнённым условиям [3] относятся:

- остановка объекта;
- смена освещения;
- наличие теней;
- динамический фон.

Для мониторинга промышленных производств смена освещения и присутствие теней менее актуальны, чем динамический фон или остановка объекта.

Согласно исследованию [4] наиболее эффективными алгоритмами с точки зрения всех ограничивающих условий встроенной системы являются: вычитание фона с адаптивным обновлением и Гауссова модель. Статистические методы (такие как Гауссова модель) относятся к параметрическим методам, в них для каждого пикселя изображения необходимо сохранять и обновлять дополнительные параметры (математическое ожидание и среднеквадратичное отклонение), что увиличивает расход памяти для выполения методов. В ESP32-CAM предусмотренно только 512 КБ встроенной оперативной памяти и это ограничение является самым критичным. Наиболее ресурсоёмким и эффективным методом в данном случае будет вычитание фона с адаптивным обновлением фоновой модели. Данный алгоритм представлен на рисунке 2.

IV. Постобработка кадров

Постобработка кадров выделяет из множества полученных пикселей изображения отдельные объекты. Она включает в себя морфологические операции, фильтрацию скоплений точек незначительного размера, разбиение на объекты, получение контуров и нахождение центроидов. Путём анализа полученных параметров объектов за промежуток времени можно сделать вывод о произошедших событиях. Например, анализируя контуры объектов возможно детектировать пересечение запрещённых зон, отслеживая центроиды и контуры можно судить о характере взаимодействия объектов.

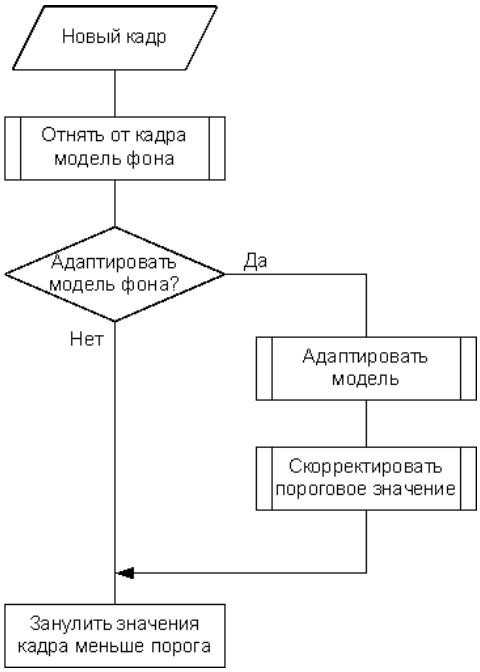


Рис. 2 – Алгоритм вычитания фона

Заключение

Предложен алгоритм удаления фона, являющийся основным системе мониторинга, влияющий на качество работы ситстемы в целом. Элемент системы мониторинга может быть внедрён для контроля перемещений в выделенных зонах, мониторинга соблюдения норм техники безопасности и охраны труда.

1. EPS32 [Электронный ресурс]. – Режим доступа: `https://www.espressif.com/en/products/socs/esp32`. – Дата доступа: 21.10.2025.
2. FreeRTOS [Электронный ресурс]. – Режим доступа: `https://www.freertos.org/`. – Дата доступа: 21.10.2025.
3. Vacavant, A. A Benchmark Dataset for Outdoor Foreground/Background Extraction / A. Vacavant, T. Chateau, A. Wilhelm, L. Lequièvre // Asian Conference on Computer Vision. – 2012. – P. 291–300.
4. Sobral, A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos / A. Sobral, A. Vacavant // Computer Vision and Image Understanding. – 2014. – Vol. 122. – P. 4–21.

СИСТЕМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ УПРАВЛЕНИЯ ПРОЦЕССОМ РИФОРМИНГА СЫРЬЕВОЙ НАФТЫ

Сержанов А. К.
Факультет информационных технологий, кафедра системного анализа и упраления
Евразийский национальный университет им. Л.Н. Гумилёва
г. Астана, Республика Казахстан
E-mail: armat.serzhanov@gmail.com

Целью этой статьи является обзор различных инструментов поддержки принятия решений, которые были разработаны для оказания помощи инженерам в принятии обоснованных решений относительно процесса каталитического риформинга. Эти инструменты предназначены для предоставления информации в режиме реального времени, прогнозирования будущих тенденций и рекомендации действий по оптимизации процесса. В статье подчеркивается важность инструментов поддержки принятия решений в нефтеперерабатывающей промышленности и то, как они могут помочь повысить эффективность процесса каталитического риформинга. Статья будет интересна инженерам-технологам и всем, кто работает в нефтеперерабатывающей промышленности и хочет улучшить свое понимание инструментов поддержки принятия решений, доступных для управления процессом каталитического риформинга.

Введение

Каталитический риформинг является важным процессом в нефтеперерабатывающей промышленности, поскольку он отвечает за превращение низкосортной нафты в высокооктановый бензин и другое ценное нефтехимическое сырье. Процесс сложный и включает в себя ряд реакций, включая дегидрирование, изомеризацию и крекинг, которые протекают при высоких температурах и давлениях в присутствии катализатора. Несмотря на свою важность, процессом каталитического риформинга сложно управлять, поскольку он требует сбалансирования нескольких конкурирующих целей, включая качество продукта, эффективность процесса и срок службы катализатора.[1]

Существует несколько текущих проблем, связанных с каталитическим риформингом в нефтеперерабатывающей промышленности:

Дезактивация катализатора. Одной из самых больших проблем каталитического риформинга является дезактивация катализатора с течением времени, что снижает его эффективность. Это может быть вызвано такими факторами, как отравление, спекание и закупорка пор.

Неэффективность процесса. Сложный характер процесса каталитического риформинга может привести к неэффективности процесса, что может привести к снижению производительности, увеличению потребления энергии и снижению продукта.

Эти текущие проблемы, связанные с каталитическим риформингом, подчеркивают необходимость эффективных инструментов поддержки принятия решений, которые помогли бы руководителям оптимизировать процесс и смягчить эти проблемы. Используя эти инструменты, инженеры могут повысить эффективность процесса, увеличить производительность, снизить затраты и свести к минимуму воздействие процесса каталитического риформинга на окружающую среду.

I. Основная часть

Системы управления катализатором – это специализированные инструменты, предназначенные для управления катализатором в процессе каталитического риформинга. Эти системы предоставляют информацию о состоянии катализатора, включая его активность, селективность и срок службы, и рекомендуют действия по оптимизации характеристик катализатора. Например, системы управления катализатором могут использоваться для мониторинга старения катализатора и прогнозирования того, когда его потребуется заменить, или для оптимизации графика замены катализатора, чтобы свести к минимуму время простоя и увеличить срок службы катализатора.

Например, алгоритмы машинного обучения могут быть использованы для прогнозирования старения катализатора и оптимального времени для замены. Эта информация может быть использована для минимизации времени простоя, необходимого для замены катализатора, и увеличения срока службы катализатора, что приводит к повышению эффективности процесса и качества продукции.

Анализ данных – это быстро развивающаяся область поддержки принятия решений для управления процессом каталитического риформинга. Эти инструменты используют алгоритмы машинного обучения для анализа больших объемов данных процесса, выявления закономерностей и тенденций и составления прогнозов относительно будущего поведения процесса. Например, аналитика данных может использоваться для прогнозирования будущей производительности процесса, выявления недостающих мест процесса и рекомендации действий по оптимизации процесса.

СППР можно использовать для оптимизации стратегии регулирования температуры с учетом множества технологических параметров, таких как состав сырья, активность катализатора и давление в реакторе. Используя алгоритмы машинного обучения для анализа исторических данных и измерений процесса в режиме реального времени, СППР может предоставить операторам полезную информацию и рекомендации по настройке температуры реактора и других переменных для поддержания оптимальных условий процесса и максимизации выхода и качества продукта риформинга.

Внедрение СППР для контроля температуры при каталитическом риформинге может привести к повышению эффективности процесса, снижению энергопотребления и повышению качества и консистенции продукта.

В исследовании Saini и Singh (2019) было обнаружено, что внедрение СППР для оптимизации процесса каталитического риформинга привело к увеличению производства высокооктанового бензина на 10%[5]. В исследовании также сообщалось о сокращении потребления природного газа на 30%, а также о сокращении количества выбросов CO2.

В другом исследовании, проведенном Rahimi et al. (2017), сообщалось об увеличении выхода высокооктанового бензина на 5,5% и снижении энергопотребления на 12% после внедрения СППР для процесса каталитического риформинга.

В тематическом исследовании IBM[3] крупная нефтегазовая компания использовала СППР для оптимизации процесса каталитического риформинга, что привело к увеличению выхода высокооктанового бензина на 10% и снижению энергопотребления на 5%.

Эти примеры демонстрируют потенциальные преимущества внедрения СППР в процесс каталитического риформинга, включая увеличение производства высокооктанового бензина, снижение энергопотребления и сокращение выбросов CO2.

II. Заключение

В заключение следует отметить, что процесс каталитического риформинга является сложным процессом в управлении, но имеется несколько инструментов поддержки принятия решений, которые помогают руководителям принимать обоснованные решения относительно процесса. Эти инструменты предоставляют информацию о процессе в режиме реального времени, прогнозируют будущие тенденции и рекомендуют действия по оптимизации процесса для повышения эффективности и качества продукции. Независимо от того, хотите ли вы оптимизировать производительность существующего процесса каталитического риформинга или рассматриваете возможность установки нового процесса, эти инструменты поддержки принятия решений помогут вам принимать наилучшие решения для вашего бизнеса.

Некоторые конкретные преимущества, наблюдаемые этими компаниями после внедрения СППР, заключаются в следующем:

Повышение эффективности процесса: Внедрение СППР привело к повышению эффективности процесса за счет предоставления информации об условиях процесса в режиме реального времени, включая температуру, давление и активность катализатора, и позволяет менеджерам принимать обоснованные решения относительно процесса. Оптимизируя технологические условия, СППР помогает повысить выход высокооктанового бензина и другого ценного нефтехимического сырья.

Повышенное качество продукции: СППР предоставляет информацию об условиях технологического процесса в режиме реального времени, что позволяет менеджерам контролировать качество продукции и вносить необходимые коррективы. Это приводит к сокращению количества нестандартных продуктов и повышению качества конечного продукта.

Сокращение времени простоя: СППР помогает выявлять и диагностировать проблемы в процессе, позволяя руководителям предпринимать корректирующие действия до возникновения серьезной проблемы. Это приводит к сокращению времени простоя и производственных потерь.

Повышенная безопасность: СППР предоставляет информацию об условиях технологического процесса в режиме реального времени, что позволяет руководителям отслеживать параметры безопасности и при необходимости принимать корректирующие действия. Это помогает предотвратить несчастные случаи и обеспечить безопасность персонала и оборудования.

III. Список литературы

1. Каталитический риформинг углеводородов: справочник / под ред. проф. Р. И. Кузьминой – Саратов: Издательство СУИ МВД России, 2010. – 252 с.
2. Промышленные установки каталитического риформинга, справочник / В. А. Гуляев [и др.] под ред. Г. А. Ластовкина. – Л.: Химия, 1984.
3. George J. Antos. Catalitic naphtha reforming: Science and Technology edited by George J. Antos, Abdullah M. Aitani, Jose M. Parera – New York – Basel – Hong Kong: Marsel Dekker, Inc., 1995.
4. Глазов Г. И. Каталитический риформинг и экстракция ароматических углеводородов [Текст] / Г. И. Глазов, В. П. Сидорин. – М.: Химия, 1981. – 188 с.
5. Fawzi, M. E. Catalytic Naphtha Reforming; Challenges for Selective Gasoline an Overview and Optimization Case Study. J. Adv. Catal. Sci. Technol. 2016, 3, 27-42.

РАЗЛИЧНЫЕ ПОДХОДЫ К РАСЧЕТУ СЕРДЕЧНИКОВ ТРАНСФОРМАТОРОВ

Свито И. Л., Тихоновецкий С. А.
Кафедра теоретических основ электротехники,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: svito@bsuir.by, tsa1954@tut.by

В работе предлагается нетрадиционный физический подход к расчету габаритной мощности сердечника через амплитудное значение синусоидальной мощности в нем, а также физический подход дополняется традиционным конструктивным.

Введение

Габаритная мощность P_{core} – это условное понятие, применяющееся при выборе сердечника трансформатора.

В данной статье, во первых, предлагается нетрадиционный физический подход к расчету габаритной мощности сердечника через амплитудное значение синусоидальной мощности в нем. Во вторых, физический подход дополняется традиционным конструктивным. Различные подходы увязываются между собой, что дает инженеру понимание того, что и почему он делает при расчете сердечника. Знание физики процессов позволяет применять формулы осмысленно, что значительно понижает вероятность ошибки.

I. Физический подход

Как известно из курса электротехники [1], энергия магнитного поля в единице объема сердечника определяется следующей формулой:

$$W_{\mu_0} = \frac{HB}{2} = \frac{\mu_a H^2}{2} = \frac{B^2}{2\mu_a}. \quad (1)$$

Здесь $B = \mu_a H$ – действующее значение индукции магнитного поля в сердечнике, H – действующее значение напряженности магнитного поля, $\mu_a = \mu_0\mu_r$ – абсолютная магнитная проницаемость материала сердечника с относительной магнитной проницаемостью μ_r, наконец, $\mu_0 = 4\pi 10^{-7} Henry/m$ – магнитная проницаемость вакуума.

Рассмотрим сердечник кольцевой формы с диаметрами D и d. Его параметры: S_c – площадь сечения сердечника, $S_o = \pi d^2/4$ – площадь окна сердечника, $l_m = \pi(D+d)/2$ – средняя длина силовой линии, $V = \pi h(D^2 - d^2)/4$ – объем сердечника, B_m – индукция насыщения ферромагитного материала.

На сердечник намотана катушка с числом витков w, по которой протекает синусоидальный ток с круговой частотой ω rad/sec: $i(t) = I_m sin\omega t$. Тогда напряженность магнитного поля в сердечнике выражается формулой $H(t) = H_m sin\omega t$. Ее амплитуда H_m может быть найдена согласно закону полного тока:

$$H_m l_m = I_m w \rightarrow H_m = I_m w/l_m. \quad (2)$$

Подставим данную напряженность в (1). Получим, что энергия магнитного поля в единице объема пульсирует по закону синуса в квадрате:

$$W_{\mu_0} = \frac{\mu_a H_m^2}{2} sin^2\omega t. \quad (3)$$

Мгновенную мощность в единице объема определим через производную от энергии:

$$p(t) = \frac{dW_{\mu_0}}{dt} = \frac{1}{2}\mu_a \omega H_m^2 sin2\omega t. \quad (4)$$

Тогда мгновенная мощность магнитного поля в сердечнике объемом V определится выражением

$$p_m(t) = P_m sin2\omega t, \quad (5)$$

где амплитудное значение мощности магнитного поля

$$P_m = \frac{1}{2}\mu_a \omega H_m^2 V. \quad (6)$$

Объем кольцевого сердечника выражается формулой

$$V = S_c l_m, \quad (7)$$

где S_c – это площадь сечения сердечника.

Предположим, что амплитуда тока такова, что индукция в сердечнике изменяется между значениями насыщения, то есть от $-B_m$ до $+B_m$ и обратно. Тогда максимально возможное значение мощности в данном сердечнике определится так:

$$P_{max} = \frac{B_m^2}{2\mu_a}\omega V. \quad (8)$$

Выбранная из числа стандартных габаритная мощность P_{core} сердечника должна быть больше, чем P_{max}.

Объем сердечника и число витков на один вольт

Оцениваются следующими формулами:

$$V_{min} = \frac{2\mu_0\mu_r P_{max}(VA)}{B_m^2(Tesla)\omega(sec^{-1})}. \quad (9)$$

$$w_0 = (4.44\Phi_m f)^{-1} = (4.44 B_m S_c f)^{-1}, \quad (10)$$

II. Конструктивный подход

В инженерных расчетах габаритную мощность P_{core} традиционно определяют как полусумму полных мощностей в первичной и вторичной обмотках.

$$P_{core} = \frac{1}{2}(U_1 I_1 + U_2 I_2) = \frac{1}{2} j (s_1 U_1 + s_2 U_2). \quad (11)$$

Здесь j – это допустимая плотность тока, одинаковая в обеих обмотках. В малых трансформаторах мощностью до 1 кВт она выбирается в пределах $(3-5) A/mm^2$. Соответственно, s_1,s_2 – это сечения проводов первичной и вторичной обмотки.

Выбранный из числа типовых сердечник должен обладать габаритной мощностью большей, чем P_{core}.

Согласно формуле трансформаторной ЭДС

$$\begin{aligned} U_1 &= 4.44 B_m S_c f w_1, \\ U_2 &= 4.44 B_m S_c f w_2. \end{aligned} \quad (12)$$

Подставляя (12) в (11), получим:

$$\begin{aligned} P_{core} &= \frac{1}{2} j (4.44 B_m S_c f)(s_1 w_1 + s_2 w_2) = \\ &= 2.22 j B_m S_c f (s_1 w_1 + s_2 w_2). \end{aligned} \quad (13)$$

Введем понятие площади меди – это суммарное сечение проводов всех витков первичной и вторичной обмотки:

$$S_m = s_1 w_1 + s_2 w_2. \quad (14)$$

Тогда (13) перепишется следующим образом:

$$P_{core} = 2.22 j B_m S_c f S_m. \quad (15)$$

Максимальная площадь меди S_m ограничена площадью окна S_o. Эта связь выражается коэффициентом заполнения:

$$\sigma = S_m / S_o \<1. \quad (16)$$

Для многослойной обмотки $\sigma = 0.15 - 0.4$, то есть, площадь меди всегда меньше площади окна.

Подставляя (16) в (15), окончательно получим:

$$P_{core} = 2.22 j B_m S_c S_o f \sigma. \quad (17)$$

Данный подход не дает возможности получить объем сердечника. Здесь только можно найти произведение сечения сердечника S_c и площади окна S_o:

$$S_c S_o = \frac{P_{core}}{2.22 j B_m f \sigma}. \quad (18)$$

В инженерных расчетах принимают $S_c = S_o$, тогда сечение сердечника

$$S_c = \sqrt{\frac{P_{core}}{2.22 j B_m f \sigma}}. \quad (19)$$

III. Вывод

При сравнении различных подходов можно видеть, что P_{max} в физическом подходе меньше, чем P_{core}, полученная при конструктивном подходе. Это естественно, поскольку это разные мощности: амплитуда мощности синусоидального магнитного поля в сердечнике и некая полная мощность P_{core}, по сути дела, абстрактная величина, обеспечивающая передачу мощности, заведомо большей, чем амплитудная P_{max}.

IV. Список литературы

1. Бессонов Л. А. Теоретические основы электротехники : Электрические цепи. Учебник для студентов вузов, 11-е издание, переработанное и дополненное. Учебное издание. Москва : Гардарики, 2007.

СИСТЕМА УПРАВЛЕНИЯ СТАНЦИЕЙ КАТОДНОЙ ЗАЩИТЫ

Стешенко П. П., Иванюкович А. В., Свито И. Л., Журавлёв В. И.
Кафедра теоретических основ электротехники
Белорусский государственный университет информатики и радиоэлектроники
ООО «АльпИндустрия»
Минск, Республика Беларусь
E-mail: sppmailbox@mail.ru

Приведен анализ существующих устройств защиты от коррозии металлических объектов. Предложена система управления станцией, используя метод катодной поляризации, в котором контролируется ток и напряжение процесса защиты конструкции.

Введение

Многие трубопроводы и металлические конструкции имеют срок службы, часто превышающий десятилетия. Для поддержания их целостности и безопасности катодная защита играет важную роль в предотвращении коррозии и увеличении долговечности систем. Обычно указанный метод применяется для защиты: подземных нефтепроводов и газопроводов; трубопроводов водоснабжения и теплосетей; оболочек электрических кабелей; металлических резервуаров, крупных объектов; подземных сооружений; судов от коррозии в воде; стальной арматуры в фундаментах, в железобетонных сваях [1].

I. Основная часть

Станции катодной защиты используются для электрохимической защиты от коррозии металлических объектов методом катодной поляризации. С развитием технологий и материалов современные системы катодной защиты могут быть более эффективными и экономически целесообразными, чем старые методы (защита покрытиями). Это поднимает вопрос о необходимости обновления и модернизации существующих систем. Станции прошлых лет представляют собой трансформаторы с тиристорными регуляторами. Современные станции – это инверторные преобразователи с микропроцессорным управлением и GSM телемеханикой.

Применяемые станции имеют ряд преимуществ и недостатков. Так станция ТДЕ(Д)-3-220 УЗ с силовым блоком на тиристорах [2] имеет относительно высокий КПД, но отсутствует цифровой интерфейс и системы телеметрии. Станция катодной зашиты «Тверца-900» [3] и ее модификации предназначена для непрерывной электрохимической зашиты подземных металлических сооружений от почвенной коррозии. Станция обеспечивает возможность как ручного, так и дистанционного управления и получения информации через встроенный GSM-модем (CSD, SMS, GPRS) или проводные линии связи RS-485. Состоит из силовых преобразователей, блока бесперебойного питания, блока защиты от перенапряжений. Однако характерна довольно высокой стоимостью и габаритами.

Архитектура разрабатываемой нами системы универсальна, что позволяет осуществлять адаптацию в соответствии с требованиями потребителя с минимальными затратами. Основные предъявляемые к разрабатываемой системе требования: минимальная стоимость системы, соответствующие в классе метрологических и точностных характеристики, низкая потребляемая мощность. Проанализировав технические условия, приведенные в ГОСТ 12997-84 Изделия ГСП, общие технические условия [4], определили основные технические требования к разрабатываемой конструкции [5]:

- автоматическая стабилизация выходного тока;
- регулирование выходного тока и напряжения;
- преобразование аналоговых сигналов с датчиков напряжения и тока;
- индикация измеренных напряжения, тока, времени наработки;
- удаленное управление и передача данных по сети Ethernet;
- удаленное управление и передача данных через встроенный модуль GSM;
- управление системой сигнализации;
- управление реле защиты.

Модернизация основана на применении программного управления и удаленного контроля, что повышает надежность системы и позволяет интегрировать ее в системы централизованной диспетчеризации объектов. На рисунке 1 представлена структурная схема станции. Сигналы с датчиков синхронно записываются во временном режиме в виде амплитудно-частотного спектра с последующей архивацией и дальнейшей детальной проработкой сигналов. Дальнейший анализ полученных сигналов возможен с использованием различных специализированных экспертных программ. Станция подключается к питающей сети (как правило, $\sim$ 220 В) и формирует ток с заданными параметрами.

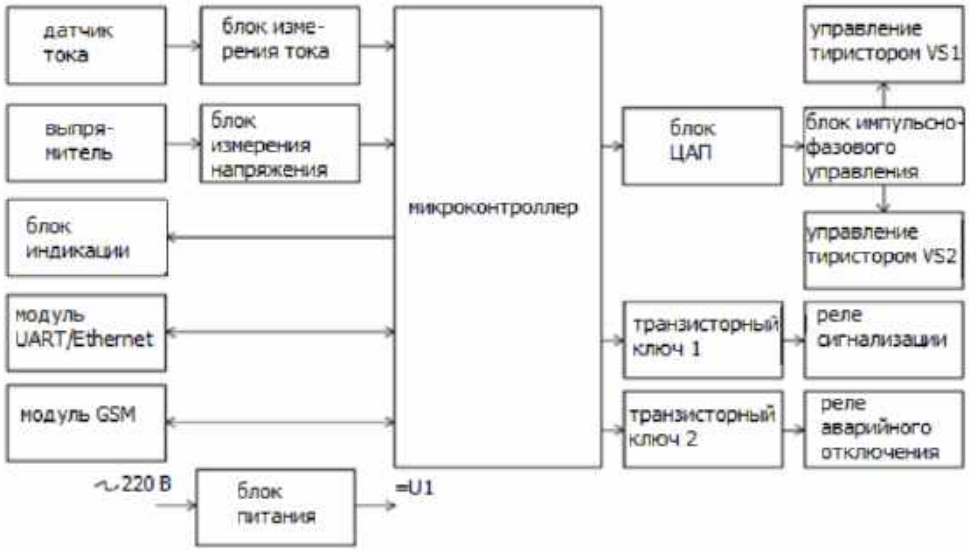


Рис. 1 – Структурная схема станции катодной защиты

Станция катодной защиты работает в условиях, изменяющихся во времени, которые могут обусловливаться воздействием блуждающих токов, изменения удельного сопротивления проводящих сред, контактных устройств или другими факторами. Поэтому разработанное устройство обеспечивает автоматическое регулирование. Блок управления реализуется на основе микроконтроллера STM32F103C8T6. Микроконтроллер имеет высокое соотношение производительность/-потребление, максимальную тактовую частоту 72 МГц и весь необходимый набор аппаратных модулей. Микроконтроллер STM32F103C8T6 представляет из себя систему, выполненную на кристалле.

Блок импульсно-фазового управления синхронизирован от питающей сети и выделяет момент начала отсчета в плюсовой или минусовой полуволне фазного напряжения и определяет интервал отсчета фазы, которая необходима для смещения управляющих импульсов (угол открытия тиристоров). Необходимый уровень выходного постоянного напряжения формирует цифро-аналоговый преобразователь ЦАП AD5170, который обеспечивает дискретность до 256 уровней изменения угла открытия тиристоров на один полупериод, что позволяет плавно изменять напряжение на нагрузке. Далее сигнал по ступает на входы реверсивного таймера угла отсечки фазы, в котором происходит формирование импульса управления тиристорным выпрямителем. Поскольку длительность управляющего сигнала очень мала, силовые ключи, в некоторых случаях могут не открыться. Для решения этого вопроса используются специальные блоки формирования импульса.

II. Заключение

Конструктивно станция представляет собой однокристальный компьютер, способный выполнять простые задачи. Использование одной микросхемы, вместо целого набора, как в случае обычных процессоров, применяемых в персональных компьютерах, значительно снижает размеры, энергопотребление и стоимость устройств, построенных на базе микроконтроллеров.

III. Список литературы

1. Современные системы защиты от электрохимической коррозии подземных коммуникаций. Том 1. Катодная защита густоразветвленной сети подземных трубопроводов / Д. Л. Рахманкулов [и др.]. – Уфа : Реактив, 2020. – 160 с.
2. Станция катодной защиты ТДЕ(Д)-3-220 УЗ [Электронный ресурс]. – Режим доступа: `http://sielectr.ru/DY_EO_ENO_1.1.pdf`.
3. СКЗ «Тверца-900» [Электронный ресурс]. – Режим доступа `http://www.eltech.tver.ru/netcat_files/tabs/101_17_T900_RE.pdf`.
4. ГОСТ 12997-84. Изделия ГСП. Общие технические условия. – Введ. 1986-07-01. – М. : Изд-во стандартов, 1986.
5. ГОСТ 15150-69. Машины, приборы и другие технические изделия. Исполнения для различных климатических районов. Категории, условия эксплуатации, хранения и транспортирования в части воздействия климатических факторов внешней среды. – Введ. 1971-01-01. – М. : Изд-во стандартов, 1981. – 76 с.

СЕРВОПРИВОД СО ВСТРОЕННЫМ В НАБЛЮДАТЕЛЬ КОМБИНИРОВАННЫМ УПРАВЛЕНИЕМ

Хаджинов М. К., Павлова А. В., Стасевич Н. А.
Кафедра систем управления,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {pavlova, stasevich }@bsuir.by

Рассматривается сервопривод подчинённого регулирования с модальным регулятором и наблюдателем. Для увеличения точности управления предлагается применить задатчик сигналов комбинированного управления, который предлагается совместить с наблюдателем. Приведены варианты настроек комбинированного управления, обеспечивающих эффективную работу сервопривода с задатчиком в наблюдателе.

Введение

Современные автоматические системы высокой динамической точности обычно строят на основе принципа комбинированного управления, сочетающего в себе принципы управления по отклонению и по возмущению. Точность работы комбинированных систем выше точности систем, использующих только один из принципов управления, причем недостатки обоих принципов при их объединении устраняются.

Высокоточный сервопривод обычно имеет структуру подчинённого регулирования с отдельными регуляторами тока, скорости и координаты [1]. Очевидно, что координированное управление сразу тремя сигналами на входах всех трёх регуляторов эффективнее управления только одной координатой [2]. Представляет несомненный интерес введение нелинейных функций в каналы комбинированного управления. В первую очередь это касается канала комбинированного управления по скорости [3].

I. Постановка задачи.

Для реализации комбинированного управления скоростью, ускорением и координатой, необходимо синтезировать алгоритм для вычисления скорости и ускорения из управляющего сигнала координатой. Эти функции будет выполнять задатчик комбинированного управления, представляющий собой дифференцирующий фильтр высокого порядка.

Для упрощения структуры электропривода поставлена задача совместить дополнительный дифференцирующий фильтр с наблюдателем или его частью и разработать приемлемые условия для такого совмещения.

II. Структура дифференцирующего фильтра задатчика в наблюдателе.

Рассмотрим последовательность задач для совмещения дифференцирующего фильтра задатчика в наблюдателе и их решения.

1 Совмещение связи дифференцирующего фильтра задатчика и наблюдателя с сигналом внешнего управления.

Для этой цели в схеме электропривода со структурой подчинённого регулирования с модальным регулятором контура управления, т.е. с отсутствием интегральных регуляторов в прямой цепи управления, есть совмещение связи наблюдателя и дифференцирующего фильтра с сигналом внешнего управления.

2 Совмещение структур дифференцирующего фильтра задатчика и наблюдателя.

Для модели объекта управления в наблюдателе следует использовать управляемую присоединённую каноническую форму, представляющею собой цепочку последовательно включённых интеграторов. Такая структура благоприятна для формирования дифференцирующего фильтра.

3 Совмещение полос пропускания дифференцирующего фильтра задатчика и наблюдателя.

Для качественного вычисления производных желательна наибольшая полоса пропускания дифференцирующего фильтра, настроенного на граничную частоту его компьютерной реализации. В наблюдателе соответствующая полоса пропускания будет зависеть от настройки модального регулятора контура управления, и будет меньше граничной частоты компьютерной реализации наблюдателя, Т.е. полоса пропускания дифференцирующего фильтра будет совпадать с полосой пропускания контура управления, что вполне приемлемо.

Задатчик комбинированного управления представлен в виде следящего оценивающего фильтра с модальным регулятором контура управления. Сигналы производных для комбинированного управления берутся со входа первого интегратора и выходов интеграторов последовательной цепочки до предпоследнего.

Коэффициенты линейного комбинированного управления образуют строку дополнительного выхода наблюдателя. Нелинейное комбинированное управление [3] потребует внешних функций вне наблюдателя, подключаемых к дополнительным выходам сигналов производных от наблюдателя.

Сигнал производной из дополнительного выхода наблюдателя подаётся на вход контура тока

электропривода как комбинированное управление.

III. Результаты моделирования

Было проведено моделирование электропривода с наблюдателем, компенсирующим десятикратное различие инерционностей модели в контроллере от реального контура тока в электроприводе. Осуществляется форсирование сравнительно медленных процессов контура тока.

Настройка модального регулятора контура управления и контура оценивания проводилась на частоту 5000 1/с. Настройка динамики системы осуществлялась с использованием полинома Бесселя, который обеспечивает монотонность процессов, минимальное перерегулирование.

На рис.1 представлены частотные и временные характеристики сервопривода. Процессы управления по длительности меньше одной милисекунды, и по форме соответствуют эталонной модели, использованной при настройки модального регулятора.

IV. Заключение

Формирование комбинированного управления в наблюдателе существенно улучшает качество электропривода без расширения его структуры.

Предлагаемая система комбинированного управления позволяет легко нарастить число используемых производных для комбинированного управления и сформировать эффективный регулятор для управления сложным объектом управления по значительно упрощенной компьютерной модели в наблюдателе.

Применение в канале комбинированного управления алгоритма нелинейной квадратичной функции с сохранением знака позволят существенно улучшить качество процессов в электроприводе.

1. Хаджинов, М. К. Сервоприводы с квадратичной обратной связью по скорости / М. К. Хаджинов, А. В. Павлова , А. Т. Доманов // Материалы международная научной конференции «Информационные технологии и системы»(ИТС 2017). – 2017.
2. Хаджинов, М. К. Сервопривод с комбинированным управлением / М. К. Хаджинов, А. В. Павлова , А. Т. Доманов// Материалы международная научной конференции «Информационные технологии и системы»(ИТС 2019). – 2019.
3. Хаджинов, М. К. Сервопривод с нелинейным комбинированным управлением / М. К. Хаджинов, А. В. Павлова , А. Т. Доманов// Материалы международная научной конференции «Информационные технологии и системы»(ИТС 2020). – 2020.

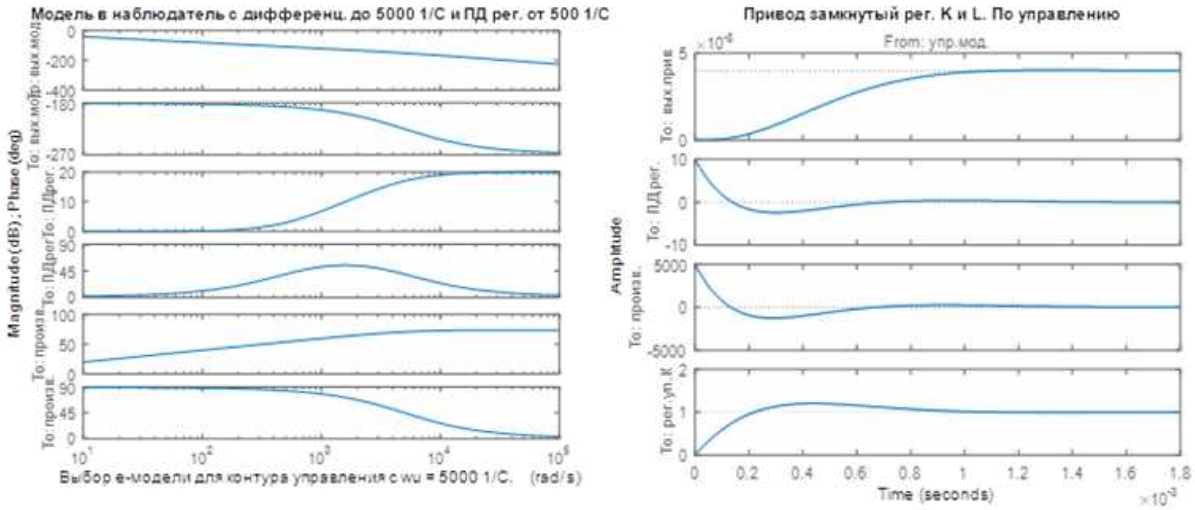


Рис. 1 – Результаты моделирования

Секция «Интеллектуальные информационные технологии»

Section «Intelligent Information Technologies»

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ТЕХНОЛОГИЙ ДОПОЛНЕННОЙ И ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ИНЖЕНЕРИИ

Акмырадова А. Я.
Кафедра компьютерные науки и информационные технологии,
Инженерно-технологический университет Туркменистана имени Огуз хана
Ашхабад, Туркменистан
E-mail: aylarakmyradova89@gmail.com

В последние годы технологии дополненной (AR) и виртуальной реальности (VR), интегрированные с искусственным интеллектом (AI), стали ключевыми инструментами в инженерии. Их использование позволяет создавать виртуальные прототипы, проводить симуляции сложных процессов и обеспечивать персонализированное обучение специалистов. Интеграция AI обеспечивает интеллектуальный анализ данных, прогнозирование потенциальных ошибок и адаптивное взаимодействие с пользователем. В данной статье рассматриваются теоретические основы AR и VR, архитектура систем, методы AI-интеграции и основные области применения в инженерной практике.

Введение

Современная инженерия сталкивается с проблемами ускоренного проектирования, обучения и контроля качества, что требует внедрения инновационных инструментов для повышения эффективности и точности рабочих процессов. Технологии дополненной (AR) и виртуальной реальности (VR) предоставляют инженерам возможность визуализировать сложные конструкции, моделировать технологические процессы и проводить интерактивное обучение без необходимости создания физических прототипов. Это значительно сокращает время разработки и снижает затраты на испытания. Искусственный интеллект (AI), интегрированный в такие системы, обеспечивает интеллектуальный анализ данных, адаптивное взаимодействие с пользователем, автоматическое выявление ошибок и прогнозирование потенциальных проблем. Совместное использование AR/VR и AI создаёт новую парадигму инженерной деятельности, где цифровые модели становятся активными участниками проектного процесса, способными обучаться и совершенствоваться на основе накопленного опыта.[1-3]

Основные преимущества интеграции AI с AR/VR включают:

- Повышение точности проектирования;
- Снижение затрат на прототипы;
- Персонализированное обучение;
- Поддержка принятия решений в режиме реального времени.

I. Архитектура систем AR, VR и AI

Современные системы AR и VR с AI включают аппаратную часть, программное обеспечение, искусственный интеллект и интерфейс пользователя.[4-6]

Таблица 1 – Схема интеграции AR/VR с AI в инженерной системе

Компонент	Описание	Пример
Аппаратная часть	Дисплеи, датчики движения, контроллеры, камеры	Шлемы VR, очки AR, сенсоры движения
Программная часть	Графические движки, алгоритмы трекинга, модели взаимодействия	Unity, Unreal Engine, OpenCV
Искусственный интеллект	Алгоритмы машинного обучения, распознавание объектов, прогнозирование	CNN, RNN, reinforcement learning
Интерфейс пользователя	Визуализация, сенсорное и голосовое управление, интерактивное взаимодействие	Контроллеры движения, голосовые команды, AR-приложения

AI обеспечивает:

- Адаптивное взаимодействие с пользователем;
- Анализ поведения и прогнозирование ошибок;
- Оптимизацию процессов моделирования.

II. Основные области применения

1. **Проектирование и моделирование.** Виртуальные прототипы позволяют инженерам проверять конструкции до физического изготовления. AI анализирует возможные нагрузки, выявляет потенциальные слабые места и предлагает оптимальные решения.

Пример: Моделирование аэродинамики в VR с AI-анализом потоков воздуха для улучшения конструкции дрона.

2. **Обучение и симуляция.** Интерактивные тренажёры с AI подстраивают сценарии под уровень знаний пользователя, моделируют аварийные ситуации и редкие события. Пример: AR-симулятор для обучения монтажу сложного оборудования с подсказками в реальном времени.
3. **Контроль качества и обслуживание.** AR позволяет видеть внутренние компоненты машин без разборки. AI анализирует состояние оборудования, прогнозирует износ и аварийные ситуации. Пример: Инспекция турбин с AR-очки, отображающие износ лопаток и прогноз ресурса работы.[7-8]

Таблица 2 – Области применения AR/VR с AI в инженерии

Область применения	Описание	Преимущество
Проектирование	Виртуальные прототипы с AI	Оптимизация конструкций, снижение затрат
Обучение	Симуляторы с AI	Безопасное и адаптивное обучение
Контроль качества	Диагностика с прогнозированием AI	Предотвращение поломок, ускорение обслуживания
Коллаборативная работа	Совместное проектирование	Удалённая работа и улучшение коммуникации

III. Методы интеграции AI

1. Обработка изображений и трекинг – AI улучшает точность распознавания объектов и движений.
2. Анализ поведения пользователя – алгоритмы машинного обучения адаптируют сценарии под индивидуальные особенности.
3. Прогнозирование ошибок и отказов – AI анализирует данные сенсоров для предсказания поломок.
4. Оптимизация взаимодействия – AI корректирует визуализацию, предоставляет рекомендации и улучшает UX.[6-7]

Заключение

Интеграция AR/VR с искусственным интеллектом значительно расширяет возможности инженерной деятельности, повышая эффективность проектирования, обучения и контроля качества. Эти технологии создают основу для формирования «умных» инженерных экосистем, где взаимодействие человека и машины становится более интуитивным и продуктивным. Перспективные направления развития включают улучшение точности трекинга и распознавания объектов, интеллектуальное прогнозирование поведения сложных систем, интеграцию с облачными вычислениями и Интернетом вещей (IoT) для удалённого моделирования и мониторинга. Особое внимание уделяется созданию персонализированных обучающих и симуляционных программ с адаптивным AI, которые способны подстраиваться под уровень подготовки пользователя. Использование этих технологий способствует снижению затрат, ускорению проектных циклов, повышению безопасности инженерных процессов и формированию новой парадигмы цифрового проектирования в эпоху индустрии 5.0.

Список литературы

1. Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. IEICE Transactions on Information and Systems, E77-D(12), 1321–1329.
2. Azuma, R. T. (1997). A survey of augmented reality. Presence: Teleoperators and Virtual Environments, 6(4), 355–385.
3. Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. Journal of Communication, 42(4), 73–93.
4. Sherman, W. R., & Craig, A. B. (2018). Understanding Virtual Reality: Interface, Application, and Design (2nd ed.). Morgan Kaufmann.
5. Craig, A. B. (2013). Understanding Augmented Reality: Concepts and Applications. Morgan Kaufmann.
6. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
7. Russell, S., & Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
8. Zhang, Z., & Wang, L. (2020). AI-driven virtual reality for engineering simulation. Journal of Engineering Applications of AI, 95, 103882.
9. Lee, J., & Kim, H. (2019). Augmented reality with machine learning for industrial training. Computers in Industry, 109, 43–55.

АРХИТЕКТУРА RETRIEVAL-AUGMENTED GENERATION КАК ОСНОВА ПОСТРОЕНИЯ ЛОКАЛЬНЫХ ИИ-СИСТЕМ С КОРПОРАТИВНЫМИ БАЗАМИ ЗНАНИЙ

Арефин В. А.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: arefin.vlad@gmail.com

Целью работы является исследование архитектуры Retrieval-Augmented Generation (RAG) как основы построения локальных интеллектуальных систем, взаимодействующих с корпоративными базами знаний. Рассмотрены ключевые компоненты архитектуры и принципы интеграции механизма извлечения знаний с языковыми моделями. Проведён анализ преимуществ RAG в контексте корпоративных информационных систем, включая повышение точности и достоверности ответов, снижение вероятности генерации ложных данных и обеспечение актуальности знаний без переобучения модели. Определены основные направления развития и практические аспекты внедрения архитектуры RAG в локальных ИИ-решениях.

I. Введение

Современные большие языковые модели (Large Language Models, LLM) демонстрируют значительный прогресс в генерации связного текста и ведении диалога, однако они имеют фундаментальные ограничения при работе с внутренними корпоративными данными. Большинство моделей обучены на открытых наборах данных и не обладают механизмом динамического обновления знаний. Это делает их использование в корпоративных сценариях неэффективным без дополнительной архитектурной поддержки.

Метод Retrieval-Augmented Generation (RAG) решает данную проблему, совмещая генерацию и поиск релевантных данных [1]. Перед формированием ответа система выполняет извлечение необходимых фрагментов из базы знаний, что повышает точность и обоснованность результатов, снижает эффект «галлюцинаций» и адаптирует модель под специфику организации [2, 5]. Подход RAG активно применяется при построении корпоративных интеллектуальных систем, чат-ботов, экспертных ассистентов и платформ поддержки принятия решений.

II. Архитектура RAG

Архитектура RAG обычно включает три ключевых этапа: (1) подготовку и индексацию базы знаний (chunking, векторизация); (2) извлечение релевантных документов по запросу (retrieval); (3) генерацию ответа с учётом найденных фрагментов (generation) [6]. На практике также используются дополнительные механизмы: reranking, фильтрация по достоверности, комбинирование источников и последующая верификация.

Подготовка корпоративной базы знаний

Корпоративная база знаний может включать регламенты, технические отчёты, инструкции, записи совещаний, внутренние чаты и инцидентные отчёты. Для эффективного применения RAG необходимо провести очистку, структурирование и семантическую разметку данных. Формирование векторного индекса на основе эмбеддингов обеспечивает быстрый поиск релевантных фрагментов, а сохранение метаданных позволяет контролировать источники и версии документов. Особое внимание уделяется вопросам безопасности, разграничению доступа и контролю за актуальностью данных.

Этап извлечения и генерации

На этапе retrieval пользовательский запрос преобразуется в векторное представление и сопоставляется с индексом знаний. Далее выбираются K наиболее релевантных фрагментов, которые подаются вместе с запросом в генеративную модель. Эта модель формирует ответ, опираясь не только на параметры своего обучения, но и на актуальные корпоративные данные [7]. Такой подход обеспечивает синтез свежих и достоверных ответов даже при ограниченном обучении основной LLM.

Архитектурный паттерн и инфраструктура

Типовая инфраструктура RAG включает:

- систему хранения и индексации документов (векторные, графовые или гибридные БД);
- модуль поиска и ранжирования релевантных фрагментов;
- генеративную модель с продуманным prompt-инжинирингом;
- подсистему контроля качества и интерпретации ответов (fact-checking, логирование источников);
- интерфейс пользователя — чат-бот, API или аналитический портал.

Для корпоративных сценариев важна возможность локального развёртывания (on-premises) с соблюдением требований безопасности

и приватности. Согласно аналитическим данным, более половины компаний, внедряющих решения на базе ИИ, рассматривают архитектуру RAG как основную при работе с внутренними данными [1].

III. Применение в локальных корпоративных системах

Кейс: корпоративный чат-бот

В крупных организациях объём внутренних документов может достигать десятков тысяч файлов, включая нормативные акты, регламенты, методические указания, отчётные формы и переписку сотрудников. Такие массивы данных являются ценным источником корпоративных знаний, однако их использование традиционными поисковыми средствами ограничено.

Применение архитектуры RAG позволяет создать интеллектуального чат-бота, который при каждом пользовательском запросе выполняет поиск релевантных документов, анализирует их содержание и формирует ответ с указанием конкретных источников [3, 6]. В результате система становится полноценным интеллектуальным помощником, способным интерпретировать контекст и связывать разрозненные сведения.

Подобный подход значительно повышает доверие пользователей, сокращает время на поиск информации и обеспечивает соответствие ответов актуальным нормативам и внутренним политикам организации.

Преимущества и вызовы

Применение RAG в корпоративной среде даёт ряд стратегических преимуществ:

- снижение вероятности генерации недостоверных или вымышленных ответов благодаря обращению к проверенным источникам;
- актуализация знаний без повторного обучения модели, что снижает затраты на сопровождение;
- возможность внедрения в изолированной корпоративной инфраструктуре с соблюдением политик безопасности и конфиденциальности;
- повышение прозрачности работы системы и доверия со стороны пользователей за счёт цитирования источников и ведения журналов обращений;

Успешное внедрение RAG требует решения ряда технологических и организационных задач. Ключевыми вызовами являются обеспечение качества индексации и полноты охвата базы знаний, масштабируемость архитектуры при росте объёмов данных, а также поддержка многоязычных документов и сложных форматов (PDF, таблицы, графики).

IV. Выводы

Архитектура RAG формирует основу для построения надёжных и адаптивных корпоративных ИИ-систем [4]. Её применение позволяет соединить преимущества больших языковых моделей и актуальных баз знаний, обеспечивая обоснованность и свежесть ответов. Перспективными направлениями развития являются интеграция мультимодальных данных, исследование самонастраивающихся retrieval-модулей, а также оптимизация производительности при обработке больших объёмов корпоративной информации [5, 2].

Список литературы

1. Yee, L., Chui, M., Roberts, R., Pometti, M., Wollner, P. What is Retrieval-Augmented Generation (RAG)? // McKinsey & Company Explainer. – 2024.
2. Zhang, H., Chen, M., Li, T. Optimizing Retrieval-Augmented Generation Pipelines for Domain-Specific Knowledge Bases // Proc. IEEE Int. Conf. on Artificial Intelligence and Knowledge Engineering (AIKE). – 2023. – P. 215-222.
3. Нестеренков, С. Н., Красовский, В. Ю., Баяк, Е. И. Искусственный интеллект и персональные ассистенты в управлении проектами // Труды БГУИР. – 2024. – № 2 (140). – С. 35-42.
4. Нестеренков, С. Н., Байчик, С. А., Голубович, Ю. И. Искусственный интеллект в процессах обучения и развития персонала // Информационные технологии и системы: материалы конф. – Минск: БГУИР, 2024. – С. 112-115.
5. Gupta, S., Ranjan, R., Singh, S. N. A Comprehensive Survey of Retrieval-Augmented Generation: Evolution, Current Landscape and Future Directions // arXiv preprint arXiv:2410.12837. – 2024.
6. Martinez, A., Bae, J., Qian, Y. Towards Reliable Corporate AI Assistants: Improving Contextual Reasoning in RAG Systems // Proceedings of the 2025 ACM Conference on Applied Artificial Intelligence (AAAI Industry Track). – 2025. – P. 88-97.
7. GenAI Adoption 2024: The Challenge with Enterprise Data // K2view Technical Report. – 2024.

СИСТЕМА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА АНАЛИЗА ДАННЫХ ТЕМАТИЧЕСКИХ САЙТОВ С ИСПОЛЬЗОВАНИЕМ RAG ТЕХНОЛОГИЙ, LLM И ГРАФА ЗНАНИЙ

Батура М. П., Кулевич А. О.
Кафедра информатики, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: bmpbel@bsuir.by, kulevich.01@gmail.com

В статье описывается подход к созданию системы анализа данных тематических сайтов с использованием технологий RAG, графовой базы данных Neo4j и библиотеки LangGraph. Представлена структура графа знаний и работа агента GraphReader, выполняющего интеллектуальный поиск и генерацию ответов с участием языковой модели. Показана реализация веб-интерфейса, который позволяет загружать данные и получать результаты анализа в режиме диалога.

Введение

Современные информационные системы обрабатывают всё большие объёмы разнородных данных, значительная часть которых представлена в текстовой форме. Для эффективного анализа таких данных требуется не только хранение информации, но и понимание её смысловых связей.

Развитие искусственного интеллекта и появление больших языковых моделей (LLM) сделали возможной автоматическую интерпретацию текстов и извлечение знаний. Однако для повышения точности анализа всё чаще применяются графовые базы данных, отражающие структуру взаимосвязей между понятиями.

Объединение LLM с графами знаний привело к появлению технологий Retrieval-Augmented Generation (RAG) [1], сочетающих поиск и генерацию ответов. Такие системы способны извлекать релевантные сведения, формировать контекст и обеспечивать точный анализ данных тематических сайтов и других источников.

I. Графовая база данных и интеллектуальный GraphReader-агент

Для хранения и анализа информации использовалась графовая база данных Neo4j, обеспечивающая представление знаний в виде взаимосвязанных узлов и отношений. Такой формат позволяет не только хранить данные, но и описывать их смысловые связи, что особенно важно при интеграции с языковыми моделями.

Исходные данные загружались с помощью скрипта на Python, где документы преобразовывались в структурированное графовое представление. Каждый документ разбивался на последовательность фрагментов текста (chunks), которые соединялись между собой отношениями NEXT, отражающими порядок следования частей текста. Далее каждый фрагмент анализировался для выделения атомарных фактов (atomic facts) — минимальных утверждений, содержащих конкретные знания.

На нижнем уровне формировались ключевые элементы (key elements), представляющие наиболее значимые понятия и темы документа. Эти элементы объединяли атомарные факты с более широкими концепциями — такими как исторические личности, географические объекты или научные термины. В результате создавался многоуровневый граф знаний, где документы, факты и ключевые элементы связаны между собой в единую семантическую сеть, общее представление которого продемонстрировано на рисунке 1.

Рис. 1 – Общее представление графовой базы данных

Такое представление данных делает возможным не просто поиск по ключевым словам, а анализ контекста и взаимосвязей между объектами. Именно это используется в работе GraphReader-агента — интеллектуального компонента, способного «читать» граф, интерпретируя его структуру при формировании ответов на запросы [2].

GraphReader-агент взаимодействует с Neo4j через библиотеку LangGraph [3], которая обеспечивает управление цепочками запросов и их обработку с помощью LLM. При поступлении запроса пользовательской системы агент выполняет последовательный обход графа: определяет ключевые элементы, наиболее близкие к запросу, анализирует связанные с ними факты и документы, формирует контекст, отражающий смысловые связи, передаёт этот контекст языковой модели для генерации итогового ответа. Структу-

ра GraphReader-агента представлена на рисунке 2.

Рис. 2 – Структура GraphReader-агента

Таким образом, GraphReader сочетает семантический поиск с возможностями языковых моделей, обеспечивая точное и интерпретируемое извлечение знаний. В отличие от традиционных подходов, где информация хранится в линейной форме, использование графа позволяет агенту работать с контекстом на уровне понятий и связей, что делает анализ данных более глубоким и гибким.

II. Интеллектуальная система анализа данных на основе RAG-архитектуры

Построенный граф знаний и агент GraphReader стали основой системы, реализующей принципы Retrieval-Augmented Generation (RAG). Эта архитектура объединяет извлечение релевантной информации из базы данных и генерацию ответов при участии языковой модели. В отличие от классического поиска, где пользователь получает набор документов, RAG-подход формирует связный и осмысленный ответ, основанный на контексте, найденном в графе.

Работа системы начинается с пользовательского запроса на естественном языке. Агент GraphReader, используя встроенные механизмы Neo4j и LangGraph, выполняет графовый поиск, выявляя узлы и связи, наиболее близкие к смыслу вопроса. Он анализирует цепочки отношений между документами, фактами и ключевыми элементами, формируя контекстное подграфовое представление — фрагмент знаний, связанный с запросом.

Затем этот контекст передаётся в LLM-модель, которая синтезирует итоговый ответ. Таким образом, модель не «угадывает» ответ из общей информации, а работает с конкретными, извлечёнными из графа фактами. Это повышает точность, снижает вероятность галлюцинаций и делает результат воспроизводимым.

Особенностью системы является её интерактивный интерфейс. Был разработан веб-сайт, позволяющий пользователю напрямую взаимодействовать с базой знаний. Через него можно загружать новые документы, которые автоматически преобразуются в графовую структуру в Neo4j, а также использовать чат-интерфейс для постановки вопросов. На рисунке 3 представлена работа пользовательского интерфейса системы.

Рис. 3 – Пример работы веб-интерфейса системы анализа данных

Запросы, поступающие через чат, обрабатываются в реальном времени: GraphReader извлекает нужные данные из графа, формирует контекст и передаёт его в языковую модель, которая генерирует ответ в человекопонятной форме. Таким образом, пользователь получает полноценный интеллектуальный инструмент для анализа тематических сайтов, где объединяются точность графовых технологий и гибкость больших языковых моделей

III. Заключение

В работе рассмотрено архитектурное решение системы анализа данных тематических сайтов, основанной на технологиях RAG, LLM и графовой базе данных Neo4j. Использование GraphReader и LangGraph позволило объединить хранение знаний в виде семантического графа и интеллектуальный поиск с генерацией ответов на естественном языке.

Такая архитектура обеспечивает контекстное понимание данных и выявление смысловых связей между понятиями, повышая точность и полноту анализа. Применение RAG-подхода делает систему эффективным инструментом для обработки растущих объёмов информации и ускоряет получение релевантных ответов.

Интеграция графовых технологий и языковых моделей создаёт основу для интеллектуальных аналитических решений, способных глубоко интерпретировать данные и поддерживать принятие решений в условиях сложных информационных потоков.

Список литературы

1. What is Retrieval-Augmented Generation? [Electronic resource]. – Mode of access: `https://aws.amazon.com/whatis/retrieval-augmented-generation/?nc1=hls`. – Date of access: 04.10.25.
2. GraphReader: Building Graph-based Agent to Enhance Long-Context Abilities of Large Language Models [Electronic resource]. – Mode of access: `https://arxiv.org/pdf/2406.14550v1` Date of access: 04.10.25.
3. LangGraph [Электронный ресурс]. – Mode of access: `https://www.langchain.com/langgraph`. – Date of access: 04.10.25.

СОВРЕМЕННЫЕ МЕТОДЫ РЕШЕНИЯ ЗАДАЧ ЛОГИСТИКИ С ПОМОЩЬЮ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ

Боброва Н. Л., Богомаз К. С.
Кафедра информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: kiril6701@gmail.com

Данная работа посвящена анализу современных методов решения задач логистики с помощью систем поддержки принятия решений (СППР). Современные системы общественного транспорта функционируют в условиях высокой сложности, неопределенности и динамичности, что требует применения передовых инструментов для планирования и управления. СППР используют математическое моделирование, оптимизацию и интеллектуальный анализ данных для повышения эффективности.

Введение

Современные системы общественного транспорта сталкиваются с проблемами высокой сложности и неопределенности, что требует применения передовых инструментов для планирования и управления. Ключевым решением в этой области является разработка и внедрение систем поддержки принятия решений (СППР), которые используют математическое моделирование и выявления закономерностей в данных для повышения эффективности. Целью данной работы является анализ современных методов решения задач оптимизации маршрутов и расписаний общественного транспорта, планирование, изменение и оптимизация графиков работы персонала и стратегическое планирование направлений и маршрутов в логистике с помощью систем поддержки принятия решений.

I. Транспорт по требованию (DRT)

Для оптимизации расходов при низком спросе хорошо зарекомендовал себя транспорт по требованию (Demand Responsive Transportation, DRT) который представляет собой гибридный подход между такси и автобусом. Данная модель хорошо показывает себя при низком, переменном или непредсказуемом спросе, например, в сельской местности [1]. Характеристиками задачи являются ограниченная вместимость транспортных средств, обслуживание запросов в определенных временных окнах, количество остановок и время в пути, минимизация эксплуатационных расходов и максимизация качества обслуживания. Поскольку DRT-системы призваны решать многокритериальные задачи, они не могут быть эффективно решены традиционными методами оптимизации. Эти проблемы сложны, особенно для решения в реальном времени.

II. Ростеринг

Планирование графиков работы персонала является критически важной задачей после составления расписания движения. При планировании графиков необходимо распределить конкретные смены между отдельными водителями с учетом их квалификации, личных предпочтений, а также сложных правил и нормативных актов. Составление ростеров - это процесс, который следует за этапом составления расписания экипажей. На этапе расписания экипажей генерируются анонимные смены, тогда как ростеринг занимается фактическим назначением этих смен водителям. Ростер представляет собой итоговое расписание для каждого водителя или группы водителей. Для поиска решений используются оптимизационные решатели (например, коммерческие MIP-решатели) или колоночная генерация (Column Generation, CG) для большого объема параметров и данных [2].

Циклический ростеринг помогает в планировании графиков для водителей с одинаковой квалификацией. Метод работает быстро и обеспечивает простоту генерации и справедливое распределение обязанностей, непопулярных смен и выходных в группе. Недостатками метода является отсутствие гибкости, метод не учитывает желания водителей и праздничные дни.

Нециклический ростеринг решает проблему циклического ростеринга, но алгоритм становится сложнее т.к. появляется необходимость учитывать потребности всех водителей.

Предыдущие 2 метода решали задачу генерации графиков. Для поиска оптимальных расстановок используются другие алгоритмы, например, коммерческие MIP-решатели или колоночная генерация для большого объема параметров и данных [2]. MIP-решатели подходят для задач, не требующих большого вычислительного времени. Минусом данного решателя является не возможность его масштабирования на объемы данных больше 200 тысяч столбцов и 150 тысяч строк.

Генерация столбцов (Column Generation, CG) решает задачи, которые не могут быть решены MIP-решателями из-за ограничений по времени или объема данных. Недостаток данного алгорит-

ма, что он в процессе выполнения может исчерпать лимит доступной памяти.

Метаэвристики решают проблемы всех предыдущих решений. Они работают достаточно быстро, используют мало памяти и подходят для большого объема данных. Недостатком является то, что качество полученных ростеров уступает CG и MIP-решателям.

III. Стратегическое планирование транспортных сетей

Эффективное стратегическое планирование, такое как открытие новых маршрутов, требует точного понимания спроса. Однако одной из главных проблем является определение точных начальных и конечных точек пассажиров или матрицы корреспонденции. Статистика перевозок между пунктами не отражает, какая часть пассажиров едет транзитом. Метод нечетких матриц корреспонденций, помогает в планировании и принятии обоснованных и взвешенных решений путем восстановления нечетких матриц [3]. С помощью данного метода и нечетких чисел можно получить интервальные и вероятностные оценки по направлениям. С помощью этих оценок маршрутный менеджер сможет принимать взвешенные решения об открытии новых или закрытии старых направлений.

IV. Оперативное диспетчерское управление

Диспетчерское управление общественным транспортом в режиме реального времени - это сложная задача, характеризующаяся неполнотой информации, быстро меняющейся дорожной обстановкой и необходимостью учета множества факторов. Для поддержки диспетчеров разрабатываются интеллектуальные системы поддержки принятия решений. Типовые дорожные ситуации, например, небольшая задержка из-за ДТП или значительное скопление пассажиров на остановке описываются как нечеткие ситуации с помощью лингвистических переменных и функций принадлежности и записывается в базу знаний [4]. При возникновении чрезвычайной ситуации на маршруте система сравнивает ситуацию с ситуациями, находящимися в базе знаний, и на основе самой подходящей ситуации/ситуаций подсказывает оператору варианты её решения. Данный подход значительно повышает оперативность, обоснованность и правильность принятия решений в реальном времени.

V. Заключение

Анализ представленных источников демонстрирует как системы поддержки принятия решений повышают эффективность и надежность общественного транспорта. Эти системы решают широкий спектр задач: от стратегического планирования маршрутной сети до управления движением и реагированием на чрезвычайные ситуации в реальном времени. Практика показывает, что оптимизация, имитационное моделирование, методы нечеткой логики позволяют управлять общественным транспортом при скачках спроса и неполных данных. Развитие таких СППР является необходимым условием для создания устойчивого и качественного общественного транспорта.

VI. Список литературы

1. Gomes, R. Pinho de Sousa, J. Teresa Galvão Dias Design and operation of Demand Responsive Transportation systems / R. Gomes, J. Pinho de Sousa // Compendium of Papers of the XV Euro Working Group on Transportation Meeting. – Paris. France – 2012. – Режим доступа: https://www.academia.edu/10242158/Design_and_operation_of_Demand_Responsive_Transportation_systems_Compendium_of_Papers_of_the_XV_Euro_Working_Group_on_Transportation_Meeting_Paris_France_September_2012. – Дата доступа: 19.10.2025.
2. Xie, L. Kliewer, N. Suhl, L A web-based Decision Support System for Crew Rostering in Public Transport. / L. Xie [и др.]. // University of Paderborn Faculty of Business and Economics Department of Information Systems Decision Support & Operations Research Lab. – Paderborn. Germany – 2013. – Режим доступа: https://www.researchgate.net/profile/Lin-Xie-3/publication/326776795_A_web-based_Decision_Support_System_for_Crew_Rostering_in_Public_Transport/links/5b62fa77a6fdccf0b207adb4/A-web-based-Decision-Support-System-for-Crew-Rostering-in-Public-Transport.pdf. – Дата доступа: 19.10.2025.
3. Нестеров, В. А. Судаков, В. А. Сыпало, К. И. Титов, Ю. П. Матрица нечетких корреспонденций модели авиационных перевозок / В. А. Нестеров [и др.]. // ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ. – 2022. – № 6. – С. 95–102.
4. В. А. Лахно, В. М. Собченко Модель автоматизированной системы поддержки принятия решений диспетчерского управления городским пассажирским автотранспортом / В. А. Лахно, В. М. Собченко. // Наука та прогрес транспорту. Вісник Дніпропетровського національного університету залізничного транспорту. – 2016. – № 2. – С. 61–69.

МЕТОДЫ ОЦЕНКИ ДОВЕРИЯ И ПРОЗРАЧНОСТИ РЕШЕНИЙ ИИ-СИСТЕМ

Вакарь Е. И., Коростелёв А. Д., Романюк М. В.
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {e.vakar, a.korostelev, romanuk}@bsuir.by

В статье рассматриваются современные подходы к оценке доверия и прозрачности решений систем искусственного интеллекта. Приведена классификация методов объяснимого ИИ (XAI), описаны метрики доверия и уверенности в предсказаниях моделей, а также анализируются факторы, влияющие на восприятие прозрачности пользователями. Отмечается актуальность разработки интерпретируемых решений для повышения безопасности и надёжности применения ИИ в критически важных областях.

Введение

Развитие искусственного интеллекта (ИИ) приводит к широкому внедрению интеллектуальных систем в различных сферах – от медицины до образования. Однако принятие решений такими системами зачастую непрозрачно для пользователя, что вызывает сомнения в достоверности и надёжности их выводов. В связи с этим возрастает необходимость разработки методов, обеспечивающих доверие к ИИ и прозрачность его решений [1].

I. Методы объяснимости и интерпретации

Современные методы объяснимого искусственного интеллекта (XAI, Explainable Artificial Intelligence [2]) делятся на локальные и глобальные. Локальные методы направлены на объяснение отдельных решений модели, то есть того, почему именно в данном случае ИИ выдал определённый результат.

К числу наиболее распространённых относятся:

- LIME (Local Interpretable Model-agnostic Explanations) – создаёт упрощённую линейную модель, приближающую поведение сложной модели в окрестности конкретного примера. Позволяет определить, какие признаки оказали наибольшее влияние на конкретный прогноз.
- SHAP (Spapley Additive Explanations) – основан на теории игр и вычисляет «вклад» каждого признака в итоговое решение модVели, что обеспечивает теоретическую интерпретируемость и универсальность подхода.
- Grad-CAM (Gradient-weighted Class Activation Mapping) – используется в свёрточных нейронных сетях для визуализации областей изображения, которые повлияли на классификацию, что особенно полезно для анализа моделей компьютерного зрения.

Глобальные методы ориентированы на интерпретацию поведения модели в целом. К ним относятся:

- Surrogate Models – построение интерпретируемой модели (например, дерева решений), которая аппроксимирует поведение сложной модели «чёрного ящика».
- Feature Importance – оценка значимости признаков, используемых моделью при обучении. Такой анализ помогает выявить наиболее критичные факторы, влияющие на решения системы.
- Partial Dependence Plots (PDP) и Accumulated Local Effects (ALE) – визуализируют зависимость прогнозов от отдельных признаков или их комбинаций, позволяя анализировать взаимное влияние входных данных.

II. Методы оценки доверия

Доверие к системам искусственного интеллекта является ключевым фактором их успешного внедрения и принятия пользователями. Если объяснимость (XAI) помогает понять почему модель выдала конкретный результат, то методы оценки доверия позволяют количественно и качественно определить, насколько можно полагаться на это решение [3].

Выделяют два основных аспекта доверия:

1. Техническое доверие – оценка корректности, устойчивости и предсказуемости поведения модели.
2. Психологическое (пользовательское) доверие – восприятие надёжности и справедливости системы со стороны человека.

Классическим подходом технической оценки доверия является анализ уверенности модели (confidence) – вероятностной оценки, выдаваемой системой вместе с результатом предсказания. Однако высокая уверенность не всегда гарантирует правильность ответа. Поэтому применяются дополнительные метрики и подходы:

- Оценка калибровки модели – проверка, насколько вероятностные оценки совпадают

с фактическими результатами. Хорошо откалиброванная модель с уверенностью 0.8 должна быть права примерно в 80% случаев. Для анализа используются такие показатели, как Expected Calibration Error (ECE) и Brier Score.

- Методы оценки неопределённости (uncertainty estimation) – позволяют количественно описывать степень неуверенности модели в своём предсказании. Применяются байесовские нейронные сети, Dropout-аппроксимация, ensemble-модели и bootstrapping.
- Оценка устойчивости (robustness) – анализ того, как изменяются результаты при незначительных изменениях входных данных. Высокая чувствительность к шуму свидетельствует о низком уровне доверия.
- Методы детектирования аномалий и выбросов – позволяют выявить ситуации, в которых модель работает вне обучающего распределения, что повышает осознанность системы относительно собственных ограничений.

Немаловажным аспектом является восприятие доверия со стороны человека, взаимодействующего с ИИ-системой. Даже технически надёжная модель может вызвать недоверие при недостатке прозрачности интерфейса или отсутствии объяснений.

Для анализа этого аспекта применяются:

- Опросы и UX-исследования – оценка субъективного восприятия доверия, удобства, справедливости и понятности решений системы.
- Методы интерактивного XAI – предоставление пользователю возможности запросить объяснение, проверить альтернативные сценарии и наблюдать реакцию модели на изменения данных.
- Визуальные интерфейсы доверия (Trust Dashboards) – отображают уверенность модели, степень неопределённости и происхождение данных, что повышает прозрачность системы в реальном времени.

Современные исследования направлены на встраивание доверительных механизмов непосредственно в архитектуру ИИ-систем. Применяются подходы, при которых модель одновременно обучается на задачу предсказания и задачу оценки собственной уверенности (self-assessment). Кроме того, развивается направление Trustworthy AI, объединяющее методы XAI, оценку неопределённости, устойчивость к атакам и этические аспекты использования ИИ.

Таким образом, оценка доверия представляет собой комплексную задачу, включающую технические, поведенческие и когнитивные аспекты. Эффективная система доверия должна не только корректно предсказывать результаты, но и уметь аргументировать их, демонстрировать устойчивость и давать пользователю возможность понять границы своей надёжности.

III. Современные тенденции

С ростом популярности больших языковых моделей (LLM), таких как GPT, всё актуальнее становятся вопросы прозрачности их ответов и интерпретируемости рассуждений. В отличие от традиционных алгоритмов, LLM используют масштабные текстовые корпуса и вероятностные зависимости, что усложняет понимание логики их работы. Активно исследуются подходы к объяснению вывода таких моделей, включая трассировку источников знаний, анализ цепочек рассуждений и визуализацию внутренних представлений. Эти методы повышают уровень доверия и делают процесс генерации более прозрачным [4].

Также активно развивается нормативное регулирование доверия к ИИ. Европейский AI Act определяет требования к прозрачности и контролю решений моделей в зависимости от степени риска их применения [5].

Заключение

Прозрачность и доверие являются ключевыми характеристиками безопасного и ответственного использования ИИ. Перспективными направлениями остаются разработка универсальных метрик доверия, интеграция XAI-инструментов в промышленные системы и создание стандартов оценки объяснимости решений.

Дальнейшее развитие в этой области требует тесного взаимодействия исследователей, разработчиков и регуляторов. Необходимы единые подходы к оценке интерпретируемости и надёжности моделей, а также внедрение принципов этичного ИИ на этапах проектирования и эксплуатации. Совместное развитие технологий, методологий и нормативных требований позволит сформировать экосистему искусственного интеллекта, основанную на прозрачности, ответственности и уверенности пользователей в результатах его работы.

IV. Список литературы

1. Long, B. Explainable AI the Latest Advancements and New Trends / B. Long, E. Liu, R. Qiu, Y. Duan // arXiv preprint arXiv:2505.07005, 2025.
2. Saarela, M. Recent Applications of Explainable AI (XAI): A Systematic Literature Review / M. Saarela, V. Podgorelec. //Appl. Sci. 2024
3. Zahra Atf. Is Trust Correlated With Explainability in AI? A Meta-Analysis. / Zahra Atf, P. R. Lewis // arXiv:2504.12529, 2025
4. Palikhe, A. Towards Transparent AI: A Survey on Explainable Large Language Models. / A. Palikhe, Z. Yu, Z. Wang, W. Zhang // arXiv preprint arXiv:2506.21812, 2025
5. European Commission / Artificial Intelligence Act – Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024

ФИЗИЧЕСКИЕ ПАРАДОКСЫ И СВЕРХМАЛЫЕ ВЕЛИЧИНЫ

Герман О. В., Герман Ю. О.
Факультет информационных технологий и управления,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: juliagerman@gmail.com

Рассматривается логика,связывающая возможность и вероятность. Возможность ассоциируется с сингулярным(-ми) событием(-иями), вероятность – с материализованной возможностью. Приводимые аксиомы могут быть использованы в построении рассуждений о прошлом с позиций настоящего, что расширяет возможности и приложения искусственного интеллекта в части обобщения прошлого опыта, а также дает объяснение известным физическим парадоксам.

Введение

В ряду неклассических логик как области искусственного интеллекта видное место занимают логики, связанные с анализом возможности и вероятности. Здесь следует отметить модальную логику и ее существующие модификации, которые оперируют модальностями возможности и необходимости и позволяют весьма общим образом связать рассуждения о формулах с их вероятностями. Говоря, что формула возможна в некотором модельном мире, имеют в виду, что ее вероятность не равна нулю в этом мире. Можно, однако, предложить иной взгляд, согласно которому если даже возможность не равна нулю, то вероятность выполнения формулы нулевая. Такая ситуация возникает, когда вероятности оцениваются a posteriori, например, относительно имевших место событий, а возможности оцениваются a priori до возникновения событий. Применительно к искусственному интеллекту такая логика позволяет строить рассуждения в настоящем о прошлом, что, очевидно представляет интерес. Это распространяется также на объяснение физических парадоксов, например, на парадокс Зенона о движущемся теле: *Если тело движется, то оно движется либо в том месте, которое оно занимает, либо в том месте, где его нет. Однако тело не может двигаться в том месте, которое оно занимает (так как в каждый момент времени оно занимает все это место), но тело также не может двигаться и в том месте, где его нет. Следовательно, движение тела невозможно.*

Объяснение этого парадокса основано на рассмотрении гипермалых чисел (величин, меньших любых действительных чисел), то есть вероятность нахождения тела в гиперамалой области равна нулю, а возможноость не равна нулю.

I. Вероятность как материализованная возможность

Это рассуждение выводит в область гипермалых (разумеется, и гипербольших) чисел, теорию которых разработал А. Робинсон в 60-х годах прошлого века [1].

Каждое гипермалое число меньше любого действительного числа. Это очень сильное допущение [2]. Если говорить о физических событиях и их вероятностях, то может ли вероятность быть гипермалым числом? Полагаем, что нет, так как физическое событие происходит в физическом мире, а из парадокса Зенона вывели, что летящее тело попадает при движении в области, где нет физического пространственно-временного мира. Физическое событие не может произойти вне пространства и времени. Итак, гипермалой вероятности нет, так как нет нужной событийной интерпретации в физическом мире.

До появления объекта должна была существовать возможность данного события. Следовательно, различаем возможность и вероятность как материализованную возможность. События и объекты материальной Вселенной до ее появления имели характер возможности.Энергия сама по себе создает возможность бесчисленного множества возможных сценариев. Материальная Вселенная – один из таких сценариев.

Пусть $P(A)$ – вероятность явления (события) A, $O(A)$ – возможность события A. Принимаем такие аксиомы, где nil означает ничто, ˜есть логическое отрицание; nonex(A) – не существующий объект в мире (либо не появившийся, либо исчезнувший, канувший в небытие). Кроме того, и это крайне важно, мы рассматриваем бесконечность как становящуюся (не завершенную). Становящаяся бесконечность – это то, что никогда не будет завершено. Перейдем к аксиомам.

II. Аксиомы

0.

$$\forall A,B\ \ (O(A \vee B)) \geq max(O(A),O(B));$$

$$O(A \vee \mathrm{nil}) = O(A); \quad O(A) \geq 0;$$

$$O(A \wedge \mathrm{nil}) = 0; \quad O(\mathrm{nil}) = 0;$$

$O(\sim \text{nil}) > 0; \quad O(A) = c{\cdot}O(A)$ ($c =$ const).

1. $P(A) > 0 \Rightarrow O(A) \neq 0$.
2. $O(A) = 0 \Rightarrow P(A) = 0; \quad A = \text{nil} \Leftrightarrow O(A) = 0$.
3. $\forall A\, O(A) \neq 0 \Rightarrow O(A) < P(A)$.
4. a) $O(A) \neq 0 \Rightarrow \infty \cdot O(A) = P(A) > 0$;

$$\frac{P(A)}{\infty} = O(A), \quad \frac{P(A)}{O(A)} = \infty.$$

   b) $O(A) = 0 \Rightarrow \infty \cdot O(A) = 0$.

5. $A \subseteq B \Rightarrow O(A) \geq O(B)$.
6. $O(A \to B) > 0 \Rightarrow O(B) \geq O(A)$.
7. $A \subseteq B \wedge O(A \to B) > 0 \Rightarrow O(B) = O(A)$.
8. $A = B \Rightarrow O(A) = O(B)$.
9. $O(\text{nil} \wedge A \neq \text{nil}) = 0$.
10. $O(A \wedge B) > 0 \Rightarrow O(A) > 0 \wedge O(B) > 0$.
11. a) $O(A|B) > 0 \Rightarrow O(A) > 0$.
    b) $O(A) = O(A \mid B) + O(A \mid \neg B) - O(B \wedge \neg B)$.
12. $P(A) = 0 \Rightarrow O(A) > 0 \vee A = \text{nil}$.
13. $O(A) > 0 \wedge \text{nonex}(A) \Rightarrow A \neq \text{nil} \wedge P(A) = 0$.
14. $A = \text{nil} \Rightarrow \text{nonex}(A)$.
15. $\text{nil} \in A \Rightarrow O(A) = 0$.
16. $P(A) = 0 \Rightarrow \text{nonex}(A)$

III. Обсуждение

Аксиома 0 говорит, что возможность дизъюнкции двух событий не ниже возможности любого из этих двух. Возможность либо есть, либо ее нет. Что касается конъюнкции, то мы вводим аксиому 10. Далее, ничто (nil) не имеет возможности, следовательно, в физическом мире nil не проявляется. $O(A) + O(A) = O(A)$. Возможность не увеличивается! $O(\sim \text{nil}) > 0$ нечто существует в возможности. Аксиома 1 говорит, что если вероятность A больше 0, то возможность A не равна 0. Аксиома 2 говорит, что если возможность A равна 0, то событие A в физическом мире не состоится (даже при бесконечном числе испытаний – см. аксиому 4).

Аксиома 3 говорит, что ненулевая возможность всегда меньше физической вероятности, какой бы малой эта физическая вероятность ни была. Это значит, что возможность есть гипермалое число или 0. Аксиома 4a говорит, что если возможность не равна 0, то в бесконечном числе испытаний событие может возникнуть по физической вероятности (т.е. появится ненулевая физическая вероятность события из гипермалой возможности). Аксиома 4b говорит, что если возможность равна 0, событие в физическом смысле не возникнет никогда. Полагая возможность $O(A \wedge \neg A) = 0$, получаем вероятность $P(A \wedge \neg A) = 0$. Однако, допуская $O(A \wedge \neg A) \neq 0$, в бесконечном числе испытаний возможна ненулевая физическая вероятность $P(A \wedge \neg A) \neq 0$. Такая трактовка не выполняет аксиомы классической теории вероятностей, но не нарушает физического смысла (например, законов квантового мира). Кроме того, пусть $P(A) < P(B)$. Тогда

$$\frac{O(A)}{\infty} < \frac{O(B).}{\infty}$$

Но мы не можем отсюда заключить, что $O(A) < O(B)$, поскольку наше представление о бесконечности вполне допускает, что $\infty + \infty = \infty$, и $c \cdot \infty = \infty$. Поэтому $\frac{O(A)}{\infty} = \frac{2 \cdot O(A)}{\infty}$, $O(A) = 2 \cdot O(A)$. Из этого следует, что $O(A)$ не есть вещественное число в обычном смысле, иначе было бы $O(A) < 2 \cdot O(A)$. Из этого далее следует, что нельзя прямо отождествить возможность с истиной (TRUE), а невозможность с ложью (FALSE), поскольку смысл TRUE $\uparrow \infty$ вообще отсутствует. Аксиома 5 говорит, что возможность части не ниже возможности целого, содержащего эту часть. Аксиома 6 говорит, что если событие A невозможно без B, то возможность B не ниже возможности A. Аксиома 7 говорит, что если A невозможно без B и A является частью B, то возможности A и B равны (противное приводит к нарушению либо аксиомы 5, либо аксиомы 6). Пример. Каждый человек есть часть мира. Но без мира нет человека. Согласно аксиоме 7, возможности мира и данного человека в нем одинаковы. Аксиома 9 означает, что ничто не дает возможности ничему. Аксиома 10 означает, что если возможен составной объект, то возможны его части. Аксиома 11 говорит, что если событие возможно при условии, то событие просто возможно.Аксиома 12 говорит, что если объекта нет в физическом мире, то его возможность в трансцендентном мире ненулевая или возможности объекта нет вообще.

IV. Заключение

Итак, трансцендентный мир нами ассоциируется с огромной энергией, которая при взрыве «вырывается» из точки и приводит к формированию физической Вселенной. Из аксиом следует, что $A = \text{nil} \to P(A) = 0$. Невозможное не случится ни при каких обстоятельствах ни в каком мире. Однако любой возможный сценарий каким-то образом «знает» ограничения (читай, законы) связанного с этим сценарием физического мира.

.

V. Список литературы

1. Robinson, A. Non-standard Analysis. / A. Robinson. // Princenton Univ. Press, 1996, 158p.
2. Bell, J. The Contiuous and the Infinitesimal in Mathematics and Philosophy / J. Bell // Polimetrica, 2005, 349p.

К ВОПРОСУ НАЗНАЧЕНИЯ ВОДИТЕЛЕЙ НА РЕЙСЫ ГОРОДСКОГО ОБЩЕСТВЕННОГО ТРАНСПОРТА НА ОСНОВЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ

Гончаров С. В., Войтешенко И. С.
Кафедра компьютерных технологий и систем, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: goncharov.sergei.21@gmail.com, voit@bsu.by

В статье представлен подход к решению задачи покрытия расписаний городского общественного транспорта путём назначения водителей на автобусные рейсы с учётом ограничений рабочего времени на основе обучения с подкреплением. Описан алгоритм обучения агента, использующий метод оптимизации проксимальной политики (PPO) для принятия решений о назначении водителей на рейсы.

Введение

Эффективное назначение водителей на рейсы с учётом ограничений, обусловленных трудовым законодательством, является одной из множества задач современного управления транспортными системами и напрямую влияет на операционные расходы транспортных компаний. Классические методы решения задачи планирования можно разделить на две основные категории: точные подходы и эвристические алгоритмы [1-2]. При этом точные методы обычно требуют значительных вычислительных ресурсов и применимы к задачам небольшого размера. Эвристические подходы позволяют решать задачи большего масштаба, но не гарантируют оптимальности решения.

Основной проблемой традиционных методов является их неспособность адаптироваться к изменяющимся условиям в режиме реального времени. Обучение с подкреплением отличается тем, что не строит статичный план заранее, а формирует адаптивную стратегию (политику) действий агента, что позволяет накапливать опыт и адаптироваться к внезапным изменениям среды, успешно решая задачи в реальном времени [3].

I. Построение MDP модели

Задачу назначения водителей на рейсы будем рассматривать как марковский процесс принятия решений (MDP) с определённым пространством состояний, пространством действий и функцией награды. Будем частично опираться на предложенный вариант модели в источнике [4]. Каждое время отправления в схеме расписания рассматривается как точка принятия решения. Метод обучения с подкреплением оценивает состояние окружения и выбирает водителя для выполнения рейса в данной точке.

Будем рассматривать расписание, состоящее только из автобусных маршрутов. Маршруты состоят из рейсов в обе стороны. Каждому рейсу соответствует пара контрольных остановок (начальная и конечная). Ограничения включают рабочее время водителя $work_{max}$, минимальное время $lunch_{min}$ в пути до обеда и длительность обеда $lunch_{dur}$, минимальное время отдыха $rest_{min}$ после выполнения рейса.

Пространство состояний. Пространство состояний в точке принятия решения состоит из параметров V_s всех контрольных остановок, параметров v_l рейса, для которого принимается решение, и параметров водителей V_d, которые являются наилучшими кандидатами для выполнения данного рейса.

Контрольная остановка s описывается следующими параметрами:

- Номер остановки n^s;
- число n^s_{lt} предстоящих рейсов с отправлением с остановки s;
- число n^s_{st} предстоящих рейсов в следующие M минут (отражает спрос в ближайшее время; M является гиперпараметром);
- число водителей n^s_d, готовых выйти в рейс в данный момент времени с остановки s.

Номер остановки n^s нужен для установления связи с рейсом. Признаки n^s_{lt} и n^s_{st} используются для получения спроса на водителей для данной контрольной остановки в будущем. Признак n^s_d необходим для получения информации о числе водителей, готовых удовлетворить запрос на выход в рейс. Вектор V_s соответственно является конкатенацией векторов параметров всех контрольных остановок.

Каждой точке принятия решений соответствует рейс l, описываемый параметрами:

- Номер начальной остановки d^l;
- номер конечной остановки a^l;
- время в пути h^l.

Каждый водитель d в точке принятия решения характеризуется следующим параметрами:

- время отдыха водителя с последней поездки r^d ($r^d = 0$, если водитель в рейсе);
- номер контрольной остановки s^d, где находится водитель ($s^d = -1$, если водитель в рейсе);
- время t^d холостой поездки в контрольную остановку, из которой принимается решение о назначении водителя на рейс;
- время t^d_w, прошедшее с выхода водителя;

– статус l^d водителя на обеде (0 или 1).

Когда $t_w^d = lunch_{min}$, у водителя начинается обед. Если в этот момент водитель находился в рейсе ($r^d = 1$), обед начинается после рейса и длится $lunch_{dur}$ минут.

Для облегчения принятия решения предварительно формируем список водителей, которых можно назначить на рейс. В список не включаются водители, которые:

– находятся в рейсе: $r^d = 0$;
– не успевают прибыть в точку отправления к моменту начала рейса: $r^v \leq rest_{min} + t^d$;
– находятся на обеде: $l^d = 1$;
– время поездки превышает оставшееся рабочее время: $h^l \geq work_{max} - t_w^d$.

Оставшееся множество водителей разбиваем на две группы (по убыванию приоритета):

1. Водители, для которых не требуется холостая поездка (отсортированные внутри группы по убыванию r^d);
2. водители, для которых требуется холостая поездка (отсортированные внутри группы по возрастанию t^d).

После разбиения выбираем K водителей из первой группы в порядке приоритета (K является гиперпараметром). Если размер первой группы меньше K, то дополняем водителями из второй группы в порядке приоритета. Вектор V_d формируется путём конкатенации векторов параметров выбранных водителей.

Пространство действий. Пространство действий состоит из множества водителей A, предоставленных для назначения на рейс в текущей точке принятия решения, то есть действие агента $a \in A$. Также в данное пространство действий включено решение о выходе на рейс нового водителя. То есть агент или выбирает водителя из предложенных, или принимает решение о выходе нового.

Функция награды. Функция награды сочетает финальное и пошаговое вознаграждения [4]. Финальное вознаграждение задаётся формулой:

$$r_{main} = -w_1 N - w_2 T$$

где N – общее количество задействованых водителей, T – общее время, потраченное на холостые поездки. Пошаговое вознаграждение вычисляется следующим образом:

$$r_{step} = -w_1 r_n - w_2 t^d + w_3 r_r - w_4 r_u$$

где r_n – штраф за выход нового водителя, t^d – штраф за необходимость совершить холостую поездку в точку отправки (время холостой поездки), r_r – поощрение за назначение водителей, которые долго отдыхают, r_u – дополнительный штраф, если для рейса был выбран водитель, которому необходимо совершить холостую поездку из контрольной остановки с большим спросом на рейсы. В обоих формулах параметры w_i являются гиперпараметрами.

II. Алгоритм обучения

Алгоритм обучения основан на методе оптимизации проксимальной политики (PPO) с архитектурой actor-critic. Система включает три компонента: нейронную сеть для извлечения признаков (State-Net), сеть актора (Actor-Net) и сеть критика (Critic-Net) [4-5].

Агент взаимодействует с симулированной средой, собирает траектории опыта, вычисляет преимущества действий и обновляет политику агента $\pi_\theta(a|s)$ [5].

III. Исходные данные

Для реализации данного подхода необходимы датасеты, которые содержат информацию о транспортных маршрутах (какие рейсы включают, начальные и конечные остановки, длительности рейсов), расписаниях рейсов (времена отправления), длительностях поездок между контрольнми остановками (для вычисления длительности холостых поездок).

Такие датасеты могут быть синтетическими [6] или собранными из открытых источников.

IV. Выводы

Предложен подход к решению задачи назначения водителей на рейсы городского общественного транспорта на основе обучения с подкреплением. Представленная MDP-модель формализует процесс принятия решений с учётом основных ограничений рабочего времени. Использование метода оптимизации проксимальной политики (PPO) позволяет обучить агента, способного формировать адаптивную стратегию назначения, минимизируя общее время холостых поездок и количество задействованных водителей. Дальнейшие шаги развития данного подхода могут включать учёт более сложных и комплексных рабочих ограничений.

1. Pardo-Peña, D., Álvarez-Martínez, D., & Escobar, J. W. (2024). A GRASP algorithm for the bus crew scheduling problem. International Journal of Industrial Engineering Computations, 15(2), 443–456. https://doi.org/10.5267/j.ijiec.2024.1.003
2. Árgilán, V. S., & Békési, J. (2025). Optimizing Bus Driver Scheduling: A Set Covering Approach for Reducing Transportation Costs. Applied System Innovation, 8(5), 122. https://doi.org/10.3390/asi8050122
3. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: a survey. Journal of Artificial Intelligence Research, 4, 237–285. https://doi.org/10.1613/jair.301
4. Liu, Y. (2024). RL-MSA: a Reinforcement Learning-based Multi-line bus Scheduling Approach. arXiv preprint arXiv:2403.06466
5. Grondman, I., Busoniu, L., Lopes, G. a. D., & Babuska, R. (2012). A survey of Actor-Critic Reinforcement Learning: Standard and Natural Policy gradients. IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews), 42(6), 1291–1307. https://doi.org/10.1109/tsmcc.2012.2218595
6. Haase, K., Desaulniers, G., & Desrosiers, J. (2001). Simultaneous vehicle and crew scheduling in urban mass transit systems. Transportation Science, 35(3), 286–303. https://doi.org/10.1287/trsc.35.3.286.10153

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ РЕШЕНИЯ ЗАДАЧ УПРАВЛЕНИЯ В ОБРАЗОВАТЕЛЬНОЙ СФЕРЕ

Гракова Н. В.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: grakova@bsuir.by

В статье проводится обзор основных задач управления в образовании, а также анализируются современные тенденции и направления. Рассматриваются направления применения интеллектуальных систем для повышения качества образования.

Введение

Современное образование находится на этапе активной цифровизации, что обусловливает необходимость разработки инновационных методов и технологий для повышения эффективности организации образовательного процесса. В данном случае особое значение приобретает применение семантических технологий, автоматизации процессов и создания интеллектуальных систем, что позволяет значительно повысить качество, персонализацию и масштабируемость образовательных проектов.

Управление в образовательной сфере – это систематическая деятельность, направленная на организацию, координацию и контроль образовательных процессов с целью обеспечения высокого качества обучения и развития образовательных учреждений.

Недостаточная адаптивность к современным тенденциям в сфере информационных технологий, ограниченность ресурсов, низкая мотивация персонала, задействованного в оказании образовательных услуг, сложность внедрения инноваций, а также недостаточная эффективность систем оценки и контроля снижают качество образовательного процесса.

Эффективное управление способствует созданию условий для реализации образовательных программ, профессионального роста педагогов и успешного развития учащихся.

I. Основные задачи управления

Рассмотрим основные задачи управления в образовании [1]:

- обеспечение качества образования;
- повышение профессионального уровня преподавателей (учителей);
- создание безопасной и комфортной среды для обучения;
- внедрение инновационных технологий и методов обучения.

К ключевым аспектам, обеспечивающим эффективное функционирование и развитие системы образования относятся [1]:

1. *Планирование и стратегическое развитие* – формулирование долгосрочных и краткосрочных целей для развития образовательных учреждений.
2. *Организация учебного процесса* – создание условий для эффективного обучения, оптимизация учебных программ, управление ресурсами (кадровыми, материальными, информационными), а также обеспечение доступности и инклюзивности образования.
3. *Контроль и оценка качества* – систематический мониторинг результатов обучения, внедрение современных систем оценки и обратной связи.
4. *Управление персоналом* – подбор, повышение квалификации, удержание квалифицированных педагогических и административных сотрудников системы образования.
5. *Внедрение инноваций* – активное использование современных технологий, методов и подходов в управлении и обучении, стимулирование инновационной деятельности и экспериментов для повышения эффективности образовательного процесса.
6. *Обеспечение соответствия стандартам и нормативам* – соблюдение законодательства, стандартов качества и этических норм.

Эти задачи являются основой для обеспечения высокого качества образования, повышения его доступности и адаптивности к современным требованиям, а также для устойчивого развития системы образования в целом.

II. Современные тенденции управления образованием

К современным тенденциям в управлении образованием можно отнести [1]:

- внедрение цифровых технологий и электронных образовательных платформ;
- развитие дистанционного и гибридного обучения;
- применение инновационных подходов к управлению качеством образования;
- повышение роли обратной связи и участия всех участников образовательного процесса.

Рассмотрим более подробно основные направления в управлении образованием [1]:

1. *Цифровизация и внедрение современных технологий* – активное использование онлайн-платформ, систем управления обучением (LMS), виртуальных классов и мобильных приложений для повышения доступности и интерактивности образовательного процесса.
2. *Индивидуализация обучения* – разработка адаптивных учебных программ, учет индивидуальных потребностей, навыков и особенностей каждого обучающегося.
3. *Инновационные методы оценки знаний* – использование портфолио, проектных работ и автоматизированных систем оценки, обеспечивающих более объективную и комплексную оценку достижений обучающихся.
4. *Обмена опытом* – участие в международных проектах и программах, что способствует внедрению передовых практик и расширению возможностей для студентов и преподавателей.

Эти тенденции способствуют повышению эффективности, доступности и качества образования, а также подготовке студентов к современным требованиям рынка труда.

III. Применение интеллектуальных технологий в сфере образования

Применение интеллектуальных технологий в сфере образования становится всё более актуальным, открывая новые возможности для повышения эффективности обучения и управления образовательным процессом.

Рассмотрим основные направления применения интеллектуальных технологий в образовании [2,3]:

1. *Персонализация обучения.* Интеллектуальные системы (ИС) позволяют адаптировать учебный материал под индивидуальные потребности и уровень каждого студента. Например, системы адаптивного обучения анализируют прогресс учащегося и подбирают задания, соответствующие его знаниям и стилю обучения.
2. *Автоматизация оценки и контроля знаний.* Использование автоматизированных систем тестирования и оценки позволяет быстро и объективно проверять знания студентов, а также выявлять слабые места в их подготовке.
3. *Обучающие чат-боты и виртуальные ассистенты.* Чат-боты помогают студентам получать ответы на вопросы, напоминать о заданиях и сроках, а также предоставлять консультации по учебным материалам в режиме 24/7.
4. *Аналитика и прогнозирование.* Интеллектуальные системы собирают и анализируют данные о поведении студентов, что позволяет преподавателям принимать обоснованные решения по улучшению учебного процесса.
5. *Разработка интерактивных и иммерсивных образовательных сред.* Использование виртуальной и дополненной реальности в сочетании с интеллектуальными технологиями создает более привлекательные и эффективные методы обучения.

К очевидным преимуществам внедрения интеллектуальных технологий в сферу образования можно отнести следующие [2,3]:

- повышение эффективности обучения за счет применения индивидуального подхода к обучению и автоматизации деятельности преподавателя, что позволит ускорить освоение материала обучающимся и снизить нагрузку на преподавателей;
- использование онлайн-форматов и виртуальных ассистентов делают образование более доступным для широкого круга обучающихся, включая тех, кто находится в удаленных регионах;
- постоянный сбор и анализ данных о прогрессе студентов помогает своевременно корректировать учебный процесс;
- применение инновационных методов (интеллектуальных технологий) в обучении стимулирует развитие новых педагогических технологий и методов.

IV. Выводы

В работе рассмотрены перспективные направления применения интеллектуальных технологий в управлении образовательной деятельности, направленные на формирование теоретических и практических основ, необходимых для создания устойчивых, адаптивных и интеллектуальных образовательных систем.

1. Подгаец С. В.Искусственный интеллект в образовании: инновации и возможности / С. В. Подгаец // «Инновационные технологии в современном образовании: стратегия, задачи, внедрение» : Материалы международной научно-практической конференции, 23 февраля 2024 года. II Книга. – Костанай: Костанайский региональный университет имени А.Байтұрсынұлы, 2024. – С. 384-389
2. Kazlova A., Halavaty A. Interoperability as A Critical Component of the Educational Process in Secondary Schools / A. Kazlova, A. Halavaty // Open Semantic Technologies for Intelligent Systems. – Minsk : Bestprint, 2024. – P. 121-126
3. Автоматизация образовательной деятельности в рамках Экосистемы OSTIS / Н. А. Гулякина [и др.] // Часть 7. Экосистема OSTIS. Технология комплексной поддержки жизненного цикла семантически совместимых интеллектуальных компьютерных систем нового поколения / под. ред. В. В. Голенкова. – Минск : Бестпринт, 2023. – С. 741-791

ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПРОГРАММ ДЛЯ ЛОКАЛЬНОГО ЗАПУСКА БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

Дичковский В. А., Ковалёв М. В.
Кафедра интеллектуальных информационных технологий,
Факультет информационных технологий и управления,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {volodya.dichkovski, michail.kovalev7}@gmail.com

В данном докладе представлен сравнительный анализ трёх популярных локальных решений для работы с LLM – vLLM, Ollama и Llama.cpp. Для оценки производительности программ при генерации ответов использовались следующие метрики: время до первого токена, сквозная задержка , задержка между токенами, скорость генерации и количество запросов в секунду.

Введение

Современные большие языковые модели (LLM) получили широкое распространение, решая задачи от генерации текста до программирования и анализа данных.

Однако применение LLM через облачные API ограничено стоимостью, зависимостью от интернета, рисками для конфиденциальности и возможными утечками данных. Актуальным становится локальный запуск языковых моделей.

Производительность программ для локального запуска LLM в значительной степени определяется вычислительными ресурсами, в частности — мощностью видеокарты и объемом её видеопамяти. В профессиональных и промышленных решения, применяются высокопроизводительные и дорогостоящие графические ускорители, такие как NVIDIA A100 или H100, обладающие объемом видеопамяти соответственно 80, 94 гигабайт и выше. Эти устройства обеспечивают максимальную эффективность при работе с крупными моделями, но их крайне высокая стоимость делает их недоступными для большинства частных пользователей, в том числе студентов.

Поэтому остается актуальной задача выбора программы для локального запуска LLM, который обеспечит приемлемую производительность на персональных компьютерах и ноутбуках среднего класса, оснащённых более доступными видеокартами.

I. Выбор и анализ программ

Поскольку развертывание LLM часто происходит в системах реального времени, требующих низкой задержки ответов и высокой пропускной способности, важен анализ и выбор лучших программ[1].

Для анализа были выбраны три локальные программы – vLLM, Ollama и Llama.cpp – они активно применяются в практических задачах и имеют большую популярность.

vLLM демонстрирует высокую производительность за счёт эффективного пакетного объединения запросов (batch processing) и оптимизации использования видеопамяти.

Ollama отличается удобным интерфейсом и ориентирована на интеграцию в пользовательские приложения.

Llama.cpp представляет собой лёгкую реализацию, рассчитанную на работу с моделями при ограниченных вычислительных ресурсах, включая использование CPU, что делает её доступной для запуска на недорогих персональных компьютерах.

Все программы используют KV-кеш для повторного применения промежуточных значений слоёв трансформера, что снижает задержку и ускоряет генерацию. Также они поддерживают совместимый с OpenAI API интерфейс, ставший мировым стандартом.

II. Используемое оборудование

Тестирование проводилось с использованием видеокарты NVIDIA GeForce RTX 3060 (12 ГБ видеопамяти). Графические процессоры NVIDIA используют архитектуру с вычислительными ядрами CUDA. Это обеспечивает эффективное выполнение параллельных вычислений и широкую поддержку в современных фреймворках машинного обучения.

III. Выбор метрик для оценки

Современные LLM представляют собой нейросети, основанные на архитектуре трансформеров. Модель Transformer обрабатывает входные токены для генерации выходных прогнозов, поэтому метрики оценки её производительности связаны с характеристиками генерации токенов[2]. В статье [3] описаны наиболее важные метрики для оценки производительности вывода LLM. Для анализа были выбраны следующие метрики: Для оценки производительности моделей использовались ключевые метрики, описанные в [4 5]. Первая из них – Time to First Token (TTFT) – характеризует задержку между отправкой запроса и появлением первого токена. Этот параметр отражает скорость реакции системы и имеет реша-

ющее значение для приложений, чувствительных к задержка[6].

End-to-End Latency (E2E) отражает полное время отклика от начала запроса до завершения генерации текста. В отличие от TTFT, эта метрика охватывает весь цикл обработки данных и наиболее показательна для серверных нагрузок[7].

Inter-Token Latency (ITL) показывает средний интервал между последовательными токенами. Чем меньше это значение, тем плавнее воспринимается ответ модели в реальном времени.

Tokens Per Second (TPS) и Requests Per Second (RPS) характеризуют пропускную способность системы[8]. TPS показывает, насколько быстро модель генерирует токены, а RPS — сколько запросов она способна обслужить за единицу времени.

IV. Использование библиотеки Locust для нагрузочного тестирования

Для оценки производительности был выбран инструмент с открытым исходным кодом — Locust. Рассматривались также инструменты k6, Gatling, Artillery и NVIDIA GenAI-Perf.

Locust представляет собой библиотеку для Python, предназначенную для нагрузочного тестирования, которая позволяет моделировать поведение множества одновременно работающих пользователей. Библиотека полностью основана на событийной модели и использует лёгкие потоки через gevent для параллельного выполнения запросов.

Преимущества Locust включают веб-интерфейс, упрощающий настройку, возможность отслеживать результаты в реальном времени и экспорт отчетов в форматах html и csv.

V. Выбор модели для тестирования

Для тестирования выбрана модель Saiga LLaMA3 8B (8 миллиардов параметров) с квантованием q4_K. Она подходит для диалогов на русском языке и обеспечивает быстрые ответы при приемлемом качестве генерации. Использовался файл в формате GGUF.

Она была выбрана за лёгкость и способность обеспечивать быстрые ответы при сохранении приемлемого качества генерации. Для запуска использовался файл в формате GGUF.

VI. Результаты тестирования

Для каждой из программ тестирование проводилось в течение 5 минут, моделировалась нагрузка в 50 пользователей.

Перед каждым тестированием использовался warm up, который, согласно [2], повышает результаты производительности (до 4-5%).

Таблица 1 – Результаты тестирования производительности LLM

Метрика	vLLM	Ollama	Llama.cpp
Количество запросов	176	90	130
Среднее TTFT (мс)	30061.1	70700.2	52040.3
Средняе e2e latency (мс)	45444	73035.1	63251.6
Средне ITL (мс)	121.1	18.4	50.0
Среднее TPS (мс)	3.44	2.83	3.10
RPS	0.30	0.17	0.22

VII. Заключение

Проведенное тестирование показало, что vLLM лидирует по пропускной способности и общей задержке, оптимален для высоконагруженных API. Ollama уступает по скорости обработки, но обеспечивает минимальную задержку между токенами, что критично для плавности генерации и интерактивных приложений. Llama.cpp является сбалансированным универсальным решением. Метрики TPS и RPS наиболее определяют пропускную способность. Нужно стремиться сводить e2e latency к минимуму.

1. Best Practices for LLM Latency Benchmarking [Electronic resource] / Newline. – Mode of access: https://www.newline.co/@zaoyang/best-practices-for-llm-latency-benchmarking--257f132d. – Date of access: 16.10.2025.
2. Zhen, R. Taming the Titans: A Survey of Efficient LLM Inference Serving / R. Zhen [et al.] // arXiv preprint. – 2025. – arXiv:2504.19720v1.
3. Sagi, S. Optimizing LLM Inference: Metrics that Matter for Real Time Applications / S. Sagi // Journal of Artificial Intelligence & Cloud Computing. – 2025. – Vol. 4, № 4. – P. 1–10. – DOI: 10.47363/jaicc/2025(4)446.
4. Rajaneesh, N., Zollo, T., Zemel, R. Test-Time Warmup for Multimodal Large Language Models / N. Rajaneesh, T. Zollo, R. Zemel // arXiv preprint. – 2020. – arXiv:2509.10641v1. – P. 1–13.
5. LLM Benchmarking: Fundamental Concepts [Electronic resource] / NVIDIA. – Mode of access: https://developer.nvidia.com/blog/llm-benchmarking-fundamental-concepts/. – Date of access: 16.10.2025.
6. Sagi, S. Optimizing LLM Inference: Metrics that Matter for Real Time Applications / S. Sagi // Journal of Artificial Intelligence & Cloud Computing. – 2025. – Vol. 4, № 1. – P. 1–4.
7. NVIDIA NIM: Large Language Models Benchmarking [Electronic resource] / NVIDIA. – Mode of access: https://docs.nvidia.com/nim/large-language-models/1.2.0/benchmarking.html. – Date of access: 16.10.2025.
8. Тестирование производительности видеокарт на примере больших языковых моделей с использованием Llama.cpp [Электронный ресурс] / Habr. – Режим доступа: https://habr.com/ru/articles/916836/. – Дата доступа: 16.10.2025.

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОГО ПЛАГИНА С ЛОКАЛЬНОЙ МОДЕЛЬЮ ГЕНЕРАЦИИ КОДА ДЛЯ ЯЗЫКА SC-КОД

Дубинина Е. О., Захаров В. В.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: eugenia.dubinina@gmail.com, zakharov@bsuir.by

Большие языковые модели демонстрируют высокую эффективность в задачах генерации кода, однако их применение ограничено популярными языками программирования. Специализированные языки, такие как SC-код платформы ОСТИС, остаются без поддержки. Более того, большинство существующих инструментов полагаются на облачную инфраструктуру, что создает проблемы с конфиденциальностью данных и требует постоянного подключения к сети. В данной работе представлен подход к разработке интеллектуального плагина для среды разработки, который интегрирует локально развернутую и оптимизированную модель для генерации SC-кода. Для решения проблемы нехватки данных был собран уникальный датасет. Модель семейства CodeGen была дообучена и оптимизирована с использованием методов квантования и дистилляции, что позволило достичь высокой производительности на потребительском оборудовании. Тестирование прототипа показало точность генерации 82,5% при средней задержке 165 мс.

Введение

Интеграция больших языковых моделей (БЯМ) в интегрированные среды разработки кардинально меняет процессы написания кода. Инструменты, такие как GitHub Copilot и JetBrains AI Assistant, ускоряют разработку, предлагая автодополнение и генерацию кода по запросу на естественном языке. По данным исследований GitHub, разработчики, использующие Copilot, выполняют задачи до 55% быстрее, а в крупных технологических компаниях ИИ-ассистенты уже генерируют от 40% до 60% нового кода.

Однако у этого подхода есть две существенные проблемы. Во-первых, существующие модели обучены преимущественно на популярных языках. Специализированные и доменно-специфические языки, к которым относится декларативный SC-код платформы ОСТИС, практически не представлены в обучающих выборках. Это приводит к синтаксически и семантически некорректным результатам генерации.

Во-вторых, большинство сервисов работают по облачной модели. Это вызывает опасения относительно конфиденциальности исходного кода и делает невозможной работу в офлайн-режиме.

Целью данной работы является разработка прототипа интеллектуального плагина для среды разработки, который решает обе проблемы: он интегрирует локально развернутую БЯМ, специально дообученную для генерации кода на языке SC-код.

I. Подход к разработке

Для достижения поставленной цели необходимо было решить три ключевые задачи:

1. Инженерия данных: Формирование и аугментация специализированного корпуса данных (датасета). Этот этап был необходим для последующего дообучения модели на синтаксических и семантических конструкциях SC-кода, который не представлен в публичных выборках.
2. Адаптация и оптимизация модели: Выбор подходящей архитектуры БЯМ, ее дообучение на собранном датасете и применение методов компрессии (таких как квантование и дистилляция) для обеспечения высокой производительности и низкого потребления ресурсов на локальном устройстве.
3. Системная интеграция: Проектирование клиент-серверной архитектуры, включающей сам плагин для среды разработки (клиент) и локальный бэкенд-сервис, который инкапсулирует оптимизированную модель и обслуживает запросы на генерацию кода

II. Подготовка данных

Ключевой проблемой при работе со специализированными языками, такими как SC-код, является острый дефицит или полное отсутствие публичных размеченных данных для обучения. Для решения этой задачи был применен гибридный подход к формированию корпуса, включающий два источника.

Во-первых, был выполнен систематический анализ и парсинг существующих публичных репозиториев платформы ОСТИС для сбора всего доступного реального кода. Во-вторых, для обеспечения полноты покрытия грамматических конструкций и типичных структур была применена техника синтетической генерации данных.

Для обеспечения возможности обучения модели генерации по текстовому запросу, каждый фрагмент кода был аннотирован (сопоставлен) с описанием на естественном (русском) языке. Ре-

зультирующий объем датасета составил около 250 000 строк кода и 70 000 текстовых описаний.

III. Выбор и оптимизация модели

Для обеспечения локальной работы была выбрана модель из семейства CodeGen, имеющая открытый исходный код и доступная в различных размерах. Модель была дообучена на собранном датасете для адаптации к специфическому синтаксису SC-кода.

1. Квантование: Веса модели были квантованы до 8-битного формата. Это позволило сократить размер модели с 1.4 ГБ до 350 МБ, что критически важно для загрузки в ОЗУ ноутбука, и ускорило вычисления.
2. Дистилляция знаний: Компактная "модель-ученик"была обучена повторять поведение "модели-учителя"большего размера, что позволило сохранить высокое качество генерации при меньшем количестве параметров.

IV. Количественные результаты

Для оценки использовался тестовый набор из 100 запросов на естественном языке, описывающих типовые задачи на SC-коде.

Точность генерации: Общая точность (доля семантически корректных фрагментов) составила 82%. Модель показала высокую точность (90%) на простых запросах (создание узлов, связей) и более низкую (65%) на сложных структурах, что указывает на необходимость расширения датасета.

Таблица 1 – Результаты оценки точности генерации

Категория запросов	Количество запросов	Успешно	Точность, %
Простые	56	55	91
Сложные	40	27	67
Итого	100	82	82

Скорость генерации: Благодаря оптимизации и архитектуре Apple Silicon, среднее время генерации составило всего 165 мс (с разбросом от 110 мс до 285 мс). Это обеспечивает практически мгновенный отклик в среде разработки.

V. Качественные результаты

Было проведено пилотное исследование с участием разработчиков, имеющих опыт работы с ОСТИС.

1. Удобство: 100% пользователей отметили интуитивность интерфейса и удобство вызова генерации.
2. Полезность: Участники подтвердили значительное ускорение рутинных задач, таких как создание стандартных rrel или nrel структур.
3. Качество: В 85% случаев код был признан релевантным. В остальных случаях требовалась незначительная ручная доработка, в основном из-за неоднозначности запросов.

VI. Заключение

В ходе работы был успешно разработан и протестирован прототип интеллектуального плагина, интегрирующего локальную БЯМ для генерации кода на специализированном языке SC-код.

Доказана жизнеспособность подхода, основанного на сборе кастомного датасета и применении методов оптимизации (квантование, дистилляция). Это позволило достичь высокой производительности (82% точность, 165 мс задержка) на стандартном потребительском оборудовании.

Данная работа демонстрирует практический путь создания автономных, конфиденциальных и эффективных ИИ-ассистентов для доменно-специфических языков программирования, что ранее было прерогативой облачных сервисов для популярных языков.

Перспективы развития включают расширение обучающего датасета для повышения точности на сложных запросах и реализацию полноценного функционала, включая контекстное автодополнение и исправление ошибок.

1. Chen, M. Evaluating Large Language Models Trained on Code / M. Chen [et al.] // arXiv preprint arXiv:2107.03374. – 2021.
2. Nijkamp, E. CodeGen: An Open Large Language Model for Code with Multi-Turn Program Synthesis / E. Nijkamp [et al.] // arXiv preprint arXiv:2203.13474. – 2022.
3. Hinton, G. Distilling the Knowledge in a Neural Network / G. Hinton [et al.] // arXiv preprint arXiv:1503.02531. – 2015.
4. Dettmers, T. LLM.int8(): 8-bit Matrix Multiplication for Transformers at Scale / T. Dettmers [et al.] // arXiv preprint arXiv:2208.07339. – 2022.
5. Golenkov, V. V. Ontology-based design of intelligent systems // Open semantic technologies for intelligent systems. – Minsk: BSUIR, 2017. – Iss. 1. – P. 37–56.
6. Голенков, В. В., Технология комплексной поддержки жизненного цикла семантически совместимых интеллектуальных компьютерных систем нового поколения. – Минск: Бестпринт, 2023.

ОПТИМИЗАЦИЯ RAG-АРХИТЕКТУР ДЛЯ ПРЕДМЕТНО-ОРИЕНТИРОВАННЫХ СПРАВОЧНЫХ СИСТЕМ ВУЗОВ

Завалей В. А., Орлов Д. В., Скиба И. Г.
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {v.zavalej, d.orlov, i.skiba}@bsuir.by

В статье рассматриваются подходы к повышению эффективности RAG-архитектур в справочных системах вузов. Проанализированы проблемы наивных реализаций, связанные с обработкой гетерогенных академических данных. Предложен комплекс методов оптимизации, включающий семантическое чанкирование документов, гибридный поиск и динамическое построение промпта. Разработана система критериев оценки, учитывающая специфику образовательной среды. Результаты работы представляют методическую основу для создания интеллектуальных ассистентов в учебных заведениях.

Введение

Современные высшие учебные заведения характеризуются сложной организационной структурой и значительным объемом внутренней документации, включающей учебные планы, академические регламенты, расписания занятий и методические указания. Обеспечение оперативного и точного доступа к этой информации для студентов, преподавателей и сотрудников представляет собой важную практическую задачу. Универсальные большие языковые модели, такие как GPT, не имеют доступа к актуальной и часто конфиденциальной информации конкретного учреждения, что может приводить к генерации недостоверных или устаревших сведений [1].

Архитектура Retrieval-Augmented Generation (RAG) представляет собой перспективный подход для решения этой проблемы, позволяя сочетать генеративные способности языковых моделей с точным доступом к релевантным внешним источникам данных. Однако стандартные реализации RAG-систем демонстрируют недостаточную эффективность при работе со специализированным контентом высших учебных заведений. Основные проблемы включают неудачное извлечение контекста вследствие неоптимального разделения документов на фрагменты, неспособность адекватно обрабатывать составные запросы и игнорирование специфической терминологии предметной области.

Целью данной работы является теоретическое обоснование и разработка методов оптимизации RAG-архитектуры, адаптированных для нужд справочных систем высших учебных заведений. В качестве основных задач определены: анализ ключевых проблемных мест в конвейере RAG, предложение конкретных методов оптимизации на этапах поиска и генерации, а также разработка системы критериев для сравнительного анализа эффективности предлагаемых решений.

I. Анализ проблем RAG-архитектур

Стандартные реализации RAG-архитектур демонстрируют ряд системных проблем при работе с корпоративными данными вуза. Академическая среда характеризуется высокой структурной и семантической гетерогенностью документов, что существенно усугубляет традиционные ограничения RAG-конвейера.

Ключевой проблемой на этапе извлечения информации является неоптимальное чанкирование документов. Статические методы разбиения текста на фрагменты фиксированной длины часто разрушают логические связи внутри документов. Например, при обработке типичного учебного плана информация о конкретной дисциплине, её кредитной стоимости, семестре изучения и форме контроля знаний может оказаться распределенной по разным чанкам. Это делает невозможным формирование целостного ответа на запрос студента о содержании конкретного курса [2].

Другой существенной проблемой является ограниченная эффективность чистого семантического поиска. Векторные модели, лежащие в основе такого поиска, могут недостаточно точно ранжировать документы при работе с узкопредметными терминами, общепринятыми аббревиатурами и составными запросами, характерными для академической среды. Типичным примером может служить запрос «Как пересдать ЭВМ осенью?», где «ЭВМ» представляет собой аббревиатуру дисциплины «Электронные вычислительные машины» [3].

На этапе генерации ответов возникает проблема игнорирования языковой моделью части предоставленного контекста. Модель может чрезмерно полагаться на собственные параметрические знания, полученные в ходе предварительного обучения, что особенно критично в условиях быстро меняющейся информации образовательного учреждения.

II. Методы оптимизации RAG-конвейера

Для решения выявленных проблемных аспектов предлагается комплекс методов оптимизации, направленных на повышение эффективности ключевых этапов работы RAG-архитектуры в условиях высшего учебного заведения.

Проблема неоптимального чанкирования документов может быть успешно решена за счет перехода от статического разбиения текста к семантическому чанкированию. Данный подход предполагает интеллектуальное разделение документов на логически целостные фрагменты на основе анализа структуры текста и его смысловых границ. Для документов с жесткой структурой, таких как учебные планы или официальные приказы, эффективно использование специально разработанных шаблонов разметки. В случае работы с неструктурированными текстами применяются алгоритмы анализа когерентности, учитывающие тематические переходы.

Для повышения точности извлекаемых документов целесообразно внедрение гибридного поиска, комбинирующего семантический и ключевой методы. Семантический поиск на основе векторных представлений обеспечивает понимание смысловых оттенков запроса, а ключевой поиск гарантирует точное соответствие специфическим терминам и аббревиатурам. Результирующий список формируется на основе взвешенной суммы оценок релевантности [4].

На этапе генерации для уменьшения вероятности игнорирования контекста предлагается метод динамического построения промпта. В отличие от простого объединения чанков, промпт структурируется для направления внимания модели на ключевые аспекты запроса. Дополнительно используется контекстуальная компрессия для снижения семантического шума.

III. Критерии оценки эффективности

Для объективного сравнительного анализа эффективности базовой и оптимизированной RAG-архитектур предлагается использовать систему метрик, ориентированную на релевантность, точность и семантическую целостность ответов.

Оценка релевантности извлеченного контекста проводится с использованием метрик Precision@K и Recall@K. Для автоматизированной оценки может использоваться метрика NDCG (Normalized Discounted Cumulative Gain), учитывающая порядок следования релевантных документов в ранжированном списке, что особенно важно для обеспечения качества финального ответа системы [5].

Качество конечного ответа системы оценивается по точности представленных фактов и семантической согласованности. Точность может измеряться путем экспертного сравнения сгенерированных утверждений с эталонными данными из доверенных источников вуза. Семантическая согласованность оценивает, насколько логически связан и целостен представленный ответ.

Дополнительным практическим критерием является коэффициент успешного завершения диалога, который отражает процент взаимодействий с исчерпывающим ответом без необходимости переформулировать запрос. Эта метрика отражает удовлетворенность пользователя и общую эффективность системы в реальных условиях.

Заключение

В работе проведен анализ системных проблем, характерных для наивных реализаций RAG-архитектур при их применении в предметно-ориентированных справочных системах высших учебных заведений. Выявлены ключевые ограничения на этапах извлечения и генерации, связанные с неоптимальным чанкированием документов, недостаточной релевантностью поиска и некорректной обработкой контекста.

Для решения указанных проблем предложен комплекс методов оптимизации, включающий семантическое чанкирование с учетом типов учебных документов, гибридный поиск, комбинирующий семантический и ключевой подходы, и динамическое построение промпта для улучшения обработки контекста языковой моделью.

Разработана система критериев для оценки эффективности RAG-архитектур, включающая метрики релевантности извлеченного контекста, точности и семантической согласованности генерируемых ответов. Перспективными направлениями являются разработка адаптивных алгоритмов чанкирования и создание специализированных моделей для академической сферы.

Предложенные решения создают методическую основу для построения более надежных и эффективных интеллектуальных справочных систем, способных качественно улучшить информационную поддержку студентов и сотрудников вузов.

IV. Список литературы

1. Lewis, P. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks / P. Lewis et al. // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 9459–9474.
2. Gao, L. et al. RAG vs Fine-tuning: A Comprehensive Comparison / L. Gao et al. // arXiv preprint arXiv:2305.16983. – 2023.
3. Hofstätter, S. et al. Hybrid Retrieval Models / S. Hofstätter et al. // ACM Transactions on Information Systems. – 2021. – Vol. 39, № 4. – P. 1–35.
4. Wang, L. et al. Precise Zero-shot Dense Retrieval without Relevance Labels / L. Wang et al. // Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics. – 2022. – P. 176–186.
5. Savchenko, A. V. Semantic Methods of Text Processing / A. V. Savchenko. – Minsk: BSU, 2020. – 215 p.

МОДЕЛЬ УПРАВЛЕНИЯ АГЕНТАМИ В ОБЩЕЙ СЕМАНТИЧЕСКОЙ ПАМЯТИ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Зотов Н. В.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: n.zotov@bsuir.by

Предложена математическая модель управления агентами в общей семантической памяти интеллектуальных систем.

Введение

Платформа OSTIS [1, 2] представляет собой комплексное решение для разработки ostis-систем. Ключевыми компонентами таких систем являются интеллектуальные агенты — автономные программные компоненты, способные реагировать на события в sc-памяти и выполнять действия в соответствии со спецификациями.

Цель настоящей работы состоит в описании математической модели управления агентами в sc-памяти ostis-систем.

I. Модель управления агентами в sc-памяти

Модель управления агентами в sc-памяти определяется как:

$$\mathcal{AM} = \langle \mathcal{A},\mathcal{S},\mathcal{E},\mathcal{C},\mathcal{T},\mathcal{R},\mathcal{I} \rangle \quad (1)$$

где $\mathcal{A}$ – множество агентов и их спецификаций; $\mathcal{S}$ – пространство состояний агентов; $\mathcal{E}$ – множество событий в sc-памяти; $\mathcal{C}$ – модель синхронизации агентов; $\mathcal{T}$ – временная модель агентов; $\mathcal{T}$ – модель исполнения агентов; $\mathcal{I}$ – программный интерфейс управления агентами.

Агенты и их спецификации образуют систему непересекающихся множеств:

$$\mathcal{A} = \bigcup_{i=1}^{n} \mathcal{A}_i, \text{ где } \mathcal{A}_i \cap \mathcal{A}_j = \emptyset \text{ для } i \neq j \quad (2)$$

Каждый агент $a \in \mathcal{A}$ определяется формально через его спецификацию:

$$\text{spec}(a) = \langle \tau,\pi,\alpha,\kappa,\rho,\sigma,\theta \rangle \quad (3)$$

где $\tau{:}\mathcal{E} \to \{0,1\}$ – предикат инициирования события в sc-памяти, определяющий первичное условие инициирования агента; $\pi{:}\mathcal{S} \to \{0,1\}$ – предикат активации агента, определяющий условие активации агента при успешном выполнении τ; $\alpha{:}\mathcal{S} \times \mathcal{E} \to \mathcal{S}$ – функция перехода между состояниями агентов, определяющая поведение агента; $\kappa \subseteq \mathcal{E}$ – множество ключевых sc-элементов (элементов в sc-памяти), необходимых для функционирования агента; $\rho{:}\mathcal{S} \to \{0,1\}$ – предикат результирующего условия агента, определяющий условие успешного выполнения агента; $\sigma \in \mathcal{P}(\mathcal{M})$ – множество объектов синхронизации.

Пространство состояний агентов задаётся:

$$\mathcal{S} = \mathcal{S}_{\text{idle}} \cup \mathcal{S}_{\text{ready}} \cup \mathcal{S}_{\text{active}} \cup \mathcal{S}_{\text{wait}} \cup \mathcal{S}_{\text{done}} \quad (4)$$

$$\forall i,j \in \{\text{idle,ready,active,wait,done}\}, i \neq j: \quad \mathcal{S}_i \cap \mathcal{S}_j = \emptyset \quad (5)$$

Функции перехода определяются как:

$$_{\text{init:}}\mathcal{S}_{\text{idle}} \times {}_{\text{spec}}(\mathcal{A}) \rightharpoonup \mathcal{S}_{\text{ready}},\ _{\text{activate:}}\mathcal{S}_{\text{ready}} \times \mathcal{E} \rightharpoonup \mathcal{S}_{\text{active}},$$
$$_{\text{suspend:}}\mathcal{S}_{\text{active}} \rightharpoonup \mathcal{S}_{\text{wait}},\ _{\text{resume:}}\mathcal{S}_{\text{wait}} \times \mathcal{E} \rightharpoonup \mathcal{S}_{\text{active}},$$
$$_{\text{complete:}}\mathcal{S}_{\text{active}} \rightharpoonup \mathcal{S}_{\text{done}},\ _{\text{destroy:}}\mathcal{S}_{\text{done}} \rightharpoonup \mathcal{S}_{\text{idle}}.$$

Модель синхронизации агентов задаётся:

$$\mathcal{C} = \langle \mathcal{L},\mathcal{Q},\mathcal{W},\Delta,\Gamma,\Phi \rangle \quad (6)$$

где $\mathcal{L}$ – множество блокировок для разделяемых ресурсов; $\mathcal{Q}{:}\mathcal{A} \to \mathbb{N}$ – функция приоритета агентов для планирования выполнения; $\mathcal{W}{:}\mathcal{A} \times \mathcal{L} \to \{0,1\}$ – предикат ожидания ресурса агентом; $\Delta{:}\mathcal{L} \times \mathcal{A} \times \mathbb{T} \to \{0,1\}$ – предикат владения блокировкой агентом в момент времени; $\Gamma{:}\mathcal{A} \times \mathcal{A} \to \{0,1\}$ – отношение зависимости между агентами (граф ожидания); $\Phi{:}\mathcal{A} \to \mathbb{R}^+$ – функция времени ожидания агента.

Модель удовлетворяет инвариантам:

Инвариант 1 (Безопасность)**.**

$$\forall l \in \mathcal{L}, \forall t \in \mathbb{T}: |\{a \in \mathcal{A} \mid \Delta(l,a,t) = 1\}| \leq 1 \quad (7)$$

Инвариант 2 (Живость)**.**

$$\forall a \in \mathcal{A}, \forall t \in \mathbb{T}: \nu(a,t) = \text{active} \implies \exists t' > t: \nu(a,t') = \text{done} \quad (8)$$

Инвариант 3 (Отсутствие дедлоков)**.**

$$\nexists (a_1,a_2,\ldots,a_k) \in \mathcal{A}^k: \Gamma(a_1,a_2) \wedge \Gamma(a_2,a_3) \wedge \wedge \cdots \wedge \Gamma(a_k,a_1) = 1 \quad (9)$$

Инвариант 4 (Справедливость)**.**

$$\forall a \in \mathcal{A}, \forall l \in \mathcal{L}, \forall t \in \mathbb{T}: \mathcal{W}(a,l) = 1 \implies \exists t' > t: \Delta(l,a,t') = 1 \quad (10)$$

Временная модель выполнения задаётся:

$$\mathcal{T} = \langle \mathbb{T}, <, \nu, \eta, \mu \rangle \tag{11}$$

где $\mathbb{T}$ – множество временных точек (дискретное или непрерывное); $<\subseteq \mathbb{T} \times \mathbb{T}$ – строгое линейное упорядочение на $\mathbb{T}$; $\nu{:}\mathcal{A} \times \mathbb{T} \to \mathcal{S}$ – функция состояния агента во времени; $\eta{:}\mathcal{E} \times \mathbb{T} \to \{0,1\}$ – предикат возникновения события в момент времени; $\mu{:}\mathcal{A} \times \mathbb{T} \to \mathbb{R}_0^+$ – функция времени выполнения агента.

Модель исполнения агентов определяется как система переходов:

$$\mathcal{R} = \langle \mathcal{S}, \Sigma, \to, s_0, \mathcal{F} \rangle \tag{12}$$

где $\Sigma = \mathcal{E} \cup \{\tau\}$ – алфавит действий (τ обозначает внутреннее действие агента); $\to \subseteq \mathcal{S} \times \Sigma \times \mathcal{S}$ – отношение переходов, определяющее возможные переходы между состояниями; $s_0 \in \mathcal{S}_{\text{idle}}$ – начальное состояние агента; $\mathcal{F} \subseteq \mathcal{S}_{\text{done}}$ – множество допустимых финальных состояний.

Программный интерфейс управления агентами определяется множеством операций:

$$\mathcal{I} = \langle \mathcal{O}_{\text{lifecycle}}, \mathcal{O}_{\text{control}}, \mathcal{O}_{\text{subscription}} \rangle \tag{13}$$

где $\mathcal{O}_{\text{lifecycle}}$ – операции управления жизненным циклом агентов; $\mathcal{O}_{\text{control}}$ – операции управления выполнением агентов; $\mathcal{O}_{\text{subscription}}$ – операции управления подписками на события.

Множество операций жизненного цикла:

$$\begin{aligned}\mathcal{O}_{\text{lifecycle}} = \{&\text{createBySpec,deleteBySpec,}\\ &\text{initialize,start,pause,}\\ &\text{resume,stop,terminate}\}\end{aligned} \tag{14}$$

где:

$$\begin{aligned}&_{\text{createBySpec:Spec}} \to \mathcal{A} \cup \{\bot\},\ _{\text{deleteBySpec:Spec}} \to \{\top, \bot\},\\ &_{\text{initialize:}}\mathcal{A} \times {}_{\text{Spec}} \to \{\top, \bot\},\ _{\text{start:}}\mathcal{A} \to \{\top, \bot\},\\ &\quad _{\text{pause:}}\mathcal{A} \to \{\top, \bot\},\ _{\text{resume:}}\mathcal{A} \to \{\top, \bot\},\\ &\quad _{\text{stop:}}\mathcal{A} \to \{\top, \bot\},\ _{\text{terminate:}}\mathcal{A} \to \{\top, \bot\}\end{aligned}$$

где $\bot$ обозначает неудачное выполнение операции, $\top$ – успешное.

Множество операций управления:

$$\begin{aligned}\mathcal{O}_{\text{control}} = \{&\text{acquireLock,}\\ &\text{releaseLock,waitForEvent}\}\end{aligned} \tag{15}$$

где:

$$\begin{aligned}\text{acquireLock}{:}\mathcal{A} \times \mathcal{L} &\to \{\top, \bot\},\\ \text{releaseLock}{:}\mathcal{A} \times \mathcal{L} &\to \{\top, \bot\},\\ \text{waitForEvent}{:}\mathcal{A} \times \mathcal{E} \times \mathbb{R}_0^+ &\to \{\top, \bot\}.\end{aligned}$$

Система подписок определяется как отношение между агентами и событиями:

$$\mathcal{SUB} \subseteq \mathcal{A} \times \mathcal{E} \times \mathcal{FP} \times \text{CF} \tag{16}$$

где FP – множество предикатов фильтрации событий; CF – множество функций обратного вызова для обработки событий.

Операции управления подписками:

$$\mathcal{O}_{\text{subscription}} = \{\text{subscribe,unsubscribe}\} \tag{17}$$

где:

$$\begin{aligned}\text{subscribe}{:}\mathcal{A} \times \mathcal{E} \times \text{FP} \times \text{CF} &\to \text{SID} \cup \{\bot\},\\ \text{unsubscribe}{:}\mathcal{A} \times \text{SID} &\to \{\top, \bot\}.\end{aligned}$$

Инвариант 5 (Уникальность подписок)**.**

$$\forall s_1, s_2 \in \text{SID}{:}s_1 \neq s_2 \implies \text{sub}(s_1) \neq \text{sub}(s_2) \tag{18}$$

Инвариант 6 (Консистентность подписок)**.**

$$\begin{aligned}\forall a \in \mathcal{A}, s \in \text{SID}{:}&\text{unsubscribe}(a,s) = \top \implies\\ &\exists e,f,c{:}(a,e,f,c) \in \mathcal{SUB}\end{aligned} \tag{19}$$

Инвариант 7 (Живость подписок)**.**

$$\begin{aligned}\forall (a,e,f,c) \in \mathcal{SUB}, \forall t \in \mathbb{T}{:}&\eta(e,t) = 1 \wedge\\ &\wedge \text{active_sub}(a,e,f,c,t) = 1 \implies\\ &\exists t' \geq t{:}|t' - t| < \Delta_{\max} \wedge c(e,t')\end{aligned} \tag{20}$$

где $\Delta_{\max}$ – максимальное время доставки уведомления.

Заключение

Представленная модель обеспечивает формальную основу для реализации модуля управления агентами в Платформе OSTIS с гарантиями корректности через систему инвариантов, поддержку событийно-ориентированной архитектуры и полный набор операций управления жизненным циклом агентов и подписками на события.

1. Zotov, N. OSTIS Platform – a Framework for Developing Intelligent Agents Based on Semantic Networks / N. Zotov // Open Semantic Technologies for Intelligent Systems (OSTIS) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : В. В. Голенков [и др.]. – Минск, 2025. – Вып. 9. – С. 113-134.
2. Зотов, Н. В. Модель управления процессами в общей семантической памяти интеллектуальных систем / Н. В. Зотов // Информационные технологии и системы 2023 (ИТС 2023) : материалы Международной научной конференции, Минск, 22 ноября 2023 / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск : БГУИР, 2023. – С. 53-54.

СРЕДСТВА РАЗРАБОТКИ АГЕНТОВ В ПРОГРАММНОЙ ПЛАТФОРМЕ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Зотов Н. В.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: n.zotov@bsuir.by

В работе рассмотрены программные средства разработки агентов в ostis-системах.

Введение

Развитие современных интеллектуальных систем характеризуется переходом от монолитных архитектур к модульным подходам, основанным на взаимодействии агентов. Ключевым требованием к таким системам является обеспечение семантической совместимости компонентов и возможность их динамической интеграции для решения комплексных задач.

Платформа OSTIS [1, 2] предлагает решение указанных проблем через использование единого языка представления знаний – SC-кода – и sc-памяти для всех компонентов системы. Такой подход обеспечивает унифицированную основу для создания семантически совместимых агентов, способных к эффективному взаимодействию и совместному решению задач.

I. Средства разработки агентов в OSTIS-системах

Платформа OSTIS предоставляет инструменты для создания, управления и интеграции агентов.

Агенты реагируют на события в sc-памяти. Агент активируется, когда происходит событие в sc-памяти, на которое он подписан. Основное условие инициирования определяет событие в sc-памяти, которое пробуждает агента. После пробуждения агент проверяет полное условие своего инициирования. В случае успеха он инициирует и выполняет действие, используя свою программу. После агент проверяет свой результат.

В рамках программного интерфейса для создания агентов и управления ими введены классы и методы для работы с агентами:

1. Два базовых класса для всех типов агентов:
   - класс `ScAgent` для реализации классов агентов, которые реагируют на любые элементарные события в sc-памяти;
   - класс `ScActionInitiatedAgent` для реализации агентов, которые реагируют на события инициированных действий в sc-памяти.
2. Класс `ScAgentContext` для работы с событиями `ScEvent`, подписками `ScEventSubscription` и ожидателями `ScWait` и `ScEventWaiter`.
3. Класс `ScAction` для обработки действий в sc-памяти.
4. Класс `ScAgentBuilder` для управления динамическими спецификациями агентов.
5. Класс `ScKeynodes` для указаний ключевых sc-элементов, используемых в агентах.
6. Класс `ScModule` для регистрации агентов в sc-памяти.

Пример использования класса `ScAgent` для создания агента представлен в листинге 1.

```
1  // Файл my_agent.hpp
2  #pragma once
3
4  #include <sc-memory/sc_agent.hpp>
5
6  // Класс агента должен наследовать класс
7  ScAgent и указывать шаблонный аргумент
8  как класс sc-события.
9  Здесь ScEventAfterGenerateIncomingArc
       <ScType::ConstPermPosArc> -
10 тип события, на которое реагирует агент.
11 class MyAgent : public ScAgent<
12   ScEventAfterGenerateIncomingArc<
         ScType::ConstPermPosArc>>
13 {
14 public:
15   // Здесь указывается класс действий,
16   которые выполняет данный агент.
17   // Здесь ‘GetActionClass‘
18   переопределяет ‘GetActionClass‘
19   в классе ‘ScAgent‘.
20   Это переопределение обязательно.
21   ScAddr GetActionClass() const
         override;
22   // Здесь указывается программа данного
23   агента. Это переопределение обязательно.
24   ScResult DoProgram(
25     ScEventAfterGenerateIncomingArc<
26       ScType::ConstPermPosArc> const
             & event,
27     ScAction & action) override;
28     // ...
29     // Другие пользовательские методы.
30 };
```

Листинг 1 – Определение агента, наследующего класс `ScAgent`

Для подписки агентов на события необходимо реализовать модуль (листинг 2) и вызвать метод `Agent` для подписки агентов (листинг 3).

```
1  // Файл my_module.hpp
2  #pragma once
3
4  #include <sc-memory/sc_module.hpp>
5
6  class MyModule : public ScModule
```

```
{
};
```

Листинг 2 – Определение класса модуля, наследующего класс ScModule

```
// Файл my_module.cpp:
#include "my-module/my_module.hpp"

#include "my-module/keynodes/
    my_keynodes.hpp"
#include "my-module/agent/my_agent.
    hpp"

SC_MODULE_REGISTER(MyModule)
  // Инициализирует статический объект
  класса `MyModule`, который может быть
  использован для вызова методов подписки
  агентов на sc-события.
  ->Agent<MyAgent>();
  // Этот метод подписывает агента и
  возвращает объект MyModule.
```

Листинг 3 – Подписка агента через класс модуля

Каждый агент имеет спецификацию. Спецификация агента представляет собой формализованный подход к управлению агентами.

Спецификация агента включает:

- его первичное условие инициирования,
- класс действий, которые он может выполнять,
- полное условие инициирования,
- условие результата,
- ключевые sc-элементы, используемые во время выполнения действия,
- и другие детали.

Спецификация агента может быть представлена в базе знаний с использованием SC-кода или программно с использованием программного интерфейса sc-памяти на C++.

Программный интерфейс для создания агентов и управления ими предоставляет два метода для реализации агентов: (1) когда спецификация агента представлена в базе знаний; (2) когда спецификация агента представлена непосредственно в коде.

Спецификации агентов могут быть статическими, динамическими или полудинамическими.

1. *Статическая спецификация агента* предоставляется извне в классе агента (путём переопределения публичных методов-геттеров классов `ScAgent` или `ScActionInitiatedAgent`). Она не хранится в базе знаний.
2. *Динамическая спецификация агента* хранится в базе знаний, либо определяется непосредственно на SCs/SCg-коде, либо загружается через API `ScAgentBuilder`.
3. *Полудинамическая спецификация агента* предоставляется в базе знаний или изначально в коде и дополняется извне.

Пример реализации агента с полудинамической спецификацией показан в листинге 4.

```
// Файл my_agent.hpp
#pragma once

#include <sc-memory/sc_agent.hpp>

class MyAgent : public ScAgent<
    ScEventAfterGenerateIncomingArc<
    ScType::ConstPermPosArc>>
{
public:
  // Статическая часть: класс действий
  агента задаётся в коде.
  ScAddr GetActionClass() const
      override
  {
    return MyKeynodes::my_action;
  }

  // Статическая часть: программа агента
  задаётся в коде.
  ScResult DoProgram(ScAction &
      action) override
  {
    // Реализация логики агента.
    return action.FinishSuccessfully
        ();
  }

  // Динамическая часть: условие
  инициирования агента извлекается из
  базы знаний.
  ScAddr GetInitiationCondition()
      const override
  {
    ScAddr result =
        SearchInitiationCondition(
        GetAbstractAgent());
    return result.IsValid() ? result
         : ScAgent::
        GetInitiationCondition();
  }

  // Другие методы могут быть реализованы
  аналогично.
};
```

Листинг 4 – Пример полудинамической спецификации агента

Заключение

Программный интерфейс для создания агентов и управления ими в общей семантической памяти, реализованный в рамках технологии OSTIS, является ключевым элементом программной платформы для разработки ostis-систем. Этот интерфейс обеспечивает создание, спецификацию и взаимодействие агентов, работающих в sc-памяти, что позволяет эффективно решать задачи обработки знаний и построения семантически совместимых систем.

1. Zotov, N. OSTIS Platform – a Framework for Developing Intelligent Agents Based on Semantic Networks / N. Zotov // Open Semantic Technologies for Intelligent Systems (OSTIS) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : В. В. Голенков [и др.]. – Минск, 2025. – Вып. 9. – С. 113-134.
2. Зотов, Н. В. Модель управления процессами в общей семантической памяти интеллектуальных систем / Н. В. Зотов // Информационные технологии и системы 2023 (ИТС 2023) : материалы Международной научной конференции, Минск, 22 ноября 2023 / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск : БГУИР, 2023. – С. 53-54.

АНАЛИЗ МЕТОДОВ РАСПОЗНАВАНИЯ ПРОФЕССИЙ В РУССКОЯЗЫЧНЫХ ТЕКСТАХ

Иванова М. Д.
Факультет прикладной математики и информатики, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: ivanova.mary76@gmail.com

В работе рассматривается задача распознавания именованных сущностей категории «Профессия» в русскоязычных текстах. Целью исследования является определение наиболее эффективного подхода к извлечению профессий из неструктурированных данных. Проведён сравнительный анализ трёх различных методов: основанных на знаниях (словарь), машинном обучении (CRF) и глубоком обучении (RuBERT). Эксперименты проведены на специально разработанном размеченном корпусе данных. Результаты могут быть использованы для интеллектуального анализа вакансий и данных о рынке труда.

Введение

Распознавание именованных сущностей (Named Entity Recognition, NER) категории «Профессия» представляет собой важное направление в области автоматического анализа текстов, поскольку позволяет извлекать из неструктурированных данных сведения о профессиональной принадлежности и трудовой деятельности людей. Такая информация имеет практическую ценность для широкого круга прикладных задач: автоматического анализа резюме, мониторинга рынка труда, социолингвистических исследований.

В отличие от стандартных категорий именованных сущностей, таких как имена собственные, организации или географические объекты, категория «Профессия» остаётся относительно малоизученной и недостаточно представленной в существующих системах распознавания. Большинство известных NER-моделей не предусматривают выделение профессий как отдельного класса. Дополнительные трудности при обработке русскоязычных текстов создают морфологическое разнообразие наименований профессий и контекстная зависимость их употребления.

Использование универсальных моделей распознавания именованных сущностей, обученных на общих корпусах, не обеспечивает достаточного качества для распознавания профессий, поскольку такие модели не учитывают доменную специфику и контекст употребления профессиональной лексики. В результате возникает необходимость в разработке специализированных решений, ориентированных на данную категорию.

Целью данного исследования является сравнительный анализ различных подходов к распознаванию сущностей категории «Профессия» в русскоязычных текстах и определение метода, обеспечивающего наилучшее качество идентификации. Для достижения этой цели были реализованы и протестированы три подхода: словарный метод, статистическая модель на основе условных случайных полей (Conditional Random Fields, CRF) и нейросетевой метод с использованием предобученной модели RuBERT, адаптированной к доменной задаче путём дообучения на специализированном корпусе данных.

I. Классификация подходов для распознавания именованных сущностей

Подходы для распознавания именованных сущностей можно классифицировать на три основные группы, отражающие их историческую эволюцию [1]. Подробное описание приведено далее.

Методы, основанные на знаниях. Исторически первые системы распознавания именованных сущностей базировались на словарях, лингвистических правилах и регулярных выражениях. Эти методы основаны на экспертном знании предметной области и формализации языковых закономерностей. Они обеспечивают высокую точность в узких доменах и прозрачность интерпретации результатов, однако требуют значительных трудозатрат при создании и адаптации, а также плохо масштабируются на новые области и не способны к обобщению непредусмотренных языковых форм.

Методы, основанные на конструировании признаков. Следующий этап развития NER связан с применением методов машинного обучения. Наиболее эффективными в этой категории показали себя условные случайные поля. Ключевое преимущество CRF перед другими алгоритмами машинного обучения заключается в способности эффективно моделировать зависимости между наблюдениями и метками, учитывая контекстные взаимосвязи [2]. Таким образом, CRF способны обобщать закономерности на основе обучающих данных и сохранять относительную прозрачность модели. Однако их качество существенно зависит от ручного конструирования признаков и полноты обучающего корпуса.

Методы, основанные на глубоком обучении. Современный этап развития NER связан с применением глубоких нейронных сетей. Нейронные сети позволяют автоматически извлекать репрезентативные признаки сущностей из больших наборов данных без явного ручного конструи-

рования. Они способны выявлять сложные лингвистические закономерности и демонстрируют высокую адаптивность к различным языковым контекстам.

II. Методология исследования

Корпус данных. Для проведения экспериментов был создан специализированный корпус русскоязычных текстов, содержащий аннотированные упоминания профессий. Корпус был сформирован на основе текстов, собранных с онлайн-платформ по поиску работы, профессиональных форумов, объявлений о вакансиях и резюме, а также дополнен синтетическими данными.

Модели и методы. В исследовании сравниваются три подхода:

1. Словарный подход. Словарный подход основан на использовании специально разработанного словаря профессий, который включает несколько тысяч наименований профессий с учётом морфологических вариаций.
2. CRF. В качестве представителя подхода, основанного на конструировании признаков, используется модель CRF. Модель учитывает последовательные зависимости между токенами и использует совокупность лексических, морфологических, орфографических и контекстуальных признаков.
3. RuBERT. В качестве нейросетевого решения применяется предобученная языковая модель RuBERT, адаптированная к задаче распознавания профессий путём дополнительного обучения на специально размеченном корпусе русскоязычных текстов.

Метрики оценки. Качество моделей оценивалось с помощью стандартных метрик: $Precision = \frac{TP}{TP+FP}$; $Recall = \frac{TP}{TP+FN}$; $F1 = 2 * \frac{Precision*Recall}{Precision+Recall}$, где TP, FP и FN обозначают истинно положительные, ложно положительные и ложно отрицательные случаи распознавания соответственно [3].

III. Результаты

Таблица 1 демонстрирует сравнительные результаты для трёх методов: словарного подхода, CRF и RuBERT.

Таблица 1 – Сравнение методов распознавания профессий

Метод	**Precision**	**Recall**	**F1-score**
Словарь	0,98	0,55	0,70
CRF	0,84	0,72	0,78
RuBERT	0,87	0,85	0,86

Результаты показывают, что словарный подход обеспечивает высокую точность за счёт жёстких правил, но не охватывает вариативность естественного языка. CRF-модель достигает более сбалансированных показателей за счёт учёта контекста, однако её эффективность сильно зависит от качества используемых признаков.

Модель RuBERT демонстрирует наилучший результат среди рассмотренных подходов. Она превышает CRF по F1-мере на 8%. RuBERT корректно обрабатывает многословные профессии и учитывает морфологическую вариативность. Качественный анализ ошибок показал, что оставшиеся проблемы связаны с редкими или новыми наименованиями профессий, метафорическим употреблением и шумом в обучающих данных.

Заключение

Проведённый сравнительный анализ подтвердил гипотезу о необходимости специализированных решений для идентификации сущностей категории «Профессия» в русскоязычных текстах. Эволюция подходов от методов, основанных на знаниях, к методам, основанным на глубоком обучении, демонстрирует устойчивое улучшение качества.

Тонкая настройка предобученной модели RuBERT показала наилучший результат по F1-мере (0,86). Данная модель демонстрирует высокую способность к обобщению и возможность эффективно обрабатывать сложные случаи: многокомпонентные наименования, морфологическую вариативность и контекстно-зависимые упоминания профессий.

Практическая ценность работы заключается в возможности применения разработанных решений в системах анализа текстовых данных: при мониторинге рынка труда, обработке вакансий и резюме, а также в социологических исследованиях. Дальнейшее развитие исследования предполагает расширение корпуса данных, совершенствование методов разметки и создание более лёгких и энергоэффективных моделей, пригодных для внедрения в реальные информационные системы. Таким образом, исследование способствует развитию технологий обработки естественного языка и подтверждает эффективность нейросетевых моделей для решения прикладных задач извлечения информации.

1. Keraghel, I. Recent Advances in Named Entity Recognition: A Comprehensive Survey and Comparative Study / I. Keraghel, S. Morbieu, M. Nadif // arXiv preprint arXiv:2401.10825. — 2024.
2. Cariello, M. C. A Comparison between Named Entity Recognition Models in the Biomedical Domain / M. C. Cariello, A. Lenci, R. Mitkov // Proceedings of the Translation and Interpreting Technology Online Conference. – 2021. – P. 76–84.
3. Метрики классификации и регрессии [Электронный ресурс]. – Режим доступа: `https://education.yandex.ru/handbook/ml/article/metriki-klassifikacii-i-regressii`. – Дата доступа: 24.10.2025.

МЕТРИКИ СМЫСЛОВЫХ ПОДПРОСТРАНСТВ

Ивашенко В. П.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: ivashenko@bsuir.by

Рассматриваются подходы к построению метрик и соответствующих подпространств для случаев: интерпретируемого пространства скрытых признаков, выделяемых методами машинного обучения, включая нейросетевые методы, информационно-операционного пространства, ориентированого на интеграцию моделей с операционной семантикой большего и малого шага, подпространств на экстенсиональных замыканиях онтологических структур и графах состояний моделей обработки знаний, включая сильносвязные.

Введение

Разнообразие видов знаний и онтологических структур с семантикой различного вида в интеллектуальных системах требуют рассмотрения соответствующих моделей смысловых подпространств и методов их интеграции.

Разные виды смысловых пространств обладают разными структурными (топологическими) свойствами в некоторых случаях достаточными для задания метрики [1] в таких пространствах.

Сформулируем задачи, связанные с построением метрик и соответствующих смысловых подпространств [2]: (1) разработать онтологию видов смысловых пространств, которые позволяют соотносить известные смысловые пространства и подпространства общего смыслового пространства с их видами; (2) разработать модели метрических и топологических смысловых пространств конечных структур с операционной семантикой и конечных сильносвязных онтологических структур, согласованных с понятием экстенсионала и экстенсионального замыкания; (3) сформулировать требования к моделям метрического конечного операционно информационного смыслового пространства и метрического смыслового пространства протомультиверсальных чисел, согласующихся с расстоянием на действительных числах.

Под согласованностью топологического и метрического пространства понимается: каждая окрестность топологического пространства является объединением либо – конечным пересечением шаров, радиусы которых выражены метрикой относительно их центров.

I. Смысловые пространства и их виды

Онтология видов смысловых пространств (СП) представлена на рисунке 1.

Важным является вопрос, связанный с выражением понятия бесконечности в СП. Понятие бесконечности на некотором конечном этапе не может быть выражено конструктивно (актуально), поэтому бесконечность может быть сведена к недетерминизму. Однако при технической реализации, даже в случае недетерминизма (между конечным и бесконечным), в памяти интеллектуальной системы может быть зафиксирована не более чем счётная бесконечность. Несчётная бесконечность может быть выражена как недетерминизм возможного. На логическом уровне это может быть представлено средствами модальных логик: $\Diamond G \Diamond \bigcirc \top$; $\Diamond G \left(\Diamond \bigcirc \alpha \wedge \Diamond \bigcirc (\neg\alpha)\right)$.

При этом, данным выражениям могут соответствовать конечные циклические структуры.

II. Метрические подпространства и их метрики

Метрика и топология онтологических структур [2], включая сильносвязные. Пусть дано топологическое пространство $\langle X,S\rangle$. Обозначим топологическое замыкание множества M как $[M]_S$. Топологически сильносвязным множеством (классом неотделимости) будем называть класс отношения эквивалентности $[\{y\}]_S = [\{x\}]_S$.

Определим $C \in \left(2^X\right)^X$: $C(x) = \{y \,|[\{x\}]_S = [\{y\}]_S\}$. Допустим, что на классе $C(x)$ определена мера $\pi \in V^{C(x)}$, тогда введём меру $\mu \in V^X$ такую, что:

$$\mu(x) = 2 * (\alpha - 1) * \lim_{z \to \sup\{\pi(y) | y \in C(x)\}} \frac{\pi(x)}{z} + \\ +2 * \sup\left(\{\mu(y) \,|[C(y)]_S \subset [C(x)]_S\}\right),$$

где $\alpha \in V$ и $\alpha > 1$, а $V \subseteq \mathbb{R}$.

При этом пусть $\pi(x) = 0$ при $\{x\} = C(x)$.

Определим метрику рекурсивно следующим образом:

$$\upsilon(\langle x,z\rangle) = \\ = \begin{cases} |\mu(x) - \mu(z)| \,|\eta(\langle x,z\rangle) \\ \inf\left(\{\upsilon(\langle x,y\rangle) + \upsilon(\langle y,z\rangle) \,|y \in X\}\right) \,|\neg\eta(\langle x,z\rangle) \end{cases} \\ \eta(\langle x,z\rangle) = ([\{z\}]_S \subset [\{x\}]_S) \vee ([\{x\}]_S \subset [\{z\}]_S)$$

Протомультиверсальные числа [2]. Для модели протомультиверсальных чисел $\langle N,L,R,F,M,A\rangle$ введём признаки:

$$\tau(\langle\chi,\gamma\rangle) = \frac{\chi+\gamma}{2}; \sigma(\langle\alpha,\beta\rangle) = |\{\emptyset \,|\alpha\}|$$

и меры:

$$\delta(\langle\chi,\gamma\rangle) = \chi - \gamma; \lambda(\langle\alpha,\beta\rangle) = \sigma(\langle\alpha,\beta\rangle) = |\{\emptyset \,|\alpha \veebar \beta\}|$$

Выразим метрику на мультиверсальных числах:

$$\rho\left(\langle m,n\rangle\right)=\psi_1\left(\langle m.top,m.bottom,n.top,n.bottom\rangle\right)\odot$$
$$\odot\psi_2\left(\langle m.up,m.down,n.up,n.down\rangle\right)\odot$$
$$\odot\psi_3\left(\langle m.hi,m.low,n.hi,n.low\rangle\right)\odot\psi_4\left(\langle m,n,A\rangle\right)$$

$$\psi_1\left(\langle\alpha,\beta,\chi,\gamma\rangle\right)=$$
$$=\psi_6\left(\langle\varepsilon\left(\{\alpha\}\right),\varepsilon\left(\{\beta\}\right),\varepsilon\left(\{\chi\}\right),\varepsilon\left(\{\gamma\}\right)\rangle\right)$$
$$\psi_2\left(\langle\alpha,\beta,\chi,\gamma\rangle\right)=\odot_{f\in F}$$
$$\odot_{r\in R}\,\psi_1\left(\langle\alpha\left(f\right)\left(r\right),\beta\left(f\right)\left(r\right),\chi\left(f\right)\left(r\right),\gamma\left(f\right)\left(r\right)\rangle\right)$$
$$\psi_6\left(\langle\alpha,\beta,\chi,\gamma\rangle\right)=$$
$$=\varphi_1\left(\langle\delta,\alpha_2,\beta_2,\chi_2,\gamma_2\rangle\right)\odot\varphi_2\left(\langle\tau,\alpha_2,\beta_2,\chi_2,\gamma_2\rangle\right)\odot$$
$$\odot\varphi_3\left(\langle\lambda,\alpha_1,\beta_1,\chi_1,\gamma_1\rangle\right)\odot\varphi_4\left(\langle\sigma,\alpha_1,\beta_1,\chi_1,\gamma_1\rangle\right)$$

$$\varphi_i\left(\langle\phi,\alpha,\beta,\chi,\gamma\rangle\right)=\left|\phi\left(\langle\alpha,\beta\rangle\right)-\phi\left(\langle\chi,\gamma\rangle\right)\right|$$

$$\alpha\odot\beta=\lim_{t\to p}\left(\sqrt[t]{\alpha^t+\beta^t}\right);\varepsilon\left(S\right)=\left\{\begin{array}{l}0\,|\emptyset=S\\ x\,|x\in S\end{array}\right.$$

Функции ψ_3, ψ_4, ψ_5 определяются аналогично ψ_2.

Операционно-информационное пространство [2] задаётся отношением перехода R между конфигурациями из множества C, множеством операций O, множеством операторов K, множеством параметров P и множеством значений V. Рассмотрим переход между конфигурациями в информационно операционном пространстве. Каждый переход связан с выполнением одной или нескольких команд соответствующих операций. Каждая операция тратит время на чтение и изъятие входных данных, а также на обработку и запись выходных данных (значений параметров). Обозначим время работы операции o, исполняющей оператор k, над входным параметром q с входными данными i с выходным параметром p и выходными данными u через $t\left(\langle k,o,q,i,p,u\rangle\right)$. Время работы операции o зависит не менее чем линейно $c>1$ от соответствующего редакционного расстояния [1,2] $\rho\left(\langle i,u\rangle\right)$: $t\left(\langle k,o,q,i,p,u\rangle\right)\geq c*\rho\left(\langle i,u\rangle\right)$. Расстояние между конфигурациями c и s не менее максимального из времён сработавших операторов k_j в c: $\rho\left(\langle c,s\rangle\right)\geq\rho_1\left(\langle c,s\rangle\right)\geq\max\left(\left\{t\left(\left\langle k_j,o_j,q_{g_j},i_{g_j},p_{h_j},u_{h_j}\right\rangle\right)|j\right\}\right)$. Если конфигурации c и s отличаются и\или множеством операторов (K_c,K_s), операций (O_c,O_s), параметров (P_c,P_s), дуг (E_c,E_s), соединяющих операторы, параметры и операции, то расстояние $\rho\left(\langle c,s\rangle\right)$ и квазиметрика $\zeta\left(\langle c,s\rangle\right)$ между ними не меньше:

$$\rho\left(\langle c,s\rangle\right)\geq\zeta\left(\langle c,s\rangle\right)\geq\kappa_1*\rho_1\left(\langle c,s\rangle\right)\odot$$
$$\odot\omega_2\left(\langle\rho_2,\left(E_s/E_c\right),\left(E_c/E_s\right)\rangle\right)\odot$$
$$\odot\omega_3\left(\langle\rho_3,\left(P_s/P_c\right),\left(P_c/P_s\right)\rangle\right)\odot$$
$$\odot\omega_4\left(\langle\rho_4,\left(K_s/K_c\right),\left(K_c/K_s\right)\rangle\right)\odot$$
$$\odot\omega_5\left(\langle\rho_5,\left(O_s/O_c\right),\left(O_c/O_s\right)\rangle\right);$$
$$k_i\geq 0;\omega_i\left(\langle\rho,\alpha,\beta\rangle\right)=\kappa_i*$$
$$*\min(\{\rho\left(x\right)|\left(f\in\beta^\alpha\right)\wedge\left(f^{-1}\in\alpha^\beta\right)\wedge$$
$$\left(x\in f\cup\{\emptyset\}\times\beta/Dom\left(f^{-1}\right)\cup\alpha/Dom\left(f\right)\times\{\emptyset\}\right)\})$$

Как вариант, расстояние ρ_5 между операциями может быть оценено ($\rho_5=\delta$):

$$\delta\left(\langle\varphi,\psi\rangle\right)=\odot\left(\{0\}\cup\{\theta\left(\rho\left(\langle\chi,\gamma\rangle\right)\right)|\langle\chi,\gamma\rangle\in\varphi-\psi\}\right).$$

Заключение

Разработана онтология видов СП, включая метрические и топологические. Разработаны модели метрического и топологического СП конечных структур с операционной семантикой и конечных сильносвязных онтологических структур, согласованных с понятием экстенсионала и экстенсионального замыкания. Сформулированы требования к моделям метрического конечного операционно- информационного СП и метрического СП протомультиверсальных чисел, согласующихся с расстоянием на действительных числах. Открытым остаётся вопрос об уточнении топологии и метрики временных онтологических структур.

1. Петровский, А. Б. Пространства измеримых множеств мультимножеств. М. : Поли Принт Сервис. – 2016. – 324 с.
2. Ivashenko, V. Approaches to the Study of Semantic Space and Integrated Logical Inference Models Using Similarity, Difference and Other Measures = Подходы к исследованию семантического пространства и моделей интегрированного логического вывода с использованием мер сходства, различия и других мер / V. Ivashenko // Open Semantic Technologies for Intelligent Systems (OSTIS) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. В. Голенков [и др.]. – Минск, 2025. – Вып. 9. – С. 97–112.

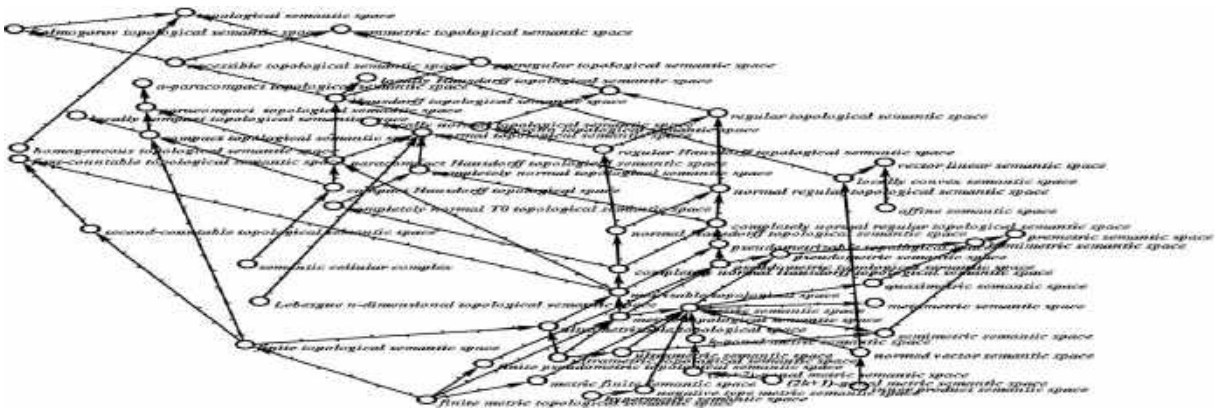


Рис. 1 – Онтология смысловых пространств

ПРИМЕНЕНИЕ АВТОКОДИРОВЩИКОВ И МЕТОДА АДАПТИРУЕМОГО КОЭФФИЦИЕНТА ОБУЧЕНИЯ ДЛЯ ПОНИЖЕНИЯ РАЗМЕРНОСТИ В СМЫСЛОВОМ ПРОСТРАНСТВЕ

Ивашенко В. П.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: ivashenko@bsuir.by

Рассматривается применение нейросетевых моделей и методов понижения размерности данных, использующих адаптивный (адаптируемый) коэффициент обучения, об элементах многомерного метрического смыслового пространствта. Используется метод точного аналитического вычисления адаптивного коэффициента обучения для линейного автокодировщика.

Введение

Многомерные метрические смысловые (под)пространства [1] (размерности δ) на некотором X образуются при построении из множества псевдометрических смысловых пространств, в соответствии со следующими теоремами 1 и 2 [2], а также при переходе от квазиметрических признаков к метрическим в условиях ограничений на связность и структуру метасвязей (максимально допустимым расстоянием m) представленных в смысловом пространстве знаний, в соответствии с выражением [3]:

$$\delta = \max \quad \left(\{1\} \cup \left\{ \left\lfloor \frac{(|X|-2)}{2*(m-1)} \right\rfloor \right\}\right).$$

Теорема 1. Пусть на множестве X заданы псевдометрическое с псевдометрикой π и псевдометрическое с метрикой μ пространства, а также $\lambda \geq 1$. Тогда $\langle X,\rho\rangle$ – псевдометрическое пространство, где

$$\rho\left(\langle\chi,\gamma\rangle\right) = \left(\left(\mu(\langle\chi,\gamma\rangle)^{\lambda}\right) + \left(\pi(\langle\chi,\gamma\rangle)^{\lambda}\right)\right)^{\left(\lambda^{-1}\right)}$$

или

$$\rho\left(\langle\chi,\gamma\rangle\right) = \max\left(\{\mu\left(\langle\chi,\gamma\rangle\right)\} \cup \{\pi\left(\langle\chi,\gamma\rangle\right)\}\right).$$

Теорема 2. Пусть на множестве X заданы псевдометрические с псевдометрикой π и псевдометрикой μ пространства, а также $\lambda \geq 1$ и $(\neg(\chi=\gamma)) \rightarrow (\mu\left(\langle\chi,\gamma\rangle\right) + \pi\left(\langle\chi,\gamma\rangle\right) \geq 0)$. Тогда $\langle X,\rho\rangle$ – метрическое пространство, где

$$\rho\left(\langle\chi,\gamma\rangle\right) = \left(\left(\mu(\langle\chi,\gamma\rangle)^{\lambda}\right) + \left(\pi(\langle\chi,\gamma\rangle)^{\lambda}\right)\right)^{\left(\lambda^{-1}\right)}$$

или

$$\rho\left(\langle\chi,\gamma\rangle\right) = \max\left(\{\mu\left(\langle\chi,\gamma\rangle\right)\} \cup \{\pi\left(\langle\chi,\gamma\rangle\right)\}\right).$$

I. Задача понижения размерности пространства

Необходимость в понижении размерности многомерного смыслвого пространства связана с необходимостью применения производительных, эффективных, высокопродуктивных алгоритмов поиска элементов и структур в смысловом пространстве. Размерность пространства сравнимая с количеством элементов в нём (линейнозависящая от их количества) является достаточным условиям неэффективности подобных алгоритмов, поэтому требуется сократить размерность соответствующих данных.

Кроме этого следует учесть, что механизмы, включая модели и методы понижения размерности смыслового пространства, с целью интерпретируемости и объяснимости получаемых с помощью них результатов, сами должны быть частью смыслового пространства [1,2,3,4].

Дано: множество образов $X = \left\{X^{(i)} \middle| X^{(i)} \in V^n\right\}$ из V^n, $p \leq n$.

Требуется: найти такие e из $\left(V^p\right)^{\left(V^n\right)}$ и d из $\left(V^n\right)^{\left(V^p\right)}$, что $\sum_i \left(e \circ d\left(X^{(i)}\right) - X^{(i)}\right)^2 \rightarrow \min$.

Задача понижения размерности пространства связана с таким направлением, как многомерное шкалирование [5]. Одним из методов для решения такой задачи являются искусственные нейросетевые модели – автокодировщики [6], которые бывают разных видов – принимающие значения из множества элементов поля с положительной натуральной характеристикой и характеристикой 0 (действительнозначные, комплекснозначные и др.)

Следует отметить, что если задана метрика μ на X ($\mu \in D^{\left(\left(V^n\right)^2\right)}$), то на $H = \left\{H^{(i)} \middle| \left(H^{(i)} = e\left(X^{(i)}\right)\right) \wedge \left(X^{(i)} \in V^n\right)\right\}$ также может быть задана метрика ρ:

$$\rho = \left\{\langle\langle\alpha,\beta\rangle, \langle d\left(\alpha\right), d\left(\beta\right)\rangle\rangle \middle| \langle\alpha,\beta\rangle \in \left(V^p\right)^2\right\} \circ \mu.$$

С целью представления их в смысловом пространстве среди комплекснознаных автокодировщиков будем выделять (комплекснозначные) рациональные автокодировщики. Все функции таких кодировщиков являются рациональными, включая функции активации f: линейная, (полу)ограниченная линейная, недоограниченная линейная (Leaky ReLU), рациональная сигмоида и

другие:

$$f(x) = \frac{1+|x|+x}{1+|x|} * \frac{x}{2}.$$

В случае линейного автокодировщика e и d имеют вид:

$$\begin{array}{l} e(X^{(i)}) = X^{(i)} * W^{(1)} - T^{(1)} \\ d(H^{(i)}) = H^{(i)} * W^{(2)} - T^{(2)} \end{array}.$$

II. Обучение линейного автокодировщика с адаптивным коэффициентом обучения

Решение задачи понижения размерности пространства сводится к обучению автокодировщика $D \subseteq \mathbb{R}$.

Пусть:

$$\begin{array}{l} H^{(i)} = \mathrm{X^{(i)}} * \mathrm{W^{(1)}} - \mathrm{T^{(1)}} \\ Y^{(i)} = H^{(i)} * \mathrm{W^{(2)}} - \mathrm{T^{(2)}} \\ D^{(i)} = X^{(i)} \\ \Delta\mathrm{T^{(2)}} = -\sum_i \left(D^{(i)} - Y^{(i)}\right) \\ \Delta\mathrm{T^{(1)}} = \Delta\mathrm{T^{(2)}} * \left(\mathrm{W^{(2)}}\right)^T \\ \Delta\mathrm{W^{(2)}} = \sum_i \left(H^{(i)}\right)^T * \left(D^{(i)} - Y^{(i)}\right) \\ \Delta\mathrm{W^{(1)}} = \left(\sum_i \left(X^{(i)}\right)^T * \left(D^{(i)} - Y^{(i)}\right)\right) * \left(\mathrm{W^{(2)}}\right)^T \\ \Delta H^{(i)} = \mathrm{X^{(i)}} * \Delta\mathrm{W^{(1)}} - \Delta\mathrm{T^{(1)}} \\ \Delta G^{(i)} = \Delta H^{(i)} * \Delta\mathrm{W^{(2)}} \\ \Delta Y^{(i)} = \Delta H^{(i)} * \mathrm{W^{(2)}} + H^{(i)} * \Delta\mathrm{W^{(2)}} - \Delta\mathrm{T^{(2)}} \\ e = \sum_i \left(Y^{(i)} - D^{(i)}\right)^2 \\ d = \sum_i \left(Y^{(i)} - D^{(i)}\right) \cdot \Delta Y^{(i)} \\ c = \sum_i \left(\Delta Y^{(i)}\right)^2 + 2 * \sum_i \left(Y^{(i)} - D^{(i)}\right) \cdot \Delta G^{(i)} \\ b = \sum_i \left(\Delta Y^{(i)} \cdot \Delta G^{(i)}\right) \\ a = \sum_i \left(\Delta G^{(i)}\right)^2 \\ r = \frac{2*a*c-3*b^2}{12*a^2} \\ q = \frac{b^3-a*b*c+2*a^2*d}{8*a^3} \end{array}$$

тогда адаптивный коэффициент обучения α:

$$\alpha \in \left\{ s - \frac{p}{s} - \frac{b}{2*a} \,\middle|\, s = \sqrt[3]{-q \pm \sqrt{r^3+q^2}} \right\}$$

и

$$\begin{array}{l} T^{(1)} \leftarrow T^{(1)} + \alpha * \Delta T^{(1)} \\ T^{(2)} \leftarrow T^{(2)} + \alpha * \Delta T^{(2)} \\ \mathrm{W^{(1)}} \leftarrow \mathrm{W^{(1)}} + \alpha * \Delta\mathrm{W^{(1)}} \\ \mathrm{W^{(2)}} \leftarrow \mathrm{W^{(2)}} + \alpha * \Delta\mathrm{W^{(2)}} \end{array}.$$

При этом область допустимых значений для α:

$$6 * a * \alpha^2 + 6 * b * \alpha + c \geq 0$$

, а ошибка достигает значения:

$$a * \alpha^4 + 2 * b * \alpha^3 + c * \alpha^2 + 2 * d * \alpha + e$$

.

В случае $a = 0$ α выражается из:

$$3 * b * \alpha^2 + c * \alpha + d = 0.$$

С целью обеспечения объяснимости и интерпретируемости результатов полученного решения множество V предполагается подмножеством модели протомультиверсальных чисел [4].

Заключение

Рассмотрена задача понижения размерности смыслового пространства, обоснована её актуальность. Предложены интерпретируемые (в рамках смыслового пространства) методы решения этой задачи на основе линейного автокодировщика с адаптивным коэффициентом обучения. Предложен метод расчёта адаптивного коэффициента обучения. На основе предложенного метода реализован алгоритм и проведено его тестирование в серии вычислительных экспериментов, результаты которых подтвердили корректность предложенного метода.

1. Manin, Yu. I., Marcolli, M. Semantic spaces. Math. Comput. Sci. – 2016. – 10, no. 4, – 459–477 pp.
2. Ivashenko, V. Approaches to the Study of Semantic Space and Integrated Logical Inference Models Using Similarity, Difference and Other Measures = Подходы к исследованию семантического пространства и моделей интегрированного логического вывода с использованием мер сходства, различия и других мер / V. Ivashenko // Open Semantic Technologies for Intelligent Systems (OSTIS) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : В. В. Голенков [и др.]. – Минск, 2025. – Вып. 9. – С. 97–112.
3. Ивашенко, В. П. Метрика, топология и персистентность в смысловом пространстве / В. П. Ивашенко // Информационные технологии и системы 2024 (ИТС 2024). – Минск: БГУИР, 2024. – С. 53–54.
4. Ivashenka, V. Semantic Space: between Continuity and Infinity / V. Ivashenka // Pattern Recognition and Information Processing (PRIP'2025) : Proceedings of the 17th International Conference, 16–18 September 2025, Minsk, Belarus / United institute of informatics problems of the National academy of sciences of Belarus. – Minsk : UIIP NASB, 2025. – P. 108–112.
5. Baldi, P. Deep Learning in Science. Cambridge: Cambridge University Press. – 2021. – 366 pp.
6. Nasir, S. A Survey on Multidimensional Scaling // S. Nasir, N. Haewoon, H. Mian, B. D. Muhammad, / ACM Computing Surveys. – 2018. – 1, no. 1, – 26 pp.
7. Головко, В. А. Нейросетевые технологии обработки данных : учеб. пособие / В. А.Головко, В. В. Краснопрошин. – Минск : БГУ, 2017. – 263 с.

СРАВНЕНИЕ ПОДХОДОВ К ПОСТРОЕНИЮ ХРОМАГРАММЫ ДЛЯ АНАЛИЗА АУДИО

Каминский А. В., Петровский Н. А.
Кафедра электронных вычислительных машин,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: sasha.pinsk2003@gmail.com, nick.petrovsky@bsuir.by

В работе проведён сравнительный анализ хромаграмм, построенных на основе мел-спектрограммы и константа-Q преобразования (CQT). Показано, что мел-хромаграмма плохо выделяет отдельные музыкальные ноты: энергия слишком равномерно распределена, границы нот не различимы. В отличие от неё, CQT-хромаграмма обеспечивает чёткое различие нот и хорошо отражает последовательность мелодии. Результаты оценивались визуально, демонстрируя преимущество CQT для анализа музыкальных сигналов с отдельными нотами.

Введение

Музыкальный сигнал представляет собой сложный нестационарный процесс, содержащий одновременно частотную, временную и гармоническую информацию.

Для анализа музыкального контента — таких задач, как определение аккордов, тональности, а также систем поиска по напеву (Query-by-Humming) — требуется устойчивое и компактное представление аудио, отражающее свойства звука независимо от тембра и октавы. Одним из наиболее распространённых способов такого представления является хромаграмма [1] (англ. chroma feature), описывающая распределение энергии по двенадцати высотным компонентам (C, C#, D, . . . , B) в каждый момент времени.

Построение хромаграммы основывается на частотно-временном анализе сигнала. Наиболее известными подходами являются использование мел-спектрограммы и константа-Q преобразование (CQT, Constant-Q Transform).

Цель работы — сравнение мел- и CQT-хромаграмм при анализе музыкальных сигналов. Сравнение выполнено теоретически и экспериментально на одном аудиофайле.

I. Теоретические основы

Мел-спектрограмма [2] представляет собой преобразование аудиосигнала, полученное на основе оконного преобразования Фурье (STFT) с последующим применением банка мел-фильтров, который аппроксимирует восприятие частоты человеческим слухом.

В отличие от линейного частотного разбиения в обычной спектрограмме, мел-шкала нелинейна: она имеет высокое разрешение на низких частотах и сниженное — на высоких, что отражает особенности слухового восприятия.

Сигнал $x[n]$ делится на перекрывающиеся участки длиной N. Каждый участок умножается на оконную функцию $w[n]$ (например, Ханна или Хэмминга), после чего вычисляется дискретное преобразование Фурье.

Чтобы перейти от линейной частотной шкалы к той, которая отражает, как человек действительно воспринимает высоту звука, используют мел-шкалу:

$$m(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right).$$

На основе этой зависимости строится банк мел-фильтров [3], состоящий из треугольных фильтров $H_i(k)$, равномерно распределённых по мел-шкале.

Каждый фильтр пропускает энергию только в своей частотной области и затухает к нулю на границах:

$$E_i(m) = \sum_{k=0}^{N/2} |X(m,k)|^2 \, H_i(k), \quad i = 1,2,\ldots,M.$$

Константа-Q преобразование [4] использует окна с одинаковым отношением частоты к полосе, обеспечивая логарифмическое распределение частот, соответствующее музыкальной шкале в отличие от линейной сетки STFT.

Частоты дискретных бинов f_k определяются как:

$$f_k = f_{\min} \cdot 2^{k/B}, \quad k = 0,1,\ldots,K-1.$$

Так как частоты f_k расположены логарифмически и совпадают с музыкальными нотами, CQT естественным образом подходит для построения хромаграммы. Каждый бин k сопоставляется высотной компоненте по формуле:

$$p = \text{round}\left(12 \log_2 \frac{f_k}{f_{\text{ref}}}\right) \bmod 12,$$

где $f_{\text{ref}} = 440$ – опорная частота (нота A4).

Энергии всех бинов, относящихся к одной нотной компоненте, суммируются, в результате чего получается 12-мерный вектор — хромавектор для текущего временного кадра.

II. Реализация построения хромаграмм

В качестве исходных данных для построения хромаграммы использовался аудиофайл, содержащий вокальную последовательность, в которой

исполняются ноты в диапазоне от C4 до B4, а затем в обратном направлении.

Хромаграмма на основе мел-спектрограммы строилась в несколько этапов. Сначала аудиосигнал нормализовался по амплитуде, далее разбивался на перекрывающиеся окна фиксированной длины с применением оконной функции (Хэмминга), после чего для каждого окна вычислялось дискретное преобразование Фурье (ДПФ) с перекрытием 50%. Амплитудный спектр каждого окна преобразовывался в энергетическую форму и подавался на вход банка мел-фильтров. Результатом этого этапа являлась мел-спектрограмма, отражающая распределение энергии по восприятию человека.

В итоге для каждого временного окна формировался вектор из 12 компонент, отражающий относительную энергию нотных компонент. Совокупность таких векторов во времени образует хромаграмму, которая является компактным представлением тональной структуры музыкального сигнала.

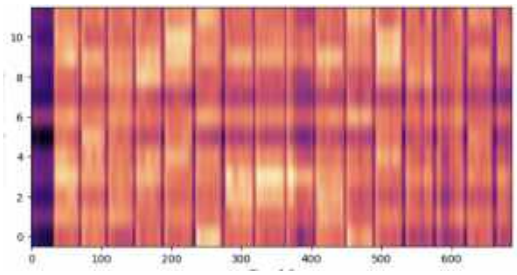


Рис. 1 – Полученная хромаграмма

Для построения хромаграммы на основе константа-Q преобразования аудиосигнал предварительно нормализовывался и разделялся на перекрывающиеся окна. Для каждого окна вычислялось константа-Q преобразование, результатом которого является спектр с логарифмическим распределением частотных бинов. Из полученной спектрограммы извлекались значения амплитуд для фиксированного числа бинов, покрывающих несколько октав. Далее для каждого бина определялась принадлежность к одному из 12 компонент высоты, соответствующих полутонам в октаве. Таким образом формировалась CQT-хромаграмма, отображающая распределение энергии по музыкальным компонентам высоты звука.

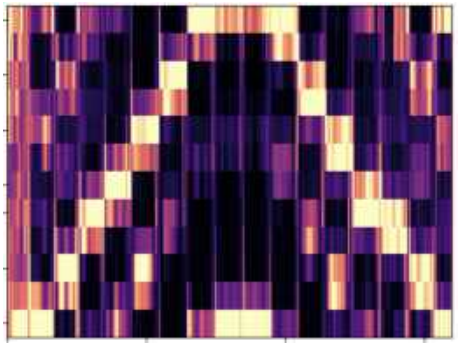


Рис. 2 – Хромаграмма после CQT

III. Результаты

Были построены хромаграммы на основе мел-спектрограммы и константа-Q преобразования. Мел-хромаграмма оказалась сглаженной и слабо различала ноты, тогда как CQT обеспечила чёткое разделение полутонов и лучшее соответствие последовательности нот.

Для количественного сравнения использовались косинусное расстояние и динамическое выравнивание по времени. CQT-хромаграмма показала большее сходство с эталоном: косинусное расстояние — 0.43 против 0.26 для мел, нормализованное DTW [5] — 0.385 против 0.735 соответственно. Это подтверждает преимущество CQT при анализе музыкальных сигналов.

1. Islam R., Tarique M. A novel convolutional neural network based dysphonic voice detection algorithm using chromagram [Electronic resource] / R. Islam, M. Tarique – University of Windsor, 2023. – Mode of access: `https://surl.li/ektwja`. – Date of access: 24.10.2025.
2. Banuroopa, K., Shanmuga Priyaa, D. A novel voiceprint using ensembled Mel-Chromagram for speaker recognition [Electronic resource] / K. Banuroopa, D. Shanmuga Priyaa. – 2022. – International Journal of Health Sciences. – Vol. 6, S4. – P. 8043–8056. – Mode of access: `https://surl.li/ukkffi`. – Date of access: 24.10.2025.
3. Vlaj, D., Kotnik, B., Horvat, B., Kačič, Z. A Computationally Efficient Mel-Filter Bank VAD Algorithm for Distributed Speech Recognition Systems / D. Vlaj, B. Kotnik, B. Horvat, Z. Kačič // EURASIP J. Advances in Signal Processing. – Vol. 2005, Article 561951 (2005). – Mode of access: `https://doi.org/10.1155/ASP.2005.487`. – Date of access: 24.10.2025.
4. Wolf-Monheim, F. Spectral and Rhythm Features for Audio Classification with Deep Convolutional Neural Networks [Electronic resource] / F. Wolf-Monheim. – 2024. – arXiv:2410.06927. – Mode of access: `https://arxiv.org/abs/2410.06927`. – Date of access: 24.10.2025.
5. Senin P. Dynamic Time Warping Algorithm Review [Electronic resource] / P. Senin – University of Hawaii at Manoa, December 2008. – Mode of access: `https://clck.ru/3Q4ohq`. – Date of access: 24.10.2025.

ИНТЕЛЛЕКТУАЛЬНЫЙ АССИТЕНТ ДЛЯ ПЕРСОНАЛИЗАЦИИ ПРОФЕССИОНАЛЬНОЙ ОРИЕНТАЦИИ УЧАЩИХСЯ

Клещукевич А. А., Владимирский П. Н.
Учреждение образования «Национальный детский технопарк»
Минск, Республика Беларусь
E-mail: kleshchukevichart@gmail.com, petrvladimirsky@yandex.ru

В работе рассматривается создание интеллектуального ассистента для оказания помощи учащимся в профессиональной ориентации. Система учитывает индивидуальные особенности пользователя и предоставляет персональные рекомендации. В основе лежит технология OSTIS, обеспечивающая онтологическое представление знаний и расширяемость. Проведено тестирование ключевых модулей, подтверждена корректность работы и практическая применимость.

Введение

Быстрые изменения на рынке труда и появление новых специальностей усложняют выбор профессии для школьников и абитуриентов. Большинство популярных профориентационных тестов формируют усреднённые рекомендации и не учитывают индивидуальные черты личности, контекст обучения и реальные образовательные траектории. Это снижает точность рекомендаций и приводит к ошибкам в выборе.[1] Поэтому подход, включающий в себя искуственный интеллект, к профориентации предполагает процесс подбора профессии и карьерного планирования, позволяет снизить риск ошибок, повысить качество конечного результата, учитывает потребности людей, нуждающихся в профориентации. [2] В работе рассматривается интеллектуальная система, которая позволяет познакомить учащихся с различными профессиями, а также помочь им с выбором своей.

I. Актуальность и постановка задачи

К ключевым проблемам современных способов профориентации относят: недостаточную персонализацию, скудную обратную связь и отсутствие прозрачных объяснений результатов. Задача исследования – построить ассистента, который:

- учитывает интересы, навыки и характерологические особенности;
- предоставляет структурированные сведения о профессиях, компетенциях, учебных заведениях и проходных баллах;
- поддерживает экспресс- и расширенные профтесты, хранит историю и позволяет давать обратную связь;
- формирует объяснимые рекомендации.

II. Методы и технологическая основа

В качестве основы интеллектуального ассистента используется технология OSTIS (Open Semantic Technology for Intelligent Systems) [3,4], являющаяся технологией, используемой для создания интеллектуальных систем, которая может эффективно работать с неструктурированными данными, облегчать представление знаний и поддерживать процессы управления знаниями.

III. Архитектура и функциональные модули

Система включает три ключевых компонента:

- **Базу знаний**, в рамках которой было выделено 5 основных предметных областей:
  - Предметная область пользователя – для хранения знаний о пользователе и информации о нём
  - Предметная область учебных заведений – для хранения знаний об учебных заведениях
  - Предметная область профессий – для хранения знаний о профессиях
  - Предметная область профориентационных тестов – для хранения знаний о профориентационных тестах, которые пользователь может пройти
- **Решатель задач, который содержит агенты**: получения списков и подробных описаний профессий; поиска учебных заведений по профессии; выдачи экспресс/расширенных тестов; анализа результатов; получения рекомендаций по профессиям; получения списка тестов.
- **Пользовательский интерфейс**, в рамках которого было реализовано 5 страниц:
  - Главная страница – начальный экран системы, содержащий навигацию по основным разделам и краткое описание возможностей.
  - Каталог профессий – раздел, предоставляющий доступ к списку профессий и их подробным описаниям.
  - Каталог учебных заведений – раздел, предоставляющий доступ к списку высших учебных заведениях и их подробным описаниям.
  - Каталог тестов – раздел, предоставляющий доступ к списку профориентаци-

онных тестов, с возможностью выбора и прохождения.
- Страница результатов тестов – страница, содержащая итоги прохождения тестов и рекомендации по выбору профессии.

Взаимодействие между модулями реализовано таким образом, что новые сущности можно добавлять.

IV. Описание ключевых агентов

- **Агент получения рекомендаций**: по результатам тестов формирует список предсказанных профессий, сопоставляя его с интересами и личными качествами, ранжирует и возвращает топ-k профессий с пояснениями на основе ответов.
- **Агент поиска профориентационных тестов**: получает все имеющиеся профориентационные тесты в базе знаний и формирует из них список.
- **Агент поиска профессий**: получает все имеющиеся профессии в базе знаний и формирует из них список.
- **Агент поиска учебных заведений**: получает все имеющиеся учебные заведения в базе знаний и формирует из них список.
- **Агент анализа профориентационного теста**: получает результаты прохождения теста и формирует на основе их анализа итоговую статистику по подходящим профессиях
- **Агент получения подробного описания по профессии**: формирует подробное описание по названию профессии
- **Агент получения профессии по названию**: получает все имеющиеся профессии и формирует список подходящих по названию профессий
- **Агент получения экспресс теста**: получает все имеющиеся профориентационные тесты в базе знаний и формирует из них список, подходящий под критерии экспресс теста.
- **Агент получения расширенного теста**: получает все имеющиеся профориентационные тесты в базе знаний и формирует из них список, подходящий под критерии расширенного теста.

V. Реализация и результаты

В данной интеллектуальной системе описаны онтологии профессий и учебных заведений; реализованы агенты: получения списка профессий, получения подробного описания профессии, получения списка вузов и поиск вузов по профессии, получения тестов (экспресс/расширенный), анализа результатов и выдачи рекомендаций. В прототипе доступны сценарии: прохождение тестов, просмотр рекомендаций. Тестирование подтвердило целостность онтологической модели, корректность работы агентов и персонализацию рекомендаций. Ассистент демонстрирует объяснимость: к каждой рекомендации прилагаются ключевые признаки, повлиявшие на выбор.

VI. Заключение

Разработан интеллектуальный ассистент для профориентации на основе технологии OSTIS. Подход обеспечивает масштабируемость, объяснимость и практическую полезность для школьников, абитуриентов и центров профориентации. В дальнейшем планируется расширение базы знаний, интеграция с образовательными платформами, улучшения методов обработки результатов профориентационных тестов с целью улучшения качества персональных рекомендаций.

Список литературы

1. Новые подходы к профессиональной ориентации в школе в условиях изменяющегося мира профессий [Электронный ресурс]. – Режим доступа: https://elar.urfu.ru/bitstream/10995/80855/1/iurp-2020-195-21.pdf?ysclid=mh3iyhns89830914557. – Дата доступа: 25.09.2025.
2. Перспективы применения искусственного интеллекта в профориентационной деятельности [Электронный ресурс]. – Режим доступа: https://cyberleninka.ru/article/n/perspektivy-primeneniya-iskusstvennogo-intellekta-v-proforientatsionnoy-deyatelnosti/viewer. – Дата доступа: 28.09.2025.
3. Проект Технологии OSTIS [Электронный ресурс]. – Режим доступа: https://github.com/ostis-ai/. – Дата доступа: 17.10.2025.
4. Open Semantic Technology for Intelligent Systems [Электронный ресурс]. – Режим доступа: http://conf.ostis.net/. – Дата доступа: 09.10.2025.

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ РАЗРАБОТКЕ И ВНЕДРЕНИИ ЭЛЕКТРОННЫХ ОБРАЗОВАТЕЛЬНЫХ РЕСУРСОВ

Ковалев М. В., Кривенков А. В., Мазаник Е. Н., Терешкова А. С., Шнейдеров Е. Н., Шункевич Д. В.
Кафедра интеллектуальных информационных технологий,
Центр развития дистанционного образования,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {kovalev, krivenkov, mazanik, tereshkova, shneiderov, shunkevich}@bsuir.by

В работе рассматриваются подходы к применению технологий искусственного интеллекта для автоматизации процессов разработки и внедрения электронных образовательных ресурсов на основе интеграции нейросетевых и семантических моделей.

Введение

В эпоху цифровизации важное значение для организации учебного процесса в учреждениях высшего образования играют информационно-коммуникационные технологии (ИКТ) и электронные образовательные ресурсы (ЭОР), применение которых делает учебный процесс более гибким и эффективным для его участников [1].

В настоящее время в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» (БГУИР) реализуется экспериментальный проект «Апробация адаптивной модели использования ИКТ в образовательной деятельности учреждения высшего образования», основной целью которого является обоснование и внедрение в учреждении высшего образования набора критериев и средств, позволяющих выстроить систему подходов к использованию ИКТ в образовательном процессе.

Одним из таких критериев является наличие ЭОР, разработка которых подразумевает создание, размещение и настройку учебных материалов в системе электронного обучения (СЭО) – платформе, позволяющей хранить и распространять учебные материалы, а также отслеживать прогресс обучения и его результаты.

Развитие технологий искусственного интеллекта (ИИ) позволяет вывести процессы разработки и внедрения ЭОР на принципиально новый уровень.

В данной работе будут рассмотрены подходы к применению технологий ИИ при разработке ЭОР, основанные на интеграции современных решений на базе искусственных нейронных сетей с открытыми семантическими технологиями.

I. Проблемы разработки и внедрения ЭОР и подходы к их решению

Опыт внедрения ЭОР в образовательный процесс БГУИР позволил выявить следующие основные технические проблемы, которые могут быть в той или иной степени решены с применением технологий искусственного интеллекта:

1. Высокая трудоемкость разработки материалов для контроля знаний, в частности, разработка больших банков вопросов или вариантов практических заданий.

В качестве решения предлагается использовать технологии ИИ для решения рутинных задач в процессе генерации учебных материалов, оценки их качества и соответствия требованиям.

Реализация такого решения предполагает использование ряда технологий ИИ:

- большие языковые модели (Large Language Models, LLM) для генерации тестов, заданий и методических материалов;
- база знаний для структурирования учебных материалов;
- средства на основе LLM в интеграции с формализованным описанием в базе знаний для автоматической проверки соответствия нормативным документам;
- средства формирований рекомендаций по улучшению материалов как на основе экспертных правил, так и на основе средств машинного обучения.

Пример использования в этом случае может выглядеть следующим образом:

- преподаватель загружает лекции и основные темы курса в систему;
- система автоматически генерирует необходимые материалы;
- преподаватель получает автоматическую оценку качества материалов;
- система предлагает рекомендации по улучшению (например, добавление практических примеров или уточнение формулировок);
- преподаватель редактирует материалы или принимает предложенные улучшения;
- после подтверждения преподаватель размещает полученные материалы в СЭО.

2. Высокая трудоемкость размещения материалов в СЭО, отсутствие навыков настройки элементов курса в СЭО.

В качестве одного из решений предлагается использовать указанные в предыдущем пункте

технологии ИИ для разработки интеллектуального помощника, который будет давать интерактивные подсказки в процессе размещения материалов и настройки элементов курса, а также частично автоматизировать данные процессы.

3. Дублирование учебных материалов и информации о результатах образовательного процесса в различных информационных системах университета.

В качестве решения данной проблемы предлагается создание централизованной базы знаний, описывающей различные информационные системы, и разработка единого интерфейса, который будет работать конкретно с этой базой знаний, а не с множеством систем. Автоматическая синхронизация между информационными системами университета избавит от необходимости дублирования информации.

Такое решение предполагает реализацию следующих ИИ-компонентов:

- база знаний для связывания различных информационных систем университета и ее автоматическая синхронизация с ними;
- специализированные интеллектуальные агенты для интеграции API различных порталов университета.

4. Низкая оригинальность материалов.

Для решения данной проблемы предлагается разработать подсистему для автоматического анализа оригинальности материала и формирования рекомендаций по улучшению его уникальности, что предполагает:

- интеграцию с антиплагиатными сервисами;
- применение LLM для генерации новых версий материалов с сохранением смысла для случаев, когда повышение оригинальности материала другими путями невозможно или нецелесообразно (например, когда часть материала касается общеизвестных фундаментальных вопросов, но исключение этой части материала нарушит целостность курса).

5. Высокая трудоемкость ручной разработки учебно-программной документации (УПД) и сопроводительных документов.

Для решения данной проблемы предлагается разработать подсистему автоматизации отдельных этапов разработки указанных документов, в которой будут применяться:

- LLM для автоматизации процесса создания УПД;
- базы знаний для описания образовательных стандартов и других нормативно-правовых актов, регламентирующих структуру и содержание УПД и сопроводительных документов;
- интеллектуальных агентов для автоматизации проверки соответствия УПД и сопроводительных документов необходимым требованиям.

II. Принципы реализации предлагаемого подхода

Как видно из представленного описания, эффективное решение выявленных проблем предполагает интеграцию:

- базы знаний для централизованного хранения и организации ЭОР и УПД, требований к их содержанию и оформлению, а также организации связей между различными информационными системами университета;
- LLM для генерации и обработки текстовых материалов;
- интеллектуальных агентов для обработки базы знаний и интеграции системы с различными сервисами посредством API.

Для реализации такого подхода предполагается использовать идеи, сформулированные в работе [2]. На рисунке 1 показана общая архитектура системы, реализующей указанные принципы интеграции.

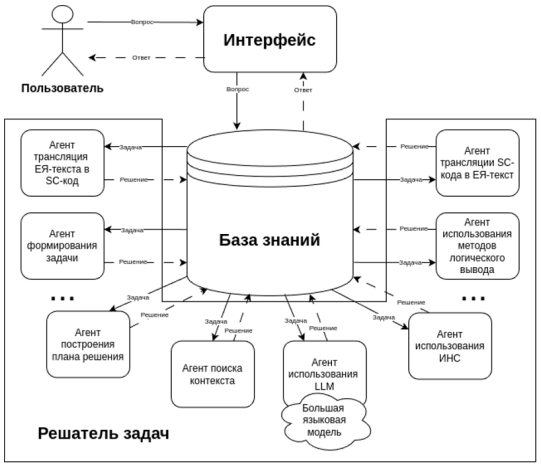


Рис. 1 – Архитектура системы

III. Заключение

В работе рассмотрены принципы применения интеллектуальных технологий для автоматизации процессов разработки и внедрения ЭОР, предложенные с учетом многолетнего опыта работы БГУИР в данной сфере.

IV. Список литературы

1. Опыт и перспективы развития дистанционного образования в БГУИР / А. В. Кривенков [и др.] // Дистанционное обучение – образовательная среда XXI века : материалы XII Международной научно-методической конференции, Минск, 26 мая 2022 г. / редкол. : Е. Н. Шнейдеров [и др.]. – – Минск : БГУИР, 2022. – С. 20.
2. Approaches of Neuro-Symbolic Integration: Large Language Models and Knowledge Bases = Подходы к нейро-символической интеграции: большие языковые модели и базы знаний / V. Golovko, V. Golenkov, M. Kovalev [et al.] // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS) : сборник научных трудов / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : В. В. Голенков [и др.]. – Минск, 2025. – Вып. 9. – С. 87–96.

РАЗРАБОТКА МОБИЛЬНОГО ПРИЛОЖЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ ПОВРЕЖДЕНИЙ НА АВТОМОБИЛЯХ

Колистратов М. В., Шинкарев И. А.
Кафедра Инфокоммуникационных технологий,
Национальный исследовательский технологический университет МИСИС
Москва, Российская Федерация
E-mail: kolistratov.mv@misis.ru, m2010469@edu.misis.ru

Рассматривается метод обнаружения повреждений автомобилей в режиме реального времени, реализованный в виде мобильного приложения. Использован гибридный подход с глубокими нейронными сетями (YOLOv8 + GNN)

Введение

Определение повреждений автомобилей с помощью мобильных устройств, оснащенных соответствующим приложением, является важной задачей анализа изображений, направленной на повышение эффективности оценки снимков повреждений для упрощения работы страховых компаний. С решением этой задачи человек справляется не всегда, поскольку для такого анализа требуется высокая точность и надежность распознавания. Например, при аренде автомобилей в каршеринговых компаниях перед началом поездки необходимо тщательно осмотреть выбранную машину на предмет повреждений. Важность решения этой задачи также актуальна для страхования автомобилей. Основные проблемы разработки метода обнаружения повреждений автомобилей связаны с необходимостью минимизировать вычислительные ресурсы, необходимые для работы мобильного приложения, но при этом точно распознавать типичные повреждения, такие как царапины, вмятины, разбитые фары, сдутые шины и т.д. Большинство существующих решений используют для детекции требующие значительных вычислительных ресурсов нейронные сети, что ограничивает их применение в мобильных приложениях. Кроме того, многие из этих решений не адаптированы к обработке реальных изображений. Несмотря на сложности реализации такого программного модуля, задача остается актуальной в сфере каршеринга и привлекает большое внимание страховых компаний, ведь ее решение может значительно улучшить качество и скорость анализа повреждений. Прогресс в данном направлении связан с появлением новых технологий детекции и оптимизации моделей нейронных сетей для мобильных устройств, которые позволяют находить эффективные решения для улучшения процессов оценки повреждений на автомобилях. Учитывая ограничения существующих решений, исследования направлены на разработку нового метода детекции повреждений на автомобилях, оптимизированного для использования в мобильных приложениях. Предлагаемый метод должен отличаться низким потреблением вычислительных ресурсов и высокой точностью детекции, что сделает его применимым в мобильных приложениях каршеринговых компаний, а также компаний, занимающихся страхованием, куплей-продажей подержанных автомобилей, где требуется оперативная и точная оценка их внешнего состояния.

I. Существующие методы решения

Для обнаружения и анализа повреждений на автомобилях популярно использование сверточных нейронных сетей (CNN). Большинство решений автоматизируют классификацию повреждений (вмятины корпуса, разбитые стекла и фары, царапины, трещины) и предсказывают их возникновение. CNN анализирует изображения на различных уровнях, автоматически выделяя особенности и выявляя мелкие дефекты, что делает ее идеальной для оценки состояния автомобилей. Классификация выполняется на датасетах с различными повреждениями. Mask R-CNN используется для сегментации, разделения и декомпозиции объектов, точно выделяя области повреждений и определяя границы и классы, при этом требуя значительных вычислительных ресурсов.

В задачах применяют глубокое обучение и классификацию изображений: обнаружение автомобилей на необработанных изображениях, поиск дефектов на выделенных авто и классификацию повреждений. Используются CNN с трансферным обучением, обеспечивающим высокую точность. Также применяется быстрая маска R-CNN для эффективного прогнозирования. Архитектура Faster R-CNN состоит из сверточной части для извлечения признаков и детектора объектов для обнаружения повреждений, извлекая уровни абстракций из изображений. Single Shot MultiBox Detector (SSD) работает быстрее, чем Faster R-CNN, включает сверточную часть и детектор [1], но имеет проблемы с точностью. Обучение CNN минимизирует функцию потерь. Сеть обучается на изображениях с повреждениями для правильной классификации и обнаружения.

Лучшими показателями детектирования повреждений обладают модели YOLO от Ultralytics. Модель YOLOv5 (2020) эффективна и проста, обрабатывает кадры в реальном времени за один

проход. YOLOv8 (2023) превосходит YOLOv5 по производительности, скорости и точности, и отлично подходит для задачи детекции повреждений [2].

II. Разработка метода обнаружения повреждений автомобилей

На текущий момент не существует модели, решающей задачу нахождения повреждений автомобилей, которая позволяла бы достичь нужной точности при низком потреблении вычислительных ресурсов и могла быть интегрирована в мобильное приложение.

Предложен метод, который отлично подходит для использования в мобильных приложениях за счет низкого потребления ресурсов, а также имеет конкурентноспособные показатели точности детекции. Разработанный метод основан на использовании сверточной нейронной сети глубокого обучения и использует разбиение результатов модели на графы для увеличения точности выявления повреждений. В роли сверточной нейронной сети выступает модель YOLOv8, которая обладает наилучшими показателями для обучения в области детекции. Так как версия YOLOv8 появилась недавно, модель на ее основе обучалась с нуля.

После обнаружения объектов на изображении с помощью модели YOLO программа получает координаты и классы найденных объектов. Каждый объект становится узлом графа и хранит в себе такие параметры как id класса и уверенность, а связи между объектами – ребрами, которые хранят расстояние между объектами. Далее применяется простая графическая нейронная сеть (GNN), определяющая достоверность полученных результатов работы модели YOLO и удаляет лишние объекты. Удаление происходит на основе определенных критериев, таких как превышение допустимого количества одного из классов и превышение допустимого расстояния между ними. На вход модели принимается матрица, в которой указаны узлы-объекты и ребра в качестве расстояния между объектами. Далее GNN проверяет матрицу и ищет нарушения, после чего удаляет неверно определенные объекты. Это способствует улучшению точности результатов детекции без лишней нагрузки на вычислительные ресурсы устройства.

III. Результаты

Для получения визуальных результатов использовался датасет, содержащий изображения 6 различных типов повреждений, таких как разбитое стекло и фары, сдутые шины, царапины, вмятины и трещины. Помимо датасета использовались реальные фотографии с мест аварий.

В рамках экспериментов сравнивались веса после каждого этапа обучения с использованием метода разбиения на графы и без него. Всего проводилось 10 этапов тренировки модели на специально созданном для обучения модели датасете. Основными критериями оценки на разных этапах обучения использовались две метрики: точность (Precision) и полнота (Recall). После 5 этапов обучения модели демонстрировали недостаточно точные результаты работы. На рис. 1 видно, что модель правильно определила царапины, однако ошибочно выделила фару. Для получения удовлетворительных результатов (уровень точности выше 0,98) произвели 10 этапов обучения. Множественные повреждения транспортных стредств выявляются с высокой веротностью (рис. 2)

Рис. 1 – Промежуточные результаты работы модели на реальных данных

Рис. 2 – Результаты работы модели на реальных данных

Разработанное приложение может быть использовано для решения таких задач, как аренда автомобилей, страхование и оценка их состояния, где критически важна точность распознавания повреждений

IV. Список литературы

1. Everingham M. et al. The pascal visual object classes (voc) challenge // International journal of computer vision. – 2010. – T.88 – P. 303-338
2. Wang, H. A novel multi-level data fusion and anomaly detection approach for infrastructure damage identification and localisation / H. Wang, G. Barone, A. Smith // Engineering Structures. – 2023 – T. 292. – P. 116473

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА ПЕРСОНАЛИЗАЦИИ ОБУЧЕНИЯ УЧАЩИХСЯ НА ПРИМЕРЕ ГЕОМЕТРИИ

Кретов А. А., Сащеко Е. А.
УО "Национальный детский технопарк"
Минск, Республика Беларусь
E-mail: {20kretovartem000, sashchekaea}@gmail.com

В работе рассмотрена интеллектуальная обучающая система по геометрии, разработанная для повышения эффективности обучения за счет персонализации образовательного процесса и адаптации содержания под потребности учащегося.

Введение

Одним из ключевых направлений повышения эффективности образовательного процесса является персонализация обучения. Актуальность данной темы обусловлена необходимостью повышения качества образования и внедрения современных подходов, способных учитывать индивидуальные особенности учащихся. В традиционной системе все учащиеся получают одинаковый объем и способ подачи материала, что не учитывает их когнитивные и физиологические особенности, уровень подготовки, темп усвоения знаний. Особенно остро эта проблема проявляется при изучении геометрии, где требуется развитое пространственное мышление, абстрактное воображение и логическое построение рассуждений. Новизна данной темы заключается в создании моделей, методов и средств персонализации обучения, ориентированных на адаптацию геометрического материала с учетом индивидуальных особенностей учащихся, уровня их подготовки и динамики освоения тем.

В данной работе рассматривается интеллектуальная система по геометрии, способная обучать пользователя, учитывая его индивидуальные особенности и характеристики, анализируя уровень прогресса и знаний в процессе обучения.

I. Структура системы

В качестве основы для интеллектуальной системы была выбрана Технология OSTIS, которая позволяет значительно упростить процесс проектирования и разработки интеллектуальных систем на протяжении всего жизненного цикла [1].

Каждая интеллектуальная система на базе технологии OSTIS состоит из 3 ключевых компонентов: базы знаний, решателя задач и пользовательского интерфейса.

В рамках решателя задач системы выделено 14 агентов:

- агент регистрации;
- агент рекомендации тем для изучения/повторения;
- агент рекомендации тем для прохождения теста/решения задачи;
- агент рекомендации урока для прохождения;
- агент рекомендации теста для прохождения;
- агент рекомендации задачи для решения;
- агент упрощения уровня;
- агент просмотра прогресса;
- агент принудительного увеличения сложности;
- агент запуска теста;
- агент подбора диагностического теста;
- агент добавления ответа к вопросу;
- агент получения следующего вопроса;
- агент завершения теста.

Вышеперечисленные агенты были написаны на языке программирования Python с использованием библиотек py-sc-client и py-sc-kpm.

Пользовательский интерфейс разработан на базе платформы Telegram в виде Telegram-бота.

Пользователю доступны следующие 13 команд:

- команда регистрации;
- команда запуска диагностического теста;
- команда открытия справочника;
- команда получения рекомендованных тем для изучения/повторения;
- команда открытия личного кабинета;
- команда начала рефлексии;
- команда получения рекомендованных тем для прохождения теста;
- команда получения рекомендованных тем для решения задачи;
- команда изменения сложности по результатам теста;
- команда принудительного изменения сложности;
- команда начала урока;
- команда начала теста;
- команда получения задачи;

II. Возможности системы

При первом взаимодействии с системой пользователь проходит регистрацию, указывая класс обучения, а также автоматически определяется уникальный ID пользователя в Telegram. После регистрации система предлагает пройти диагностический тест, включающий в себя вопросы по всему перечню тем геометрии, соответствующих школьной программе 9 класса, который определя-

ет уровень знаний пользователя как в общем, так и по отдельным темам. Во время тестирования система адаптирует сложность вопросов: она может как повышать, так и снижать уровень сложности, а также завершить тест в случае достаточности полученных ответов для оценки знаний. После теста она сообщает результаты и итоговый уровень знаний по прохождению теста.

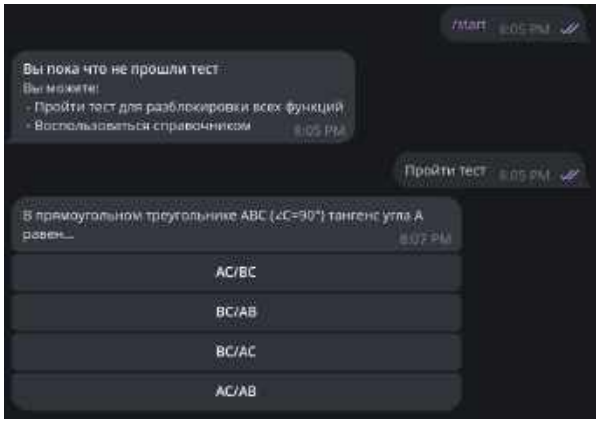


Рис. 1 – Прохождение теста пользователем.

Пользователь имеет доступ к личному кабинету, где отображаются его данные: имя, ID, класс обучения, количество достижений, уровень знаний и уровень самооценки. В личном кабинете пользователь может самостоятельно оценить свой уровень знаний, а система в ответ рассчитает величину "отклонения"самооценки от реального уровня знаний. Помимо этого, пользователь может изменить сложность вопросов, с помощью прохождения диагностического теста – этот вариант предлагается исключительно при желании повысить сложность. В случае успешного прохождения теста уровень знаний пользователя повышается, и он получает соответствующее достижение. Возможен и альтернативный вариант смены сложности без тестирования, однако за это достижение не присуждается.

Система также предоставляет рекомендации на основе результатов всех тестов. Анализируется сложность и успешность прохождения, изученные ранее темы, и на этой основе формируются персонализированные рекомендации. Система может предложить пользователю темы для повторения или освоения, а также задания или тесты, относящиеся к определённым темам, что способствует углублению знаний и повышению эффективности обучения. Рекомендации отображаются либо при введении соответствующих команд или сообщений, либо при нажатии соответствующих кнопок.

Примеры взаимодействия пользователя с системой представлены на рисунках 1 и 2.

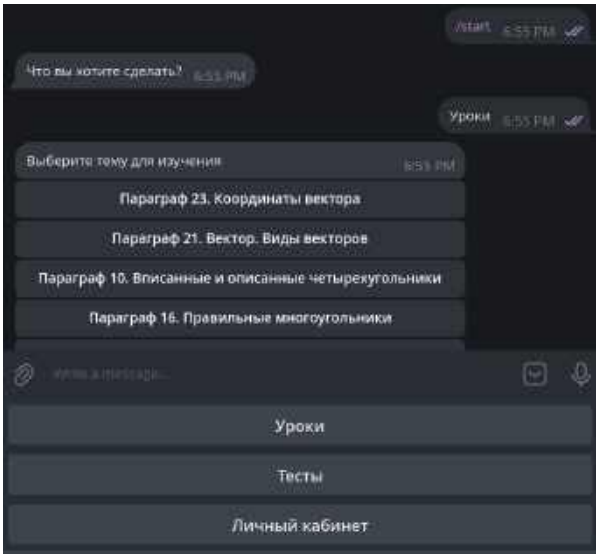


Рис. 2 – Пример диалога с системой.

III. Заключение

В рамках работы рассмотрена интеллектуальная система для персонального обучения в области геометрии, разработанная с целью повышения эффективности обучения за счет персонализации контента для пользователя и самооценки, т.е. механизма рефлексии. Система является полностью открытой, ее исходный код находится в свободном доступе и может быть дополнен. Пользовательская оболочка системы также находится в открытом доступе. В перспективе, планируется разработать подсистему визуализации геометрических чертежей, увеличить объем тем охватываемых системой и улучшить механизмы рекомендаций для наилучшего обучения [2][3].

IV. Список литературы

1. Голенков, В. В. Открытая технология онтологического проектирования, производства и эксплуатации семантически совместимых гибридных интеллектуальных компьютерных систем / В. В. Голенков, Н. А. Гулякина, Д. В. Шункевич. – Минск : Бестпринт, 2021. – 690 с.
2. Проект «Интеллектуальная обучающая система по геометрии» [Электронный ресурс] / «Интеллектуальная обучающая система по геометрии». – Режим доступа: `https://github.com/kretoffer/geometry.ostis`. – Дата доступа: 20.10.2025.
3. Проект «Пользовательский интерфейс для интеллектуальной обучающей системе по геометрии» [Электронный ресурс] / «Пользовательский интерфейс для интеллектуальной обучающей системе по геометрии». – Режим доступа: `https://github.com/kretoffer/tg-bot.geometry.ostis`. – Дата доступа: 20.10.2025.

РАЗРАБОТКА МОДУЛЯ ДЛЯ ПРЕДОБРАБОТКИ ДОКУМЕНТОВ ОБНАРУЖЕНИЯ ЗАИМСТВОВАНИЙ В ТЕКСТЕ

Крез К. С., Ефименко Д. Д., Войнилович Н. Ю.
Кафедра проектирования информационно-компьютерных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: k.krez@bsuir.by

*В статье рассматриваются основные этапы предварительной обработки *.pdf и *.docx-документов для обнаружения заимствований в тексте и принципы их работы. Выявлены ключевые проблемы предварительной обработки *.pdf и *.docx-документов. Предложены пути их решения в том числе использование специализированных библиотек для языка Python.*

Введение

В условиях роста объемов цифровой информации и широкого распространения автоматизированных систем анализа текстов особую значимость приобретает задача эффективной предварительной обработки данных. Одним из ключевых этапов данного процесса является очистка текста, направленная на удаление «шумовых» и «нерелевантных» элементов, препятствующих корректному проведению лингвистического и семантического анализа [1].

Особую сложность представляет обработка файлов в форматах *.pdf и *.docx, поскольку они включают не только текстовые данные, но и графические материалы, метаданные и элементы верстки. В *.pdf-документах текст, как правило, представлен в виде множества независимых блоков, разделенных линиями и координатными сетками, что делает затруднительным его прямое использование в алгоритме. В документах *.docx значительная часть информации содержится в технических метках и стилях, не имеющих смысловой нагрузки, но влияющих на корректность извлечения контента.

В связи с этим актуальной задачей является разработка модуля, обеспечивающих автоматическую очистку текстов от технических и структурных артефактов, с сохранением только релевантной смысловой информации.

I. Реализация процесса очистки

Процесс очистки текстовых данных в разработанной системе реализован в модуле text_cleaning.py [2], который представляет собой совокупность последовательно выполняемых процедур, направленных на удаление «шумовых, графических, технических и структурных» элементов документа [3]. Реализация данного модуля основана на принципах модульности, расширяемости и независимости от конкретного формата исходного файла. Это позволяет применять его как к документам формата *.pdf, так и к *.docx, обеспечивая единый подход к формированию очищенного текста. Этапы очистки текстового документа:

1. Удаление ссылок на графические элементы направлен на исключение всех упоминаний графических объектов – рисунков, схем, диаграмм и таблиц, – которые не несут содержательной информации.
   Для удаления подобных ссылок используются регулярные выражения, позволяющие находить и удалять подписи и текстовые отсылки к иллюстрациям. Наиболее часто применяемые паттерны:
   - r'(Рисунок|Таблица|рисунок|таблица) \s*\d+ ([–]?\d+) *\s*[–]?\s*.*' – удаление подписей к рисункам и таблицам;
   - r'\b(на рисунке|На рисунке|На рисунках)\b.*?[\n.]*' – удаление ссылок на графические элементы внутри текста.
2. Удаление технических элементов документа часто присутствуют в отчетах, учебных и нормативных документах. Они не несут смысловой информации, однако могут существенно влиять на корректность токенизации и разметки текста. Для устранения подобных артефактов используются следующие паттерн:
   - удаление номеров страниц: r'\n\s*\d+\s*\n';
   - удаление стандартных заголовков (например, шапок государственных документов): r'Министерство образования Республики Беларусь.*?Минск\s*\d{4}';
   - удаление служебных обозначений и маркировок: r'ПР\s+[А-Я](\s+[−]+)?'.
3. Удаление структурных элементов документа связанных с организационной структурой документа, таких как оглавления (содержание), списки использованных источников, приложения, ведомость и т.д. Эти части не несут содержательной значимости и лишь дублируют информацию, уже содержащуюся в структуре текста.

Удаление осуществляется по ключевым словам (например, «СОДЕРЖАНИЕ», «СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ», «ПРИЛОЖЕНИЕ», «ВЕДОМОСТЬ») и по регулярным выражениям, описывающим строки, содержащие только последовательности точек.

4. Нормализация пробелов и переносов строк, когда после удаления структурных и технических элементов выполняется нормализация текста. Цель данной этапа – обеспечение единообразного форматирования для последующих этапов анализа.
   Множественные пробелы, табуляции и повторяющиеся переносы строк заменяются на одиночные пробелы с помощью регулярного выражения r'\s+'. В результате формируется ровный, непрерывный текст без лишних разрывов и неформатированных фрагментов.
5. Обработка заголовков и структурирование текста направлены на обеспечение логической целостности документа посредством корректной идентификации разделов и глав. Функция is_chapter_heading() анализирует каждый текстовый блок на соответствие определённым паттернам, отражающим структуру документа:
   - r'^\d+\s+[А-Я\s]+$' – нумерованные заголовки, записанные прописными буквами;
   - r'^\d+\.\d+\s+[А-Я][а-я\s]+' – иерархические заголовки;
   - r'^[А-Я\s]+$' – ненумерованные заголовки, оформленные заглавными буквами.

   На основе найденных заголовков выполняется разделение текста по разделам, где каждая глава представлена заголовком и соответствующим ей очищенным содержанием.

В отличие от *.pdf и *.docx, где очистка выполняется после извлечения содержимого, для файлов формата *.docx очистка интегрирована непосредственно в процесс извлечения. Система извлекает только текстовые параграфы, игнорируя технические метки и элементы форматирования, не влияющие на смысловую структуру. Это способствует значительному снижению объёма постобработки и повышает качество итогового текста.

II. Практическая часть

Практическая часть исследования направлена на комплексную оценку производительности разработанного модуля текстовой очистки text_cleaning.py. Для обеспечения репрезентативности результатов была сформирована тестовая выборка, включающая: тип документа (курсовая и дипломная работа), формат документа (*.pdf и *.docx), кол-во страниц.

Заключение

Представленный модуль очистки *.pdf и *.docx документа использует минимальный набор этапов таких как: удаление ссылок на графические элементы, удаление технических элементов документа, удаление структурных элементов документа, нормализация пробелов и переносов строк, обработка заголовков и структурирование текста. В ходе работы была рассмотрена реализация данного процесса в рамках модуля text_clining.py.

Практическая часть показала, что: статистически значимые показатели производительности модуля со средним временем обработки 3,31 ± 0,02 мс/страницу; эмпирически подтверждена линейная зависимость временных характеристик обработки от объема документа.

Таким образом, предложенное решение представляет законченный модуль предобработки текстовых данных, обладающую теоретической обоснованностью и практической эффективностью для интеграции в программное средство для проверки академической добросовестности.

1. Разработка программного модуля распознавания документов на основе машинного обучения [Электронный ресурс]. – Режим доступа: https://apni.ru/article/11072-razrabotka-programmnogo-modulya-raspoznavaniya-dokumentov-na-osnove-mashinnogo-obucheniya. – Дата доступа: 14.09.2023.
2. Textclean: Text cleaning tools [Electronic resource]. – Mode of access: https://github.com/trinker/textclean. – Date of access: 14.09.2023.
3. Хобсон Лейн, Ханнес Хапке, Коул Ховард. Обработка естественного языка в действии / Л. Хобсон, Х. Хапке, К. Ховард. – СПб.: Питер, 2020. – С. 68–140.

Таблица 1 – Результаты очистки текста в документах

Тип документа	Кол-во стр.	Измерение, мс.						Сред. знач., мс.
		1	2	3	4	5	6	
Курс. работа 1 (*.pdf)	67	374,94	374,83	374,93	374,83	374,76	374,85	374,87
Курс. работа 2 (*.docx)	67	379,23	379,12	379,24	379,13	379,21	379,23	379,19
Диплом. работа 1 (*.pdf)	89	387,34	387,31	387,29	387,34	387,31	387,31	387,31
Диплом. работа 2 (*.docx)	89	387,59	387,61	387,58	387,59	387,59	387,61	387,60

ОБЗОР И КЛАССИФИКАЦИЯ УГРОЗ ИНФОРМАЦИОННЫМ СИСТЕМАМ СО СТОРОНЫ КВАНТОВЫХ АЛГОРИТМОВ

Крупенич Е. Г., Вергель И. В.
Кафедра программного обеспечения информационных технологий, Факультет компьютерных систем и сетей,
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {s.nesterenkov, e.krupenich, i.vergel}@bsuir.by

В статье рассматриваются квантовые алгоритмы, представляющие угрозу современным криптографическим системам. Среди них особое место занимают алгоритм Шора, позволяющий эффективно решать задачу факторизации целых чисел и вычисления дискретного логарифма, и алгоритм Гровера, обеспечивающий квадратичное ускорение поиска в неструктурированных массивах данных. Также предложен один из возможных вариантов классификации квантовых угроз по типу криптографического примитива, характеру воздействия, горизонту актуализации и степени критичности для инфраструктуры.

Введение

Квантовый алгоритм представляет собой классический алгоритм, который задает последовательность унитарных операций с указанием, над какими именно кубитами их надо совершать. Результат работы квантового алгоритма носит вероятностный характер. За счёт небольшого увеличения количества операций в алгоритме можно сколь угодно приблизить вероятность получения правильного результата к единице. Множества задач, допускающих решение на квантовом компьютере и на классическом, совпадают. Квантовый компьютер, таким образом, не увеличивает число алгоритмически разрешимых задач. Весь смысл применения квантового компьютера в том, что некоторые задачи он способен решить существенно быстрее, чем любой из классических.

I. Квантовые угрозы

ддВ последние годы появились конкретные алгоритмы, представляющие угрозу существующим средствам защиты информации. Наибольшую известность получили алгоритм Шора, подрывающий безопасность асимметричных криптосистем, и алгоритм Гровера, снижающий стойкость симметричных шифров и хэш-функций. Кроме того, в практику входят атаки типа «собери сейчас – расшифруй потом», что делает проблему ещё более актуальной. Подобные угрозы приобретают стратегический характер, поскольку затрагивают критическую инфраструктуру, системы электронных платежей, защищённые каналы связи и цифровые подписи.

II. Алгоритм Шора

Алгоритм Шора – это алгоритм разложения числа на простые множители, позволяющий разложить число M за $O(log^3 N)$, используя $O(\log N)$ логических кубитов. Его применение представляет угрозу для криптографических систем с открытым ключом, безопасность которых основана на сложности задач факторизации и вычисления дискретного логарифма. К таким системам относятся широко используемые алгоритмы RSA, Diffie–Hellman и криптография на эллиптических кривых (ECC). Именно они лежат в основе протоколов TLS/SSL, VPN, цифровых подписей и инфраструктуры открытых ключей (PKI). При этом другие классы асимметричных алгоритмов, использующие иные математические основы (например, решётки или кодовые конструкции), не подвержены воздействию алгоритма Шора.

Алгоритм Шора сводит задачу факторизации числа к задаче нахождения периода (порядка) функции.

1. Выбираем случайное число a, где $1 < a < N$.
2. Вычисляем $R = gcd(a,N)$.
   - Если K, то это нетривиальный делитель числа $N \neq 1$, а второй равен N/K. Алгоритм завершен.
3. Иначе используем квантовую подпрограмму, чтобы найти порядок r числа a.

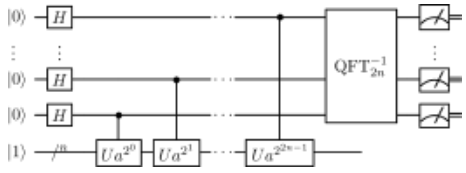


Рис. 1 – Квантовая подпрограмма в алгоритме Шора

4. Если r нечетное, возвращаемся к шагу 1.
5. Вычисляем $g = gcd(N,a^{r/2}+1)$
   - Если g нетривиален, то другой делитель равен N/g. Алгоритм завершен.
   - Иначе возвращаемся к шагу 1.

Алгоритм с высокой вероятностью успешно разложит N на множители после первого вызова квантовой подпрограммы, хотя теоретически может потребоваться несколько повторов.

III. Алгоритм Гровера

Алгоритм Гровера – это квантовый алгоритм поиска, позволяющий найти искомый элемент в неструктурированной базе данных из N элементов за $O(\sqrt{N})$ шагов вместо $O(N)$, требуемых классическим перебором. Для его реализации необходимо $O(\log N)$ кубитов. С его помощью становится возможным ускоренный перебор ключей и поиск предобразов для хэш-функций, что снижает криптостойкость симметричных алгоритмов шифрования и функций хэширования. В первую очередь под угрозой оказываются такие алгоритмы, как AES и SHA-2/SHA-3, что может привести к компрометации систем аутентификации и целостности данных.

Алгоритм Гровера основан на многократном применении так называемого оператора инверсии амплитуды, который усиливает вероятность правильного ответа:

1. Инициализируем систему до равномерной суперпозиции по всем состояниям.

$$|s\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle.$$

2. Применяем оператор U_ω.
3. Измерь полученное квантовое состояние в вычислительном базисе.

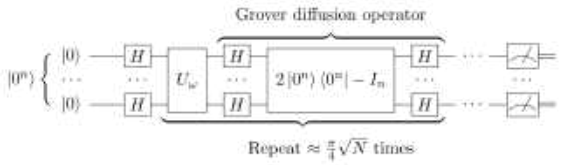


Рис. 2 – Представление алгоритма Гровера в виде квантовой схемы

После порядка $\frac{\pi}{4}\sqrt{N}$ итераций вероятность нахождения искомого элемента становится близкой к единице.

IV. Классификация угроз

Для систематизации квантовых угроз информационной безопасности можно выделить несколько ключевых признаков. Один из них связан с типом криптографического примитива: асимметричные алгоритмы, основанные на задачах факторизации и вычисления дискретного логарифма, оказываются уязвимыми перед алгоритмом Шора, тогда как симметричные шифры и хэш-функции подвержены ускоренному перебору с помощью алгоритма Гровера.

Другим важным признаком является характер воздействия: если Шор полностью разрушает математическую указанных алгоритмов, то Гровер лишь снижает их эффективную стойкость, уменьшая длину ключа или сложность поиска предобраза.

Важен и временной горизонт угроз: атаки на асимметричные системы строго доказаны и станут критичными с появлением масштабных квантовых компьютеров, тогда как атаки на симметричные примитивы возможны уже при меньших ресурсах.

Наиболее критичны для инфраструктуры системы обмена ключами, подписей и доверенных каналов, тогда как симметричные шифры и хэш-функции сохраняют стойкость при увеличении параметров.

Сводное распределение существующих криптографических систем по выделенным признакам представлено в таблице 1.

1. M. Ekerå. On completely factoring any integer efficiently in a single run of an order-finding algorithm. / M. Ekerå // Quantum Information Processing. – 2021, – Vol. 205, № 20, P. 1-14.
2. Bernstein D. J. Cost analysis of hash collisions: Will quantum computers make SHARCS obsolete? / D. J. Bernstein // Conference Proceedings for Special-purpose Hardware for Attacking Cryptographic Systems. – 2021. – №9. – P. 105-117.
3. Sinitsyn N. Topologically protected Grover's oracle for the partition problem. / N. Sinitsyn, B. Yan // Physical Review A. – 2023. – Vol. 108.
4. Bernstein D. J. Post-quantum RSA. / D. J. Bernstein, N. Heninger, P. Lou; L. Valenta // Luke Post-Quantum Cryptography. Lecture Notes in Computer Science. – 2017. – Vol. 10346. – P. 311-329.
5. Exman I., Pérez-Castillo R., Piattini M., Felderer M. Quantum Software: Aspects of Theory and System Design. / I. Exman, R. Pérez-Castillo, M. Piattini, M. Felderer // Springer Nature. — 2024. ISBN 978-3-031-64136-7.

Таблица 1 – Классификация угроз информационной безопасности со стороны квантовых алгоритмов

Класс системы	Примеры	Алгоритм	Эффект
Асимметричные	RSA, ECC, Diffie–Hellman	Шор	Полная уязвимость
Симметричные	AES-128/192/256	Гровер	Сокращение эффективной длины ключа
Хеш-функции	SHA-2, SHA-3	Гровер	Быстрый поиск коллизий
Инфраструктурные протоколы	TLS/SSL, VPN, PKI	Шор + Гровер	Ослабление защиты
Квантовая криптография	QKD (BB84, E91)	Шор/Гровер	Устойчивы
Асимметричные (постквантовые)	NTRU, Kyber, McEliece	Шор/Гровер	Устойчивы

АНАЛИЗ ПОДХОДОВ К ПОСТРОЕНИЮ КРИПТОГРАФИЧЕСКИХ СХЕМ, УСТОЙЧИВЫХ К КВАНТОВЫМ АТАКАМ

Крупенич Е. Г., Лукьянов А. А.
Кафедра программного обеспечения информационных технологий, Факультет компьютерных систем и сетей,
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {e.krupenich, a.lukianov}@bsuir.by

Статья даёт обзор современных подходов к построению постквантовых криптографических схем: решётчатых, кодовых, многочленных, хеш-основанных конструкций и изогений. Рассмотрены основные математические предпосылки, оценены устойчивость к классическим и квантовым атакам, сопоставлены производительность и практические ограничения. Приведены рекомендации по применимости в различных классах систем и указаны ключевые источники для углублённого изучения.

Введение

Появление масштабируемых квантовых вычислений ставит под сомнение безопасность классических асимметричных схем на основе факторизации и дискретного логарифма. В ответ разработаны несколько семей постквантовых примитивов, различающихся по математической предпосылке, эффективности и области применения. Цель статьи – систематизировать эти подходы, сопоставить их по ключевым критериям (стойкость, размеры, производительность, масштабируемость) и сформулировать практические рекомендации для поэтапной миграции протоколов и систем.

I. Обзор моделей

Решётчатые (Lattice-based). Опираются на SVP/CVP и LWE/Ring-LWE; дают широкий набор примитивов: KEM, подписи, шифрование, гомоморфия. Сильная теоретическая база и гибкость параметризации; основные ограничения – чувствительность к выбору параметров и увеличенные размеры ключей/шифротекстов в ряде реализаций.

Кодовые криптосхемы(Code-based). Основаны на сложности декодирования линейных кодов (классика – McEliece). Отличаются высокой скоростью шифрования/дешифрования и устойчивой практикой, но страдают очень большими открытыми ключами и менее гибкой параметризацией.

Многочленные и структурированные алгебраические конструкции (Ring/Module-LWE, NTRU-класс). Используют кольцевую структуру для уменьшения размеров и ускорения арифметики; позволяют компактные реализации для встроенных устройств. Риск – появление структурных слабостей при небрежной параметризации.

Хеш-основанные подписи. Lamport, Merkle, XMSS, LMS, SPHINCS+ полагаются на стойкость хеш-функций; дают простую и формально выверяемую модель безопасности. Минусы – управление состоянием (stateful решения) или увеличенная длина подписей (stateless схемы).

Isogeny-schemes. Схемы на основе изогений эллиптических кривых над сложными полями (пример SIKE) обеспечивают крайне компактные ключи, но операции относительно медленны, доказательная база и практическая зрелость ограничены; подходят в сценариях с критическим ограничением размера ключа.

II. Сравнение моделей

Для практического сравнения постквантовых схем использованы следующие критерии: (i) математические предпосылки и наличие доказательной базы; (ii) размеры ключей и подписи; (iii) производительность и затраты ресурсов; (iv) масштабируемость и архитектурные ограничения. Эти критерии позволяют соотнести безопасность с ресурсными требованиями и выбрать подходящие классы схем для конкретных приложений. Сводная оценка по указанным критериям приведена в Таблице 1.

III. Применимость в реальных системах

- Сетевые протоколы (TLS, VPN): предпочтение – решётчатые KEM и NTRU-подобные; внедрять гибридно для совместимости.
- Встроенные/IoT-устройства: приоритет – кольцевые/многочленные реализации (Module-LWE, NTRU) с оптимизацией памяти и CPU.
- Долговременная архивация и юридические подписи: предпочтение – хеш-основанные схемы; обеспечить процедуру управления состоянием.
- Серверная инфраструктура и HSM: кодовые схемы и оптимизированные решётчатые реализации; предусмотреть хранение и распределение крупных ключей.
- Миграция и эксплуатация: поэтапная интеграция, гибридные ключи, тестирование реализаций и мониторинг стандартов (NIST и др.).

IV. Вывод

Сравнительный анализ показывает, что универсального «лучшего» подхода не существует. Выбор зависит от требований: если важна гибкость и широкий набор примитивов – решётчатые схемы; при необходимости очень быстрых операций шифрования – кодовые; для долгосрочной гарантийной подписи – хеш-основанные; для компактных ключей при допускаемой медлительности – isogeny-методы. При практической миграции важно проводить системную оценку: анализ угроз, оценку ресурсов, тестирование реализации и план постепенного развёртывания. На переходном этапе стоит использовать гибридные схемы (классические + постквантовые) и опираться на параметры, рекомендованные профильными стандартами и профессиональными сообществами.

V. Список литературы

1. Москвин, В. С., Богатырёв, В. А. Постквантовые алгоритмы электронной цифровой подписи и их применение в распределённом реестре / В. С. Москвин, В. А. Богатырёв // H&ES Research. – 2022. – Vol. 14, № 4. – P. 47–53.
2. Лежинский, М. В., Мокряков, А. В. Алгоритмы шифрования, устойчивые к взлому в условиях квантового превосходства / М. В. Лежинский, А. В. Мокряков // Сборник научных трудов кафедры прикладной математики и программирования по результатам работы постоянного семинара «Теория систем». – 2021. – С. 141–147.
3. Шемякина, М. А. Анализ применения квантовых технологий в криптографии / М. А. Шемякина // International Journal of the Humanities and Natural Sciences. – 2019. – № 5-4. – С. 59–62.
4. National Institute of Standards and Technology. Status Report on the First Round of the NIST Post-Quantum Cryptography Standardization Process / Alagic J. et al. // NISTIR – 8240. – Gaithersburg, MD: NIST, 2019. – 27 pp.
5. Bos, J., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Presern, T., Schwabe, P., Seiler, G., Stehlé, D. CRYSTALS-Kyber: a CCA-secure module-LWE based KEM / J. Bos et al. // Proc. PQCrypto 2020. – 2020. – P. 1–23.
6. Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Neves, S., Seiler, G., Stehlé, D. CRYSTALS-Dilithium: a lattice-based digital signature scheme / L. Ducas, E. Kiltz, T. Lepoint etal.// Proc. Eurocrypt/PQCrypto 2020 – 2020. – P. 24–45.
7. Bernstein, D. J., Hülsing, A., Käsper, E., Lange, T., Schwabe, P. SPHINCS+: stateless hash-based signatures / D.J.Bernstein, A. Hülsing, E. Käsper etal.// Proc. PQCrypto 2020 / IACR ePrint Archive – 2020. – P. 1–31.

Таблица 1 – Сравнение моделей

Класс	Предпосылка безопасности	Размеры ключей /подписей	Производительность и ресурсы	Масштабируемость и ограничения
Решётчатые	LWE/Ring-LWE, SVP/CVP; богатая доказательная база	открытые ключи: 1–5 КБ; подписи: 1–3 КБ	хорошая при оптимизациях; подходит для аппаратного ускорения	требует памяти/пропускной способности для больших параметров; легко масштабируются на серверы
Кодовые	декодирование линейных кодов; устойчивая практика	открытые ключи: 100 КБ–1 МБ; подписи/-шифротексты: десятки–несколько сотен байт	очень быстрые шифрование/-дешифрование; генерация ключей дороже	хранение и передача больших ключей; предпочтительны серверные инфраструктуры
Кольца /многочлены	Ring/Module-LWE, NTRU-класс; риск структурных атак	открытые ключи: 1–2 КБ; подписи: 1–2 КБ	высокая эффективность при полиномиальной арифметике; низкая латентность	хороши для встраиваемых/IoT при оптимизации; требуют аккуратной параметризации
Хеш-основанные	безопасность сводится к стойкости хеш-функций; простая модель анализа	ключи: 32–64 байта; подписи: 10–40 КБ, stateless ближе к верхней границе	производительность зависит от конструкции; stateful обычно дешевле	подходят для архивирования и контролируемых сред; управление состоянием критично
Isogeny	сложность вычисления изогений; ограниченная доказательная база	открытые ключи: 200–400 байт; подписи/обмены: 100–300 байт	относительно медленные операции; высокие вычислительные накладные расходы	ограниченная практическая зрелость; пригодны при жёстких ограничениях по размеру ключей

ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ГЕНЕРАЦИИ НАРРАТИВОВ В ИНТЕРАКТИВНЫХ 2D-ПРИЛОЖЕНИЯХ

Курочка К. С., Овчинина О. В.
Кафедра информационные технологии,
Гомельский государственный технически университет имени П. О. Сухого
Гомель, Республика Беларусь
E-mail: kurochka@gstu.by, chess.rs56@gmail.com

В данной работе рассматриваются алгоритмы машинного обучения, в частности трансформерные модели, применяемые для автоматизированной генерации адаптивных сюжетов в интерактивных 2D-играх. Исследуется использование механизмов самовнимания и методов промпт-инжиниринга для построения динамических повествований, реагирующих на действия пользователя. Особое внимание уделяется архитектурам нейросетевых моделей и способам их интеграции в программные средства разработки интерактивного контента.

Введение

Современные интерактивные приложения характеризуются растущим вниманием к сюжетам, способным адаптироваться под решения игрока. Создание таких сценариев требует значительных трудозатрат со стороны сценаристов [1]. В связи с этим актуальной становится задача автоматизации генерации нарратива с использованием алгоритмов машинного обучения и искусственного интеллекта.

Целью данной работы является исследование возможностей применения нейросетевых моделей для синтеза сюжетов и динамически изменяющихся сюжетных элементов в интерактивных 2D-приложениях.

Актуальность темы обусловлена быстрым развитием генеративных моделей, таких как GPT, T5 и LLaMA, которые способны создавать связные текстовые последовательности на основе краткого описания. Применение таких систем в интерактивных продуктах открывает возможности для автоматического формирования сюжетных ветвлений, диалогов и описаний игровых событий. Это позволяет значительно сократить участие человека в процессе написания сценариев, обеспечивая при этом разнообразие и адаптивность игрового опыта.

В рамках исследования рассматриваются современные подходы к интеграции языковых моделей в программные среды, а также анализируются методы управления их генерацией с помощью параметров, таких как температура и длина вывода. Особое внимание уделяется вопросам согласованности текста, логической связности повествования и возможности масштабирования алгоритмов для использования в реальном времени.

I. Архитектура и процесс генерации нарратива

Генерация нарратива в интерактивных 2D-приложениях осуществляется с использованием алгоритмов машинного обучения. Одним из наиболее эффективных подходов являются трансформерные модели, такие как GPT, T5 и LLaMA, основанные на механизме самовнимания [2]. Механизм самовнимания позволяет модели учитывать взаимосвязи между всеми элементами текста одновременно, обеспечивая последовательность и логическую связанность событий. Модели обучаются на больших объемах данных, что позволяет формировать сцены, диалоги и ветвящиеся сюжетные линии, объединенные общей тематикой.

Для генерации нарратива алгоритм должен учитывать не только последовательность событий, но и интерактивность – влияние действий пользователя на ход сюжета. Это реализуется с помощью контекстных подсказок (prompts), которые могут включать: описание персонажей, цели и мотивации, конфликты, локации и ключевые события [3]. Использование структурированных prompts позволяет задавать моделью желаемый стиль повествования, жанр и глубину ветвлений сюжета. На основе этих подсказок формируется сценарий с несколькими вариантами развития событий.

Процесс генерации нарратива включает следующие этапы.

1. Обработка входного текста – выделение ключевых объектов, действий и событий;
2. построение событийной последовательности – формирование логики и последовательности развития сюжета;
3. формирование диалогов и взаимодействий между персонажами;
4. проверка согласованности событий – устранение логических ошибок, противоречий и повторов;
5. постобработка и корректировка – фильтрация некорректных фраз, согласование действий персонажей и настройка длины текстовых блоков.

Архитектура системы состоит из модулей обработки текста, генерации сюжета и проверки логики. Входные данные подаются в языковую модель, которая возвращает последовательность

событий и диалогов. Ключевые параметры модели, влияющие на результат, включают размер контекста, глубину внимания, количество слоев трансформера, длину генерируемого текста и температуру генерации, отвечающую за степень вариативности сценариев. При низких значениях температуры модель создает предсказуемый и логически устойчивый текст, а при высоких – более креативные, но менее стабильные повествования.

Дополнительно могут применяться методы дообучения моделей на специализированных корпусах – наборах игровых сценариев, диалогах или сюжетных шаблонах. Это позволяет адаптировать систему под конкретные жанры и стили повествования. Также возможно использование алгоритмов оценки качества текста, основанных на метриках BLEU, ROUGE и BERTScore, обеспечивая автоматическую проверку связности и оригинальности генерируемого контента.

II. Методы управления нарративом и перспективы развития

Для повышения управляемости процесса применяются алгоритмы итеративной генерации сюжета, при которых модель поэтапно развивает историю, сохраняя контекст предыдущих событий. Такой подход реализуется через механизм context window и позволяет системе поддерживать внутреннюю логику повествования на протяжении всей интерактивной сессии.

Перспективным направлением является использование иерархических трансформеров, где различные уровни модели отвечают за разные аспекты нарратива.

- Верхний уровень – за общую структуру истории (завязка, конфликт, развязка);
- средний – за сцены и взаимодействия;
- нижний – за диалоги и локальные детали.

Это повышает согласованность текста и дает возможность масштабировать систему под проекты различной сложности.

Дополнительно для управления сюжетом могут применяться следующие методы.

- Адаптивное ветвление – модель формирует несколько параллельных сюжетных линий, из которых в реальном времени выбирается наиболее релевантная в зависимости от действий пользователя;
- обратная связь и корректировка – автоматически оцениваются сгенерированные блоки текста по критериям логики, эмоциональной насыщенности и тематической целостности, и при необходимости происходит корректировка следующего шага генерации;
- управление стилем и жанром – в контекстных подсказках задаются параметры, определяющие жанр истории, эмоциональный тон и глубину ветвлений, позволяя поддерживать единый стиль повествования на протяжении всей интерактивной сессии.

В дальнейшем возможна интеграция генерации сюжета с системами симуляции поведения персонажей и анализом эмоций, что позволит создавать более естественные и контекстно зависимые истории, формирующиеся в ответ на действия пользователя.

Заключение

В тезисе рассмотрены современные алгоритмы машинного обучения для генерации нарратива в интерактивных 2D-приложениях. Основное внимание уделено трансформерным моделям, способным создавать связные сюжеты и диалоги на основе краткого текстового описания с учетом интерактивности пользователя. Описаны этапы процесса генерации, включая обработку текста, построение последовательности событий, формирование диалогов и проверку согласованности.

Выявлено, что использование контекстных подсказок и настройка ключевых параметров модели, таких как размер контекста и температура генерации, позволяет получать адаптивные и вариативные сценарии. Применение таких подходов сокращает трудозатраты на создание сюжетов и обеспечивает возможность автоматизированного формирования разветвленных нарративов в интерактивных приложениях.

Применение алгоритмов оценки качества текста (BLEU, ROUGE, BERTScore) и дообучение на специализированных корпусах позволяет повысить связность, оригинальность и стилистическую согласованность генерируемого нарратива.

В целом, использование современных подходов к генерации сюжетов позволяет существенно сократить трудозатраты на создание интерактивных сценариев, создавать разветвленные и адаптивные нарративы, повышающие вариативность и глубину интерактивного опыта в 2D-приложениях. Перспективным направлением является дальнейшая интеграция с системами анализа поведения пользователей и оценкой эмоционального отклика для формирования более естественных и контекстно зависимых историй.

1. Курочка, К. С. Нейросетевая модель автогенерации тестов для студентов в системе Moodle на основе анализа методических материалов / К. С. Курочка, Ю. С. Башаримов //Цифровая трансформация. – 2025. – Т. 31 – № 3 – С. 66–75.
2. Transformer models: an introduction and catalog [Электронный ресурс] – Режим доступа: `https://arxiv.org/abs/2302.07730`. – Дата доступа: 11.10.2025.
3. Hernández, J. A. ChatGPT Learning prompt engineering with 100+ examples / J. A. Hernández [etc.] // Madrid – 2024. – P. 161.

МЕТОДИКА ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ И ПОДГОТОВКИ ДАТАСЕТА ДЛЯ АНАЛИЗА ПРЕПАРАТОВ ЖИДКОСТНОЙ ЦИТОЛОГИИ С ИСПОЛЬЗОВАНИЕМ МАШИННОГО ОБУЧЕНИЯ

Курочка К. С., Зайцева Л. П., Гуменников Е. Д.
Кафедра информационных технологий, Гомельский государственный университет имени П. О. Сухого;
Учреждение Гомельский областной клинический онкологический диспансер,
централизованная цитологическая лаборатория
Гомель, Республика Беларусь
E-mail: guma178@gstu.by

В данной работе рассматриваются методы и подходы к подготовке датасета для обучения моделей машинного обучения в задачах анализа препаратов жидкостной цитологии. Особое внимание уделяется процессам сбора, очистки, аннотирования и форматирования данных, необходимым для обеспечения высокой точности и надежности автоматизированных систем диагностики. Представлены рекомендации по оптимизации этапов предварительной обработки данных, а также обсуждаются особенности работы с медицинскими изображениями и цитологическими образцами.

Введение

Исследование клеточных структур методом автоматизированного анализа изображений цифровой микроскопии играет важную роль в биомедицинских исследованиях, особенно в области онкологии. Одним из ключевых элементов, участвующих в клеточном метаболизме и пролиферации, являются ядрышковые организаторы. Эти участки хромосом содержат гены, отвечающие за синтез рибосомальной РНК (рРНК), что делает их активность важным показателем уровня рибосомного синтеза и общей активности клетки. Анализ изменений в структуре и количестве ядрышковых организаторов может служить ценным диагностическим критерием при оценке злокачественных новообразований и прогнозировании их агрессивности.

В публикации рассматриваются методы подготовки датасета для обучения ИИ в анализе жидкостной цитологии, включая сбор, аннотирование и обработку данных, что способствует качественному обучению моделей искусственного интеллекта и повышает эффективность автоматизированных решений.

I. Обзор исходных данных, постановка задачи

На рисунке 1 приведен пример изображения препарата полученного методом цифровой микроскопии. Подобные изображения являются исходными данными для искомой модели машинного обучения. Перед такой компьютерной программой стоит задача локализации клеточных структур и их компонентов, а именно ядрышковых организаторов.

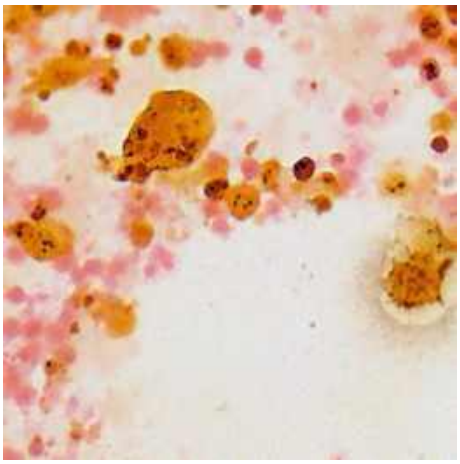


Рис. 1 – Цитологический препарат

На данном изображении клеточные структуры представлены эллиптическими областями повышенной контрастности. Однако формы и размеры этих фрагментов в значительной мере отличаются относительно друг друга. Так же присутствуют элементы окрашенные в аналогичные цвета и обладающие нечеткой границей, которые не являются искомыми клеточными структурами. Таким образом задача локализации таких объектов является трудно формализуемой. Похожее описание применимо к другому типу искомых объектов – ядрышковых организаторов. Лишь с тем отличием что их контрастность куда выше контрастности клеточных структур и их размер значительно меньше сравнительно с первым описанным классом. На изображении также встречаются фрагменты похожего цвета и контрастности не являющиеся искомыми объектами. Ядрышковые организаторы всегда содержатся внутри клеточных структур. Традиционные алгоритмы основаны на методах компьютерного зрения: пороговая сегментация, морфологические операции, фильтрация и анализ контуров. Эти

подходы позволяют выделить объекты по цвету, форме и текстуре. Однако они чувствительны к шуму, вариативности окрашивания и артефактам. При незначительных изменениях входных данных такие методы требуют ручной настройки параметров, что снижает их универсальность.

Более продвинутые методы используют машинное обучение, где объекты распознаются на основе заранее извлечённых признаков – например, гистограмм градиентов, локальных бинарных шаблонов или др. Классификаторы, такие как SVM или случайные леса, обучаются на размеченных данных. Хотя эти методы демонстрируют лучшую устойчивость, они не всегда справляются с сложными случаями перекрытия или деформации клеток.

Наиболее эффективным решением задачи распознавания клеток и ядрошковых организаторов являются современные архитектуры глубоких нейронных сетей, относящиеся к классу одноэтапных детекторов. Эти модели объединяют этапы локализации и классификации объектов в единую структуру, что позволяет обрабатывать изображения с высокой точностью. Они обучаются напрямую на изображениях, автоматически извлекая релевантные признаки, и демонстрируют устойчивость к шуму, вариациям формы и плотности объектов.

Благодаря своей архитектуре, одноэтапные детекторы особенно хорошо подходят для задач, где требуется быстрое и точное обнаружение множества мелких объектов — как в случае с клетками и ядрошковыми организаторами.

В задачах анализа цитологических изображений локализация объектов оказывается более практичным и эффективным подходом по сравнению с сегментацией. Метод локализации позволяет быстро и надёжно обнаруживать клетки и ядрошковые организаторы с помощью ограничивающих рамок, обеспечивая высокую точность при значительно меньших затратах на аннотирование и обучение модели. В отличие от сегментации локализация демонстрирует лучшую масштабируемость, устойчивость к вариативности изображений и высокую скорость обработки.

Для обучении модели искусственного интеллекта необходим датасет, содержащий изображения препаратов и аннатирующий файл, содержащий N строк. Кажадая строка содержит информацию о классе объекта и его габаритной рамки в виде «N X1 Y1 X2 Y2», где N код класса, X1 и Y1 координаты нижнего левого угла габарита объекта, а X2 и Y2 верхнего правого.

II. Конвертация существующего датасета предназначенного для задачи сегментации

Для обучения модели локализации клеточных структур и ядрышковых организаторов возможно использование датасета предназначенного для инстанс-сегментации аналогичных объектов. В существующем наборе данных каждое изображение аннотируется текстовым файлом содержащим количество строк равное количеству размеченных объектов искомых классов на соответствующем изображении. Каждая строка содержит код класса и пары вещественных чисел кодирующих координаты вершин полигонов ограничивающих сегмент изображения содержащий объект.

Путем простых математических манипуляций возможно рассчитать границы рамок для задачи локализации и сформировать датасет для решения описанной задачи. Расчет координат границ рамок возможно вычислить по следующим формулам:

$$x_{ul} = min\,(X)\,; y_{ul} = min\,(Y)$$

$$x_{br} = max\,(X)\,; y_{br} = max\,(Y)$$

где X – координаты X вершин полигона;

Y – координаты Y вершин полигона;

x_{ul} – координата x верхнего левого угла рамки;

y_{ul} – координата y верхнего левого угла рамки;

x_{br} – координата x нижнего правого угла рамки;

y_{br} – координата y нижнего правого угла рамки;

Таким образом конвертируется датасет ориентированный на задачу сегментации для задачи локализации.

III. Контекстно-ориентированная фильтрация артефактов изображений

Для повышения качества обучающего датасета при распознавании клеток и ядрошковых организаторов в цитологических препаратах разумно применение метод предварительной обработки изображений, заключающийся в размытии ложных или обрезанных объектов. Такие структуры, часто встречающиеся на краях изображений или в виде артефактов, могут негативно влиять на обучение модели, создавая шум и снижая точность распознавания. Размытие позволяет нейтрализовать их влияние, сохраняя при этом естественную текстуру фона, что делает изображения более приближенными к реальным условиям микроскопии. Этот подход способствует улучшению устойчивости модели и снижению количества ложных срабатываний.

1. Курочка К. С., Ковалев В. А. Организация распределенных вычислений при обработке цифровых биомедицинских изображений // Информатика. – 2018. – №. 4 (24). – С. 66-73.
2. Weeks, S. E. The nucleolus: a central response hub for the stressors that drive cancer progression / S. E. Weeks, B. J. Metge, R. S. Samant // Cellular and Molecular Life Sciences. – 2019. – Vol. 76. – С. 4511-4524.
3. Богуш, Р. П. Обнаружение объектов на изображениях с большим разрешением на основе их пирамидально-блочной обработки / Р. П. Богуш, И. Ю. Захарова, С. В. Абламейко // Информатика. – 2020. – Т. 17, № 2. – С. 7-16.

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КАЛИБРОВКИ МЭМС-АКСЕЛЕРОМЕТРОВ В МИКРОКОНТРОЛЛЕРНЫХ ИНЕРЦИАЛЬНЫХ СИСТЕМАХ

Кузнецов К. С.
Высшая школа компьютерных технологий и информационных систем
Санкт-Петербургский политехнический университет Петра Великого
Санкт-Петербург, Россия
E-mail: kuznetsov.ks@edu.sbpstu.ru

В работе рассматривается проблема калибровки МЭМС-акселерометров, применяемых в микроконтроллерных навигационных системах. Показано, что традиционные методы компенсации ошибок (устранение смещения нуля и погрешности масштаба) имеют ограниченную эффективность. Предложено использование методов машинного обучения, позволяющих адаптивно снижать систематические и случайные ошибки измерений. Для иллюстрации подхода приведено моделирование сигнала акселерометра с добавлением шума и смещения, а также его коррекция с применением линейной регрессии. Полученные результаты демонстрируют снижение среднеквадратической ошибки и подтверждают перспективность использования машинного обучения для повышения точности инерциальных навигационных систем.

Введение

Микроэлектромеханические датчики (МЭМС) занимают ключевое место в современных навигационных и управляющих системах благодаря низкой стоимости, компактности и возможности интеграции с микроконтроллерами [1]. Однако точность МЭМС-акселерометров ограничивается рядом факторов: смещением нуля, температурным дрейфом, шумами и перекрёстными погрешностями. Классические методы калибровки, такие как шестипозиционный тест и использование фильтра Калмана, позволяют частично компенсировать данные ошибки, но имеют ограничения при учёте нелинейных и нестабильных составляющих. В связи с этим актуальной задачей является применение методов машинного обучения (ML) для калибровки и коррекции выходных данных МЭМС-акселерометров в составе микроконтроллерных инерциальных навигационных систем (INS).

I. Ошибки МЭМС-акселерометров

К типовым источникам погрешностей акселерометров относят [2]:

- смещение нуля (bias), изменяющееся со временем и температурой;
- масштабный коэффициент (scale factor), зависящий от условий эксплуатации;
- перекрёстные чувствительности (misalignment);
- случайные шумы.

На рисунке 1 представлена упрощённая схема формирования ошибок МЭМС-акселерометра.

Рис. 1 – Схема формирования ошибок МЭМС-акселерометра

II. Методы калибровки

Классические подходы включают [3]:

- статическую калибровку в нескольких фиксированных положениях (шестипозиционный тест);
- динамическую калибровку с использованием эталонных траекторий;
- применение фильтра Калмана для совместной оценки параметров модели и состояния системы.

Ограничением указанных методов является их недостаточная эффективность при наличии нелинейных и изменяющихся ошибок.

Таблица 1 – Сравнение методов калибровки

Метод	Учет нелинейностей	Вычислительные ресурсы	Точность
Шестипозиционный тест	низкий	низкие	средняя
Фильтр Калмана	частичный	средние	выше
Машинное обучение	высокий	зависит от модели (TinyML – низкие)	высокая

III. Применение и моделирование методов машинного обучения

Методы машинного обучения позволяют моделировать сложные зависимости между «сырыми» данными акселерометра и истинными значениями ускорений. Перспективными являются следующие направления:

- использование регрессионных моделей и нейросетей для предсказания смещения и масштабных коэффициентов [4];
- гибридные подходы (ML-корректор + фильтр Калмана) [5];

– внедрение технологий TinyML, позволяющих выполнять коррекцию непосредственно на микроконтроллере с ограниченными ресурсами [6].

Общий алгоритм применения ML можно описать следующим образом:

1. Сбор данных с акселерометра в различных условиях (температура, ориентация).
2. Обучение ML-модели на эталонных данных.
3. Интеграция обученной модели в прошивку микроконтроллера.
4. Использование модели для коррекции измерений в реальном времени.

Сигнал идеального акселерометра можно описать выражением

$$a_{true}(t) = A \cdot \sin(\omega t). \tag{1}$$

С учётом смещения и шумов измеренный сигнал имеет вид

$$a_{measured}(t) = (1 + \epsilon) \cdot a_{true}(t) + b + n(t), \tag{2}$$

где ϵ – ошибка масштаба, b – смещение нуля, $n(t)$ – гауссовский шум.

Калибровка выполняется методом линейной регрессии:

$$a_{calibrated}(t) = \alpha \cdot a_{measured}(t) + \beta, \tag{3}$$

где коэффициенты α, β подбираются методом наименьших квадратов.

Качество восстановления оценивается по среднеквадратической ошибке:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} \left(a_{true}(t_i) - a_{signal}(t_i) \right)^2. \tag{4}$$

Используя соотношения (1)–(4), был проведён модельный эксперимент, в ходе которого сгенерирован сигнал идеального акселерометра $a_{true}(t)$ и искажённый сигнал $a_{measured}(t)$ с добавлением смещения и шумов. Для коррекции измерений применялся метод линейной регрессии (3), позволяющий подобрать коэффициенты α и β таким образом, чтобы минимизировать среднеквадратическую ошибку (4).

Результаты эксперимента показали, что применение регрессионной модели позволяет эффективно компенсировать систематические погрешности и уменьшить уровень шума. На рисунке 2 приведено сравнение исходного и откалиброванного сигналов, из которого видно, что восстановленный сигнал значительно ближе к истинным значениям ускорений.

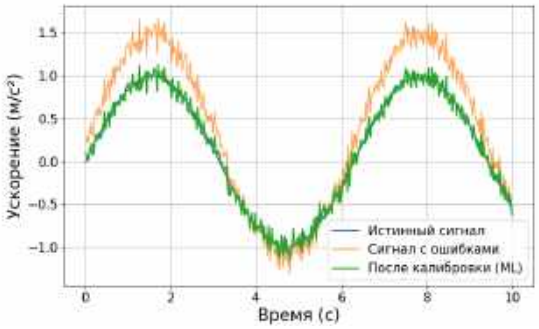


Рис. 2 – Сравнение измеренного и откалиброванного сигналов акселерометра

В модельном эксперименте использовались параметры: $N = 500$, $b = 0.2$ м/с^2, scale factor $\epsilon = 1.3$, $n(t) \sim \mathcal{N}(0, (0.1 \text{ м/с}^2)^2)$. После калибровки методом линейной регрессии ($\alpha = 0.7613$, $\beta = -0.1509$) среднеквадратическая ошибка снизилась более чем на 95%, а восстановленный сигнал практически совпал с истинным, что подтверждает эффективность ML-подхода.

IV. Заключение

Применение методов машинного обучения для калибровки МЭМС-акселерометров в составе микроконтроллерных INS является перспективным направлением, позволяющим повысить точность навигационных решений. В отличие от классических методов, ML-подходы способны учитывать нелинейные и изменяющиеся погрешности. Развитие технологий TinyML делает возможной реализацию подобных алгоритмов даже на маломощных микроконтроллерах. Перспективными задачами являются исследование оптимальных архитектур нейросетей, адаптивная калибровка в реальном времени и интеграция ML-методов в комбинированные INS.

V. Список литературы

1. Mammadov,A. Z. The model of inertial navigation system on base of MEMS sensors for unmanned aerial vehicles / A. Z. Mammadov // Кронос. – 2019. – № 10 (37).
2. Аксенов,В. В. Инструментальные средства для экспериментального построения модели погрешностей МЭМС акселерометра / В. В. Аксенов // ИВД. – 2020. – № 11 (71).
3. Белоус,А. И. МЭМС: конструкции, технологии, приложения / А. И. Белоус, С. А. Чижик. – М.: ТЕХНОСФЕРА, 2024. – 620 с. – ISBN 978-5-94836-688-3.
4. Оксенчук,Т. А. Алгоритмы машинного обучения для регрессионного анализа и прогнозирования числовых данных / Т. А. Оксенчук // Вестник науки. – 2025. – № 6 (87).
5. Тихонова, А. И. Применение методов машинного обучения для анализа вектора состояний РТК / А. И. Тихонова, Г. А. Мамедов, Д. А. Косневич, В. С. Радабольский, М. А. Гисмельбари // XXVII Международная конференция по мягким вычислениям и измерениям (SCM'2024), 22–24 мая 2024, Санкт-Петербург, СПбГЭТУ «ЛЭТИ».
6. TinyML – машинное обучение на микроконтроллерах // Хабр, 2024. – Режим доступа: `https://habr.com/ru/companies/first/articles/665932/`.

ИИ КАК ИНСТРУМЕНТ ОБОСНОВАНИЯ СТРАТЕГИЧЕСКИХ И ТАКТИЧЕСКИХ РЕШЕНИЙ В СИСТЕМЕ ЖКХ РЕСПУБЛИКИ БЕЛАРУСЬ

Лазаревич И. М., Мосин Н. А., Савицкий Р. М.
Кафедра экономики, экономической информатики
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: i.lazarevich@bsuir.by

В статье обоснована необходимость применения жилищно-коммунальным хозяйством Республики Беларусь ИИ при обработке заявок населения, составлен предположительый план реализации проекта, выделены возможные выгоды от его внедрения.

Введение

Оптимизация процессов оказания услуг жилищно-коммунальным хозяйством (ЖКХ) является первоочередной задачей, нацеленной на повышение качества обслуживания населения. На текущий момент у ЖКХ Республики Беларусь сформированы механизм получения информации о проблемах в обслуживании населения и информационная база заявок. С целью устранения повторяющихся проблем, возникших при обслуживании тех или иных объектов, возможно применение интеллектуальных информационных технологий, например искусственный интеллект (ИИ).

I. Основная часть

ИИ представляет собой комплекс инструментов, позволяющих решать задачи (восприятие, обучение, рассуждение, выявление проблем и принятие решений по их устранению), выполняемые человеком или группой лиц, и реализованных компьютерными системами.

В функции ЖКХ входит обработка обращений от населения, связанных с авариями и поломками, качеством обслуживания. Часто работники ЖКХ устраняют проблему временно, не выявляя основные причины и закономерности, что способствует возникновению различного рода проблем: повторение обращения граждан, сокращение эффективности работы организации, повышение тарифов населения и др. Следовательно традиционные методы обработки обращений и устранения повторяющихся проблем не эффективны и им требуется доработка.

При том, что ИИ имеет высокую цену разработки; представляет собой сложную систему, состоящую из огромного количества компонентов, которую нужно создавать и поддерживать; хранение самого машинного интеллекта и пакетов данных на серверах требуют огромного количества места, бесконечную отладку и модификацию, использовать нейронную сеть надо в тех областях, где процессы могут быть полностью оптимизированы [1]. ИИ позволит сократить время на рутинные задачи, мониторинг и устранение сбоев. Для этого целесообразно выявить закономерности между обращениями граждан и системой работы ЖКХ, сформировать модель ИИ с блоком анализа и устранения различного рода проблем системы. Поэтому для реализации проекта необходимо использовать [2]:

1. Машинное обучение – для анализа заявок;
2. Нейронные сети – для прогнозирования аварий;
3. Обработка естественного языка – для обработки и понимания текстов обращения граждан;
4. Аналитические системы – для проведения глубокого исследования с целью предложения новых методов оптимизации работы.

Но стоит понимать, чтобы интеграция произошла успешно и была эффективна, необходимо не только собрать статистические данные и построить алгоритмы, но и грамотно контролировать сам этап обучения, исключая возможность появления статистически несвязанных результатов, обеспечивающих низкую эффективность работ. Возможные этапы достижения цели представлены в таблице [3]:

Таблица 1 – Этапы реализации проекта

Этап	Название
1	Анализ текущего состояния цифровизации и автоматизации управления в ЖКХ
2	Определение основных групп обращения граждан
3	Разделить основные группы обращений на кластеры
4	Разработать концептуальную модель интеллектуальной системы обработки кластеров на основе машинного обучения и анализа тестовых данных
5	Сформировать основные подходы для аналитического выявления повторяющихся и критических проблем в системе ЖКХ
6	Разработать и протестировать основные алгоритмы построения стратегических решений
7	Провести профилактику неисправностей на основе прогнозных данных модели
8	Оценить экономическую эффективность внедрения модели в систему ЖКХ

Возможные выгоды от внедрения ИИ в систему ЖКХ:

1. Решение типовых проблем системы (оперативное реагирование на заявки);
2. Оптимизация маршрута движения дежурной бригады. Например, если поступило 5-7 одинаковых заявок из жилого комплекса о том, что нет горячей воды, дежурная группа будет автоматически направлена на общий узел подачи горячей воды. Данный подход позволит сократить время на выяснение причины поломки и уменьшить время простоя служб;
3. Контроль нагрузки на структурные подразделения ЖЭС. В случаях выявления неравномерной нагрузки работников (бригад) ИИ составит рекомендации по перенаправлению трудовых ресурсов. Данный подход позволит сделать систему гибкой и эффективной, повышая интенсивность труда и снижая время обработки заявок [4];
4. Составление плана по оптимизации расходов. Например, если в пределах жилищно ремонтно-эксплуатационных объединений (ЖРЭО) или жилищно-эксплуатационных служб (ЖЭС) происходят регулярные поломки оборудования, то ИИ предоставит план ремонта, согласно которому работа специалистов не прекратиться. Данный подход позволит поэтапно заменять оборудование на новое, не прекращая обслуживание населения;
5. Подготовка возможного плана развития ЖКХ в целом и его структурных подразделений (ЖРЭО, ЖЭС) и др.

Не смотря на преимущества интеграции ИИ в систему ЖКХ, его развертывание сопровождается рядом сложностей:

1. Необходим большой объем статистических данных. Для успешного обучения модели необходимо подготовить набор данных, включающий в себя не только идеальные данные, но и ошибочные. Такой подход поможет сделать анализ более успешным [5].
2. Нужны постоянные инвестиции в разработку и развитие ИИ, а также нужно успешно интегрировать его в существующие системы, при этом не нарушив работу.
3. Необходима переподготовка части персонала. По сколько ИИ является новой технологией, части персонала будет трудно адаптироваться к работе [6]. Поэтому необходимо будет ввести обучающие курсы, следовательно это может снизить эффективность работы ЖКХ, так как часть персонала будет на переобучении.

Необходимо утвердить стандарты сбора, очистки, а также настроить процедуры валидации и обмена данными между подразделениями и механизмы, скорытия персональных данных. Эти инструменты помимо точности прогнозов снижают риск санкционированного использования сервисов систем, которые не соответствуют требованиям законодательства.При грамотном подходе, структурированном анализе и последовательному выполнению работы данные сложности будут устранены.

II. Заключение

Внедрение в работу системы ЖКХ ИИ позволит повысить качество и скорость обработки обращений граждан, формировать стратегический и тактический планы профилактики и устранения цикличных неисправностей и проблем системы, прогнозировать объем финансовых расходов на устранение системных проблем сферы ЖКХ. ИИ позволит снизить нагрузку на систему ЖКХ, способствуя повышению эффективности принятий решений и улучшения качества обслуживания населения, а также начнёт этап формирования «умного города», где большинство процессов будет автоматизировано и работать на основе ИИ.

III. Список литературы

1. Мосин Н. А. Влияние искусственного интеллекта на экономические процессы / Н.А. Мосин, Н.А. Ходаренок // актуальные вопросы экономики и информационных технологий: м-лы 60-й юбил. науч. конференции аспирантов, магистрантов и студентов УО «Белорусский государственный университет информатики и радиоэлектроники», 22 апреля 2024 г., Минск, Беларусь. – 2024. – С. 110-113.
2. Sergey Ablameyko Использование систем искусственного интеллекта при обеспечении общественной безопасности в «умном городе»: юридические аспекты / Nadzeya V. Shakel, Rykhard Bohush // Vestnik of Polotsk State University Part D Economic and legal sciences, Полоцк, Беларусь. – 2021.
3. John J. Artificial Intelligence for Smart Cities: A Comprehensive Review / J. John // MDPI. – 2025.
4. Zhu D. City AI: a strategic framework for urban artificial intelligence / D. Zhu // Springer. – 2025
5. Artificial Intelligence in Smart Cities – Applications, Barriers and Future Directions: A Review / (A comprehensive review paper) // ResearchGate. – 2024
6. OECD. AI for advancing smart cities – Issues note (desk review + 70 case studies) / OECD. – 2025
7. Sanchez-Gallegos D. D. Smart Cities: A Systematic Review of Emerging Technologies / D.D. Sanchez-Gallegos // MDPI. – 2025.
8. Шиюй Т. Использование цифровых технологий в развитии умных городов: опыт Республики Беларусь / Т. Шиюй // CyberLeninka. – 2023
9. Синь В. Искусственный интеллект как фактор развития «умного города» / В. Синь // Elib.BSU. – 2024

РЕАЛИЗАЦИЯ РЕКОМЕНДАТЕЛЬНОЙ СИСТЕМЫ В МАРКЕТПЛЕЙСЕ НА ОСНОВЕ ЭМБЕДДИНГОВ

Нестеренков С. Н., Лебедевич А. В.
Факультет компьютерных систем и сетей,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: nesterenkov@bsuir.by, artyomlebedevich@gmail.com

В работе предлагается подход к построению гибридной рекомендательной системы для электронной коммерции на основе векторных представлений. Модель использует компактные 32-мерные эмбеддинги, объединяющие текстовую информацию, атрибуты товаров и данные о поведении пользователей. Ключевая особенность решения – интеграция контентных векторных сигналов с традиционными методами коллаборативной фильтрации через механизм гибридного переранжирования. Экспериментальная оценка на реальных данных маркетплейса показала увеличение CTR на 6,8%, покрытия каталога на 3% и эффективное решение проблемы холодного старта для новых товаров.

Введение

Современные маркетплейсы сталкиваются с задачей предоставления персонализированных рекомендаций из каталогов, содержащих десятки и сотни тысяч товарных позиций. Классические методы коллаборативной фильтрации демонстрируют высокую эффективность для популярных товаров, но испытывают серьёзные трудности с новыми позициями (проблема холодного старта) и редкими товарами длинного хвоста [1, 2].

Векторные представления (эмбеддинги) открывают возможность объединения коллаборативных и контентных подходов: кодирование товаров в едином семантическом пространстве позволяет измерять их близость независимо от истории взаимодействий [3]. Дополнительным преимуществом является возможность применения единой модели не только в рекомендациях, но и в функции поиска [4].

I. Математическая модель векторных представлений

Пусть $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$ – множество товаров в каталоге, $\mathcal{U} = \{u_1, u_2, \ldots, u_M\}$ – множество пользователей. Целью является построение отображения $f: \mathcal{P} \to \mathbb{R}^d$, которое кодирует товар p_i в вектор $\mathbf{e}_i \in \mathbb{R}^d$ фиксированной размерности $d = 32$.

Формирование товарных эмбеддингов

Представление товара $\mathbf{e}_i$ формируется как взвешенная комбинация признаков:

$$\mathbf{e}_i = W_t \mathbf{t}_i + W_a \mathbf{a}_i + W_b \mathbf{b}_i \tag{1}$$

где $\mathbf{t}_i \in \mathbb{R}^{d_t}$ – текстовые признаки (название, описание, бренд), $\mathbf{a}_i \in \mathbb{R}^{d_a}$ – атрибутивные признаки (категория, цена, характеристики), $\mathbf{b}_i \in \mathbb{R}^{d_b}$ – поведенческие признаки (просмотры, покупки), W_t, W_a, W_b – обучаемые матрицы проекции.

Текстовые признаки извлекаются через предобученный трансформер-энкодер для русского языка [7]:

$$\mathbf{t}_i = \text{Encoder}(\text{concat}(name_i, desc_i, brand_i))$$

Атрибутивные признаки нормализуются:

$$\mathbf{a}_i = [\text{embed}(cat_i), \frac{price_i - \mu_p}{\sigma_p}, \mathbf{attr}_i]$$

где $\text{embed}(cat_i)$ – эмбеддинг категории, нормализация цены выполняется по среднему μ_p и стандартному отклонению σ_p в категории.

Поведенческие признаки кодируют популярность:

$$\mathbf{b}_i = [\log(1 + v_i^{7d}), \log(1 + v_i^{30d}), \log(1 + c_i^{90d})]$$

где v_i^{7d}, v_i^{30d} – число просмотров за 7 и 30 дней, c_i^{90d} – число покупок за 90 дней. Логарифмическое преобразование сглаживает влияние экстремально популярных товаров.

Выбор размерности $d = 32$ обусловлен балансом между выразительной способностью и вычислительной эффективностью.

Формирование пользовательских эмбеддингов

Представление пользователя строится через взвешенное агрегирование:

$$\mathbf{u}_j = \frac{\sum_{i \in \mathcal{I}_j} w_{ij} \mathbf{e}_i}{\sum_{i \in \mathcal{I}_j} w_{ij}} \tag{2}$$

где $\mathcal{I}_j$ – множество товаров в истории пользователя. Веса учитывают тип взаимодействия и давность:

$$w_{ij} = \alpha_{type(ij)} \cdot \exp(-\lambda \Delta t_{ij})$$

где α_{type} – базовый вес (покупка: 3.0, корзина: 2.0, просмотр: 1.0), $\lambda = 0.05$ – коэффициент затухания, Δt_{ij} – время с момента взаимодействия в днях.

Метрика близости

Для измерения релевантности используется косинусное сходство:

$$s_{\text{embed}}(i,j) = \frac{\mathbf{e}_i \cdot \mathbf{u}_j}{\|\mathbf{e}_i\|\|\mathbf{u}_j\|} \tag{3}$$

Для эффективного поиска применяется индекс FAISS HNSW [5], который обеспечивает приближённый поиск за время $O(\log N)$ вместо линейного перебора.

II. Гибридная функция скоринга

Финальный скор объединяет коллаборативные и контентные сигналы:

$$S(i,j) = \alpha_i s_{\text{collab}}(i,j) + (1-\alpha_i) s_{\text{embed}}(i,j) \tag{4}$$

Коллаборативный скор вычисляется через алгоритм Swing [6]:

$$s_{\text{collab}}(i,k) = \sum_{u,v \in \mathcal{U}} \frac{w_u \cdot w_v}{\alpha_{swing} + |\mathcal{I}_u \cap \mathcal{I}_v| - 1}$$

где суммирование ведётся по парам пользователей, купивших товары i и k; w_u, w_v – веса пользователей; $\mathcal{I}_u$, $\mathcal{I}_v$ – множества купленных товаров; $\alpha_{swing} = 5$ – параметр сглаживания.

Параметр баланса α_i адаптируется индивидуально:

$$\alpha_i = \begin{cases} 0.2, & \text{если } n_i < 5 \\ 0.5, & \text{если } 5 \le n_i < 20 \\ 0.75, & \text{если } n_i \ge 20 \end{cases}$$

где n_i – число взаимодействий с товаром за 90 дней. Схема обеспечивает переход от контентного подхода для новых товаров к коллаборативному для зрелых позиций.

III. Обучение векторной модели

Модель обучается методом контрастивного обучения на исторических данных взаимодействий. Положительные примеры извлекаются из логов: если пользователь купил или кликнул товар, формируется положительная пара (u_j, p_i).

Отрицательные примеры формируются тремя стратегиями: случайные негативы из всего каталога (60%), жёсткие негативы из той же категории (30%), внутрипакетные негативы – положительные примеры других пользователей (10%). Соотношение положительных и отрицательных примеров составляет 1:5.

Текстовый энкодер основан на предобученной модели BERT для русского языка с 6 слоями трансформера. Проекционные слои реализованы как двухслойные MLP с активацией GELU и dropout 0.1.

Оптимизация выполняется алгоритмом AdamW с learning rate 3×10^{-4} и weight decay 10^{-2}. Обучение проводится на данных за 6 месяцев в течение 20 эпох с размером батча 512.

IV. Применение в точках контакта

Персональные рекомендации. На главной странице генерируются кандидаты из трёх источников: коллаборативная фильтрация (300 товаров), контентная близость через FAISS (200 товаров), популярные новинки (100 товаров). Финальный топ-20 формируется через переранжирование по формуле (4).

Похожие товары на карточке. Кандидаты извлекаются через Swing-алгоритм (до 500 товаров) и переранжируются с учётом семантической близости. Для новых товаров без истории ($n_i < 3$) используются только эмбеддинги с поиском в той же категории.

Рекомендации в корзине. Для каждого товара строятся дополнения через граф совместных покупок. Кандидаты ранжируются по релевантности всей корзине через среднее сходство с вектором корзины $\mathbf{e}_{\text{basket}}$.

Персонализация промо. Акционные товары ранжируются по релевантности интересам пользователя, избегая показа нерелевантных скидок.

V. Экспериментальная оценка

Система оценивалась на данных маркетплейса с каталогом 78 тыс. товаров и 850 тыс. активных пользователей за 3 месяца. Проводилось A/B-тестирование на 15% трафика.

Онлайн-метрики

Таблица 1 – Результаты A/B-теста гибридной системы

Метрика	Базовая	Гибридная	Прирост
CTR@10	7,82%	8,35%	+6,8%
Conversion@10	1,64%	1,79%	+9,1%
Add-to-Cart	4,21%	4,53%	+7,6%

Увеличение CTR на 6,8% статистически значимо (p-value < 0.01) и указывает на повышение релевантности. Прирост конверсии на 9,1% напрямую влияет на выручку платформы.

Покрытие и холодный старт

Базовая система обеспечивала попадание в рекомендации 54,2% товаров за месяц. Гибридная система увеличила покрытие до 57,2% (+3,0 п.п.). Длинный хвост (товары с менее 10 взаимодействиями) увеличил представленность с 11,3% до 19,7%.

Для новых товаров (первые 7 дней в каталоге): базовая система показала попадание 14,8%, CTR 2,3%; гибридная – 61,2% и 5,1% соответственно. Рост попадания в 4,1 раза подтверждает эффективность контентного подхода.

Разнообразие выдачи

Среднее число уникальных категорий в топ-10 увеличилось с 2,8 до 3,4. Индекс концентрации Джини снизился с 0,71 до 0,66, что указывает на более равномерное распределение внимания между товарами.

Анализ параметра α показал оптимум при 0,5–0,7. Адаптивная схема позволяет динамически выбирать оптимальный баланс для каждого товара.

Заключение

В работе предложен гибридный подход к построению рекомендательной системы, объединяющий коллаборативную фильтрацию и контентные методы через векторные представления. Использование компактных 32-мерных эмбеддингов обеспечивает вычислительную эффективность при сохранении различительной способности.

Ключевые результаты: увеличение CTR на 6,8%, конверсии на 9,1%, покрытия каталога на 3,0 п.п., эффективное решение проблемы холодного старта с четырёхкратным ростом видимости новых товаров. Механизм адаптивного взвешивания позволяет автоматически выбирать оптимальную стратегию в зависимости от доступности статистики [8].

Дальнейшие направления включают интеграцию визуальной модальности через кодирование изображений товаров, разработку специализированных архитектур для электронной коммерции и применение методов активного обучения для адаптации к сезонным трендам.

1. Koren, Y. Matrix Factorization Techniques for Recommender Systems / Y. Koren, R. Bell, C. Volinsky // IEEE Computer. – 2009. – Vol. 42, № 8. – P. 30–37.
2. Huang, J. Embedding-based Retrieval in Facebook Search / J. Huang [et al.] // Proceedings of the 26th ACM SIGKDD Conference. – 2020. – P. 2553–2561.
3. Covington, P. Deep Neural Networks for YouTube Recommendations / P. Covington, J. Adams, E. Sargin // Proceedings of the 10th ACM Conference on Recommender Systems. – 2016. – P. 191–198.
4. Devlin, J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin [et al.] // Proceedings of NAACL-HLT. – 2019. – P. 4171–4186.
5. Malkov, Y. A. Efficient and robust approximate nearest neighbor search using Hierarchical Navigable Small World graphs / Y. A. Malkov, D. A. Yashunin // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2020. – Vol. 42, № 4. – P. 824–836.
6. Sanz-Cruzado, J. Structural novelty and diversity in link prediction / J. Sanz-Cruzado [et al.] // Proceedings of the 9th International Workshop on News Recommender Systems. – 2018. – P. 12–19.
7. Devlin, J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding / J. Devlin [et al.] // Proceedings of NAACL-HLT. – 2019. – P. 4171–4186.
8. Нестеренков, С. Н. Методы построения рекомендательных систем на основе глубокого обучения / С. Н. Нестеренков, А. И. Петров // Доклады БГУИР. – 2023. – Т. 21, № 4. – С. 62–69.

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ В УПРАВЛЕНИИ СКЛАДСКИМИ ЗАПАСАМИ

Логвинович М. В., Коростелёв А. Д., Марков А. Н.
Центр информатизации и инновационных разработок, кафедра информатики,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {m.logvinovich, a.korostelev, a.n.markov}@bsuir.by

В условиях растущих объемов данных и постоянно меняющегося рынка стандартные методы управления складскими запасами часто оказываются неэффективными. Данная статья рассматривает, как машинное обучение (ML) может модернизировать эту область, обеспечивая более точное прогнозирование спроса и оптимизацию уровней запасов. В работе анализируются ключевые ML-алгоритмы и их преимущества по сравнению с традиционными статистическими методами.

Введение

Современные логистические системы и складские хозяйства сталкиваются с растущими объемами информации и необходимостью оптимизации процессов управления запасами. Эффективное управление складом напрямую влияет на снижение издержек, повышение скорости обработки заказов и удовлетворённость клиентов. В условиях нестабильного спроса классические подходы планирования и учёта запасов, основанные на статистических моделях и экспертных оценках, часто оказываются недостаточно точными и адаптивными.

В последние годы наблюдается стремительное развитие методов машинного обучения (Machine Learning, ML), которые позволяют анализировать большие объёмы данных, выявлять скрытые закономерности и строить прогнозы с высокой точностью [1–2]. Использование ML в управлении запасами даёт возможность прогнозировать спрос на товары, оптимизировать уровни запасов, снижать избыточные и недостающие запасы, а также повышать эффективность распределения ресурсов на складе.

I. Проблематика управления запасами

Обычно методы прогнозирования складских запасов опираются в первую очередь на анализ исторических данных и статистические модели временных рядов. Наиболее распространённые подходы включают скользящие средние, экспоненциальное сглаживание и регрессионные модели. Несмотря на простоту, такие методы имеют серьёзные ограничения:

- Недостаточная гибкость. Стандартные модели плохо учитывают резкие изменения спроса, вызванные сезонностью, промоакциями, изменением поведения покупателей или внешними экономическими факторами.
- Линейные зависимости. Методы чаще всего предполагают наличие линейных зависимостей между параметрами, тогда как реальные процессы в логистике гораздо сложнее.
- Отсутствие адаптивности. При изменении рыночной ситуации требуется заново перенастраивать параметры моделей.

В результате компании сталкиваются либо с избыточными запасами, которые увеличивают расходы на хранение и замораживают оборотный капитал, либо с дефицитом товаров, что приводит к потерям клиентов и снижению прибыли.

II. Методы машинного обучения в прогнозировании спроса

В отличие от стандартных статистических подходов, методы машинного обучения позволяют строить нелинейные модели, учитывать множество факторов одновременно и адаптироваться к изменяющимся условиям. В практике управления складскими запасами наиболее часто применяются следующие алгоритмы:

1. Решающие деревья и ансамблевые методы (Random Forest, Gradient Boosting, XGBoost). Эти алгоритмы хорошо работают с табличными данными и способны выявлять сложные зависимости между факторами. Они учитывают такие параметры, как сезонность, маркетинговые акции, макроэкономические показатели и поведение покупателей.
2. Нейронные сети (в частности, рекуррентные нейронные сети LSTM). Данный класс моделей особенно эффективен при прогнозировании временных рядов. LSTM-сети способны учитывать долгосрочные зависимости и лучше предсказывать динамику спроса, чем базовые статистические модели.
3. Методы кластеризации (например, k-means). Используются для сегментации ассортимента по характеристикам спроса и оптимизации стратегий управления для каждой группы товаров.
4. Обучение с подкреплением (Reinforcement Learning). Хотя этот подход пока используется ограниченно, он позволяет моделировать адаптивные системы управления запасами, которые принимают решения на основе обратной связи из окружающей среды.

III. Применение ML в проектировании и оптимизации склада

Помимо прогнозирования спроса, машинное обучение активно применяется для поддержки стратегических решений в области проектирования и планирования склада. Анализ больших объёмов данных из систем управления складом (WMS) позволяет создавать классификаторы, которые могут предсказывать оптимальные параметры для различных компонентов системы. К таким компонентам относятся:

- Технология хранения (ST). Алгоритмы ML могут рекомендовать подходящую систему хранения, например, автоматизированные системы хранения и поиска (AS/RS), консольные стеллажи или стеллажи для поддонов.
- Система обработки материалов (MHS). Машинное обучение помогает выбрать оптимальный набор ресурсов для перемещения товаров, таких как вилочные погрузчики или операторы.
- Стратегия размещения товаров (SAS). ML-модели могут определять наилучшие политики размещения, например, дублирование мест хранения для одного и того же SKU (Stock Keeping Unit, единица складского учёта) с целью ускорения операций комплектации.
- Политика комплектации (PP). Алгоритмы прогнозируют, как следует организовывать операции по комплектации заказов (например, по одному заказу или по нескольким).

Этот подход, основанный на анализе данных из реальных складских систем, позволяет оптимизировать процессы, повысить производительность и снизить затраты.

IV. Практические результаты применения ML

Ряд исследований подтверждает эффективность внедрения алгоритмов машинного обучения в управление складскими запасами. Так, в работе [3] проведено сравнение устоявшихся методов прогнозирования с современными ML-алгоритмами на основе исторических данных о продажах, включая сезонность, праздничные периоды и маркетинговые акции. Результаты показали, что ансамблевые методы, такие как Random Forest и XGBoost, позволили сократить избыточные запасы примерно на 20% по сравнению с классическими статистическими подходами. Использование LSTM-сетей обеспечило наибольшую точность прогнозов в условиях резких изменений спроса. Внедрение ML-моделей в систему управления запасами также способствовало снижению частоты дефицита товаров и повышению уровня обслуживания клиентов.

Кроме того, в исследовании [4] показано, что методы машинного обучения могут применяться не только для прогнозирования спроса, но и для поддержки стратегических решений в проектировании склада. Алгоритмы позволяют предсказывать подходящие конфигурации систем хранения и обработки материалов, что повышает эффективность планирования логистической инфраструктуры.

V. Выводы

Применение методов машинного обучения в управлении складскими запасами открывает новые возможности для повышения эффективности логистических процессов. В отличие от традиционных статистических моделей, ML-алгоритмы способны учитывать множество факторов, выявлять скрытые зависимости и адаптироваться к изменяющимся условиям рынка.

Практические исследования показывают, что использование таких моделей, как Random Forest, XGBoost и LSTM, позволяет существенно увеличить точность прогнозов спроса и снизить издержки, связанные с хранением избыточных запасов. Кроме того, ML-подход может быть использован для решения стратегических задач проектирования склада, таких как выбор оптимальных технологий хранения и систем комплектации, что способствует переходу от реактивного управления к проактивному, основанному на данных.

Перспективными направлениями дальнейших исследований являются применение методов глубокого обучения для интеграции различных источников данных, а также использование методов обучения с подкреплением для построения адаптивных систем управления складом.

VI. Список литературы

1. Иващеко, М. Г. Эффективное использование нейронных сетей в решении задач автоматизации в логистике / М. Г. Иващенко, С. Н. Нестеренков, А. В. Ситников // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. X Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13 марта 2024 года). В 2 ч. Ч. 1 / редкол. : В. А. Богуш [и др.]. – Минск : БГУИР, 2024. – С. 368–372.
2. Furlan de Assis, R., Frias, A. F., Thomasset-Laperriére, V., Santa-Eulalia, L. A., Ouhimmou, M., de Paula Ferreira, W. (2024). Machine Learning in Warehouse Management: A Survey. Procedia Computer Science, 232, 2790-2799. doi:10.1016/j.procs.2024.03.364
3. Bendhi, M. R. (2023). Inventory Management: Machine Learning Predicts Demand, Reducing Excess Stock by up to 20%. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 4(2), 67-83. doi:10.63282/3050-9262.IJAIDSML-V412P108
4. Tufano, A., Accorsi, R., Manzini, R. (2022). A machine learning approach for predictive warehouse design. The International Journal of Advanced Manufacturing Technology, 119, 2369-2392. doi:10.1007/s00170-021-08035-w

ИНТЕЛЛЕКТУАЛЬНЫЕ АЛГОРИТМЫ ДЛЯ АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ КАНДИДАТОВ ПО ФОТОИЗОБРАЖЕНИЮ

Малащук М. Ю.
Кафедра информационных технологий автоматизированных систем
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: maximopolnate@gmail.com

В статье приведен обзор подходов к построению интеллектуальных алгоритмов, позволяющих проводить автоматизированную оценку кандидатов по фотоизображениям. Рассматриваются методы извлечения признаков лица, применения регрессионных моделей и методов машинного обучения, а также перспективы интеграции данных технологий.

Введение

В современном мире искусственный интеллект все больше находит свое применение в различных сферах и способен решать новые задачи. Одна из них – необходимость автоматизировать оценку кандидатов. Помимо традиционных методов анализа резюме и интервью, активно исследуются технологии компьютерного зрения и искусственного интеллекта. Одним из актуальных направлений является автоматизированная оценка характеристик человека по фотоизображению, в частности уровня интеллекта или профессиональной пригодности.

I. Методы извлечения признаков

Для оценки кандидатов по фотоизображению необходимо выделить признаки, которые отражают ключевые особенности лица и могут быть связаны с когнитивными характеристиками. От качества выделенных признаков напрямую зависит точность последующего анализа и надежность предсказаний. На сегодняшний день можно выделить три основные группы признаков.

В первую очередь выделяются геометрические признаки лица. Этот подход основывается на анализе формы и пропорций лица. Ключевыми параметрами выступают расстояния между глазами, длина и форма носа, контуры подбородка, соотношение ширины и высоты лица, а также степень симметрии [1]. Геометрические характеристики позволяют формализовать структуру лица и выявить особенности, которые традиционно ассоциируются с восприятием интеллекта и других личностных качеств. Дополнительно могут использоваться методы антропометрического анализа и алгоритмы для построения «каркаса лица».

Далее необходимо выделить текстурные признаки. Данный подход направлен на исследование текстуры кожи, наличия и характера морщин, микровыражений, а также других деталей поверхности лица. Текстурный анализ помогает зафиксировать устойчивые визуальные паттерны, которые отражают эмоциональные и поведенческие особенности. Для извлечения таких признаков часто применяются методы фильтрации (например, фильтры Габора, локальные бинарные паттерны) и вейвлет-преобразование, которые позволяют учитывать как локальные, так и глобальные особенности текстур.

Завершающую группу образуют грубинные признаки. Современный уровень анализа изображений в значительной степени связан с применением глубокого обучения. Сверточные нейронные сети, обученные на больших выборках лицевых изображений, способны автоматически извлекать сложные и высокоуровневые представления. Такие признаки учитывают не только геометрию и текстуру, но и скрытые взаимосвязи между ними. Использование предобученных моделей (например, FaceNet, ResNet) позволяет ускорить процесс разработки и повысить точность анализа, а настройка этих моделей под конкретную задачу делает систему более специализированной.

Поэтому выделение признаков является первым ключевым этапом построения интеллектуального алгоритма оценки кандидатов. Комбинация различных категорий признаков позволит повысить точность анализа и сделать систему более устойчивой к внешним факторам, таким как освещенность, ракурс и качество изображения.

II. Модели анализа

После того как признаки были выделены и представлены в числовом виде, следующим этапом является их анализ с помощью математических моделей. На данном этапе необходимо выявить зависимости между признаками и когнитивными характеристиками кандидата, а также построить алгоритм, способный выдавать групповые оценки. В зависимости от поставленной задачи и характера данных могут применяться различные группы моделей.

В первую очередь применяются регрессионные модели. Регрессия используется для предсказания числового показателя, например уровня интеллекта по определенной шкале. Наиболее

распространенными подходами являются Support Vector Regression (SVR), линейная и полиномиальная регрессия. Эти методы позволяют установить зависимость между набором признаков и целевой переменной, а также давать количественную оценку когнитивных характеристик. Преимуществом регрессионных моделей является их интерпретируемость и возможность выдавать конкретный результат, а не только категориальное суждение. Однако они требуют корректной подготовки данных и чувствительны к шумам в признаках.

Далее используются классификационные модели. Классификация применяется для разделения кандидатов на группы, например:«выше среднего»,«средний»,«ниже среднего». Эти подходы удобны в ситуациях, когда требуется обобщенное решение. Среди наиболее часто используемых алгоритмов можно выделить логистическую регрессию, решающие деревья и методы градиентного бустинга. Такие модели хорошо работают при наличии достаточно больших и хорошо сбалансированных выборок, позволяя выявлять закономерности в данных и использовать для принятия решений.

Интегрированные модели позволяют объединять результаты нескольких алгоритмов для повышения общей точности и устойчивости системы. Наиболее известные методы включают бустинг, стекинг и различные схемы голосования. Использование интеграции позволяет снизить риск переобучения и делает систему более надежной при работе с разнородными данными. Такой подход особенно актуален при анализе больших массивов изображений, когда отдельная модель может быть недостаточно универсальной. Кроме того, интегрированные модели позволяют комбинировать как простые алгоритмы, так и более сложные, усиливая их сильные стороны.

Поэтому выбор модели анализа зависит от поставленной задачи. Если требуется числовая оценка, то используется регрессия. Если нужно группировать кандидатов, то применяются классификационные модели. Для максимальной точности и устойчивости системы применяются интегрированные модели.

III. Этические аспекты

Применение интеллектуальных алгоритмов для оценки кандидатов по фотографиям связано с рядом этических вопросов. Во-первых, существует риск предвзятого отношения алгоритма: система может некорректно интерпретировать особенности внешности, что приведет к дискриминации отдельных групп. Во-вторых, обработка фотографий затрагивает вопросы конфиденциальности и права на личную неприкосновенность. Поэтому, согласно законодательству, требуется согласие кандидатов, а также высокий уровень защиты информации [2].

Отдельное внимание следует уделить прозрачности решений алгоритма. Если система автоматически формирует оценку интеллекта или профессиональной пригодности, кандидат имеет право знать, на каких признаках был сделан вывод. Это повышает доверие и снижает вероятность необоснованного отказа.

В перспективе особое значение приобретает формирование единых этических стандартов и нормативных регламентов использования подобных технологий. Их разработка предполагает междисциплинарное взаимодействие специалистов в области искусственного интеллекта, юриспруденции и социальной психологии, что позволит обеспечить баланс между инновационностью решений и необходимостью соблюдения принципов справедливости, прозрачности и недискриминации.

Таким образом, при внедрении подобных систем необходимо сочетать техническую эффективность с этическими нормами, чтобы избежать ошибок, несправедливости и снижения доверия со стороны общества.

IV. Заключение

Интеллектуальные алгоритмы анализа кандидатов по фотоизображениям представляют собой перспективное направление исследований. Они могут стать вспомогательным инструментом в HR-практиках. Однако их применение требуют осторожности из-за вопросов этики, приватности и законодательства. Дальнейшие исследования будут направлены на повышение интерпретируемости моделей и расширение набора признаков для оценки.

V. Список литературы

1. Wei X., Stillwel D. How smart does your profile image look? Estimating intelligence form social network profile images. / Psychomeytics Centre, University of Cambridge, School of Managementб University of Bath, June 2016, [Electronic resource]. – Mode of access: `https://www.researchgate.net/publication/304642532`. – Date of access: 20.09.2025.
2. Floridi L., Cowls J. A unified framework of Five Principles for AI in Society. Harvard Data Science Review, 2019, [Electronic resource]. – Mode of access: `https://hdsr.mitpress.mit.edu/pub/l0jsh9d1/release/8`. – Date of access: 30.09.2025.

ОБНАРУЖЕНИЕ ПРОГРАММ-ВЫМОГАТЕЛЕЙ НА ОСНОВЕ КОМБИНИРОВАННОЙ НЕЙРОННОЙ СЕТИ И АНАЛИЗА ЭНТРОПИИ ДАННЫХ

Мальцев В. Л., Возмитель В. В., Матяс Е. В.
Кафедра защиты информации,
Белорусский государственный университет информатики и радиоэлектороники
Национальный детский технопарк
Минск, Республика Беларусь
E-mail: {viktor.maltsevlul, racfor4}@gmail.com

В условиях растущего числа изощренных кибератак программы-вымогатели представляют собой одну из наиболее серьезных угроз информационной безопасности. Традиционные методы защиты и методы, основанные на сигнатурах и анализе энтропии данных, нередко оказываются неэффективными из-за высокого уровня ложных срабатываний и неспособности надежно отличить зашифрованные данные от легитимных файлов со сжатием. В качестве решения данной проблемы предлагается двухуровневый подход анализа данных, совмещающий в себе разные методы. На первом этапе выполняется быстрая фильтрация файлов по порогу энтропии для отбора «подозрительных» объектов. На втором этапе эти файлы анализируются гибридной нейронной сетью, объединяющей сверточную сеть для выявления локальных паттернов и архитектуру Transformer для анализа глобальных контекстных зависимостей в данных.

Введение

В современном цифровом мире, где информация является одним из самых ценных активов, киберугрозы и кибератаки постоянно эволюционируют, становясь все более изощренными и опасными. Одной из наиболее серьезных и распространенных угроз являются кибератаки с использованием программ-вымогателей. В настоящее время количество таких атак неумолимо растет, нанося колоссальный ущерб как крупным компаниям, так и обычным пользователям. С развитием искусственного интеллекта стали возможными и кибератаки с использованием искусственных нейронных сетей. Одним из перспективных подходов к выявлению несанкционированного шифрования является анализ энтропии данных. Этот метод основывается на алгоритме энтропии Шеннона, который является мерой неопределенности или случайности данных [1][2].

Основной раздел

Идея заключается в том, что зашифрованные данные обладают очень высокой степенью случайности, и их показатель энтропии приближается к максимальному значению, равному примерно 8. Однако этот метод имеет существенный недостаток. Высокую энтропию также имеют и многие популярные и абсолютно легитимные типы файлов. К ним относятся, например, архивы (ZIP, RAR), многие форматы документов (например, .docx), а также видео и графические файлы, которые используют сжатие. Этот факт делает невозможным точное выявление работы программ-вымогателей на основе одного лишь порога энтропии, так как это привело бы к огромному количеству ложных срабатываний. Чтобы наглядно продемонстрировать эту проблему энтропии, был проведен анализ множества файлов в двух состояниях: до и после шифрования. Результаты, представленные на графике распределения энтропии для зашифрованных и незашифрованных файлов, показывают степень их энтропии.

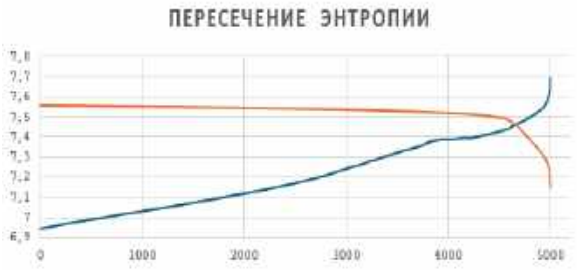


Рис. 1 – График распределения энтропии для зашифрованных и незашифрованных файлов

На графике (рисунок 1) представлено пересечение двух кривых, где синяя линия соответствует обычным, незашифрованным файлам, а оранжевая – тем же файлам, но уже в зашифрованном виде. Эта иллюстрация наглядно показывает две ключевые проблемы:

1. Значительное количество обычных, незашифрованных файлов имеет энтропию, равную или даже превышающую показатели многих зашифрованных файлов.

2. Некоторые зашифрованные файлы, в свою очередь, могут иметь низкий или относительно низкий показатель энтропии. Это может быть связано с использованием алгоритмов шифрования с низким уровнем энтропии на начальных этапах процесса или с частичным шифрованием файла.

Если провести эксперимент по выявлению зашифрованных файлов, установив оптимальное

пороговое значение энтропии на уровне 7.45, можно получить следующие результаты:

- ложно положительные срабатывания: 5.42 процента;
- ложно отрицательные срабатывания: 0.34 процента;
- истинно положительные срабатывания: 94.58 процентов;
- истинно отрицательные срабатывания: 99.66 процентов.

Результаты этого теста показывают, что практически каждый 20-й файл из тестового набора был бы ошибочно опознан как зашифрованный. Такая высокая погрешность делает невозможным практическое применение данного метода для корректного выявления программ-вымогателей в реальных условиях.

Для решения проблемы выявления была разработана и обучена нейронная сеть передовой комбинированной архитектуры, состоящая из двух ключевых модулей: CNN (Convolutional Neural Network, сверточная нейронная сеть) и Transformer.

Модуль CNN: модуль специализируется на выявлении локальных зависимостей и паттернов в последовательности данных. В контексте анализа файла он способен выделить из потока байтов характерные микро-признаки, которые могут указывать на процесс шифрования. Поле свертки позволяет эффективно анализировать локальные зависимости между байтами и сократить длину самой последовательности в два раза для последующей обработки.

Модуль Transformer: архитектура, созданная для обработки естественного языка, превосходно справляется с выявлением глобальных зависимостей в данных. После предварительной обработки модулем CNN, Transformer анализирует последовательность в целом, оценивая взаимосвязи между удаленными друг от друга участками файла. Это позволяет нейросети формировать целостное «понимание» структуры файла, а не просто реагировать на отдельные подозрительные фрагменты [2].

Комбинация этих двух модулей позволяет создать высокопроизводительную и точную нейронную сеть, способную улавливать как локальные, так и глобальные признаки вредоносной активности. Для обучения нейронной сети был использован обширный набор данных из открытого источника – NapierOne, суммарное количество файлов в котором составило порядка 300 тысяч. Этот датасет включает в себя как вредоносные, так и безопасные файлы, что позволяет модели «научиться» их различать.

Тестирование нейросети проводилось на специально подобранной выборке зашифрованных и незашифрованных файлов, ключевой особенностью которой было то, что все файлы имели энтропию выше значения 7.4. Оценивать работу на данных с более низкой энтропией не имеет смысла, так как именно в «спорном» диапазоне традиционные методы дают сбой, а количество зашифрованных файлов с низкой энтропией стремится к нулю.

По итогам тестирования на сложных случаях нейросеть продемонстрировала следующие показатели точности:

- ложно положительные срабатывания: 3.75 процентов (ошибочное определение безопасного файла как вредоносного);
- ложно отрицательные срабатывания: 1.63 процента (пропуск реальной угрозы);
- истинно положительные срабатывания: 96.25 процентов (корректное обнаружение угрозы);
- истинно отрицательные срабатывания: 98.37 процентов (корректное определение безопасного файла).

Если же применить двухуровневый подход (сначала быстрый тест на энтропию, а затем анализ нейросетью только «подозрительных» файлов), то итоговые показатели эффективности в абсолютных числах становятся еще более убедительными:

- ложно положительные срабатывания: 0.2 процента;
- ложно отрицательные срабатывания: 0.006 процента;
- истинно положительные срабатывания: 99.8 процентов;
- истинно отрицательные срабатывания: 99.994 процента.

Данные показатели являются абсолютно приемлемыми для использования в защитных решениях на среднестатистических персональных компьютерах.

Заключение

Таким образом, предложенный подход, сочетающий в себе сильные стороны сверточных сетей и архитектуры Transformer, способен практически безошибочно определять факт вредоносного шифрования и, следовательно, эффективно выявлять и блокировать атаки программ-вымогателей.

Список литературы

1. ESET discovers PromptLock, the first AI-powered ransomware[Electronic resource] / ESET North America.: 2025. – Mode of access: `https://www.eset.com/us/about/newsroom/research/eset-discovers-promptlock-the-first-ai-powered-ransomware/`. – Date of access: 02.10.2025.
2. Abásolo, D., Hornero, R., Espino P. Entropy analysis of the EEG background activity in Alzheimer's disease patients // Physiological Measure-ment. 2006. Vol. 27(3). P.241-253. /University of Surrey. – Mode of access: `epubs.surrey.ac.uk/39603/6-006.pdf`. – Date of access: 02.10.2025.
3. Козлов, С. Transformer – новая архитектура нейросетей для работы с последовательностями. – Никосия, Кипр, 1998–2024. – Режим доступа: `https://habr.com/ru/articles/341240/`. – Дата доступа: 02.10.2025.

ПОСТРОЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ ЧАТ-БОТОВ НА ОСНОВЕ МОДЕЛЕЙ ГЛУБОКОГО ОБУЧЕНИЯ

Мамченко К. А., Ситников А. В., Романюк М. В.
Факультет компьютерных систем и сетей,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: kmamcenko@gmail.com, a.sitnikov@bsuir.by, romanuk@bsuir.by

В статье рассматриваются современные подходы к построению интеллектуальных чат-ботов на основе моделей глубокого обучения. Проанализированы этапы подготовки диалоговых корпусов, архитектуры нейронных моделей, методы дообучения, а также метрики оценки качества диалога. Особое внимание уделено вопросам интерпретируемости, этики и тенденциям развития диалоговых систем. Обозначены ключевые направления исследований, направленные на повышение осознанности и управляемости искусственного интеллекта в человеко-машинных коммуникациях.

Введение

Появление мощных моделей глубокого обучения радикально изменило представление о возможностях диалоговых систем. Если ранее чатботы представляли собой последовательности правил или деревья принятия решений, то сегодня они используют сложные вероятностные представления языка, способные улавливать контекст, эмоциональную окраску и намерения пользователя [1].

Глубокие языковые модели, такие как GPT, BERT и их производные, продемонстрировали способность генерировать естественные, связные и стилистически согласованные тексты. Это привело к значительному росту интереса к внедрению интеллектуальных ассистентов в различные сферы – от образования и здравоохранения до электронной коммерции и государственной службы.

I. Предобработка данных и формирование корпусов

Формирование диалогового корпуса является одним из наиболее трудоёмких этапов. Важна не только количественная составляющая данных, но и их репрезентативность. Например, в корпусах технической поддержки ключевым фактором является баланс между типами запросов: информационными, эмоциональными и инструктивными.

Помимо стандартной токенизации, современные системы используют морфологический и синтаксический анализ для сохранения грамматических связей между элементами высказывания. Для языков с богатой морфологией (например, русского) это особенно важно, так как неправильная сегментация может привести к потере смысла [2].

Перспективным направлением становится создание мультидоменных корпусов, способных обеспечивать перенос знаний между различными предметными областями. Это особенно актуально при разработке универсальных моделей, предназначенных для общения в разнообразных контекстах.

II. Обучение и дообучение (fine-tuning)

Обучение диалоговых моделей включает не только настройку параметров, но и оптимизацию архитектур. Наиболее распространённым решением остаются трансформеры, обеспечивающие высокое качество генерации за счёт механизма самовнимания. При этом исследователи активно изучают способы уменьшения вычислительных затрат, связанных с обучением больших языковых моделей.

Fine-tuning на доменных данных позволяет значительно повысить релевантность ответов. Например, чат-бот для юридической консультации обучается на корпусах судебных решений и нормативных актов, что позволяет ему адекватно интерпретировать юридические термины.

Отдельное внимание уделяется обучению с подкреплением (RLHF), где человеческие оценки формируют сигнал обратной связи. Такой подход позволяет моделям не только воспроизводить язык, но и следовать коммуникативным нормам. Это особенно важно для задач, где требуется поддержание этичного и безопасного взаимодействия.

Дополнительно применяются гибридные подходы, сочетающие статистические и нейросетевые методы. Такая стратегия повышает устойчивость моделей при работе с неоднородными данными и позволяет адаптироваться к стилю и эмоциональному тону пользователя. Формируется направление «контекстно-адаптивного обучения», где параметры модели могут изменяться динамически, в зависимости от типа диалога или социальной роли собеседника. Это открывает возможности для создания персонализированных чат-ботов, способных подстраиваться под поведение и уровень подготовки пользователя.

III. Метрики оценки качества диалога

Проблема объективной оценки качества диалога остаётся одной из наиболее дискуссионных

в области ИИ. Несмотря на активное развитие автоматических метрик, их интерпретация часто требует контекстного анализа.

Современные методы включают использование embedding-моделей для вычисления семантической близости между ответами модели и эталоном. Метрика BERTScore показала, что такие подходы лучше отражают смысловую релевантность по сравнению с традиционными BLEU и ROUGE.

В то же время появляется тенденция к созданию комплексных метрик, объединяющих когерентность, информативность и эмоциональную естественность. В исследованиях всё чаще применяются человеческие опросы, где оценивается «естественность общения» – субъективный, но важный параметр, который напрямую влияет на восприятие модели пользователями.

IV. Тенденции и направления развития

Текущее поколение диалоговых систем развивается в сторону контекстно-зависимого мышления и долговременной памяти. Механизмы retrieval-augmentation позволяют системам обращаться к внешним источникам, включая базы знаний и поисковые системы, что обеспечивает точность и актуальность ответов [3].

Большое внимание уделяется мультимодальности: чат-боты начинают обрабатывать не только текст, но и изображения, аудиопотоки и видео. Такие системы применяются в виртуальных музеях, медицинской диагностике и обучающих средах, где важно учитывать зрительный или звуковой контекст.

Другим направлением стала энергоэффективность. Методы distillation и sparsity уменьшают размер моделей без потери качества, что делает возможным их применение в мобильных устройствах и робототехнических платформах.

Особое значение приобретают вопросы интерпретируемости и доверия. Исследования направлены на визуализацию процессов внимания и объяснение внутренних решений моделей. Это помогает не только анализировать ошибки, но и выявлять предвзятость, возникающую из-за несбалансированных данных [4].

Отдельного рассмотрения требуют социально-этические аспекты. Разработка фильтров токсичности, анонимизация пользовательских данных и внедрение принципов «ответственного ИИ» становятся неотъемлемой частью жизненного цикла чат-бота. В этом контексте исследовательская деятельность всё больше ориентируется на междисциплинарное взаимодействие – сочетание информатики, лингвистики, когнитивной психологии и этики.

Отдельное направление составляет использование чат-ботов в образовательных и когнитивных практиках, где они выполняют функции цифровых наставников и ассистентов преподавателя. Такие системы демонстрируют потенциал для формирования метапредметных компетенций и развития критического мышления у обучающихся. В перспективе подобные решения могут стать частью интеллектуальных обучающих сред, интегрированных с платформами адаптивного обучения и инструментами оценки прогресса студентов.

Заключение

Интеллектуальные чат-боты на основе глубокого обучения представляют собой один из наиболее быстро развивающихся классов искусственного интеллекта. Их эффективность определяется качеством обучающих данных, архитектурой нейросетей и продуманностью стратегий дообучения.

В то же время задачи объяснимости, энергоэффективности и этичности продолжают оставаться открытыми. Их решение требует синтеза инженерных и гуманитарных подходов, направленных на создание систем, которые не только понимают язык, но и способны безопасно и осмысленно взаимодействовать с человеком.

Современные тенденции указывают на переход от узкоспециализированных ботов к универсальным агентам, интегрированным в экосистему цифрового общества. Исследовательский интерес к этой области, вероятно, будет только усиливаться в ближайшие годы.

1. Данилина, Е. К. Обучение с помощью технологии чат-бот: терминология и общая характеристика / Е. К. Данилина // Вестник Волгоградского государственного университета. – 2021. – № 7. – С. 55-61.
2. Комашко, М. Н. ChatGPT, текст, информация: критический анализ / М. Н. Комашко // Труды Института системного анализа Российской академии наук. – 2023. – № 1. – С. 10-20.
3. Чат-боты как одна из тенденций развития современного образования / Е. И. Лебедева [и др.] // ExpEducation. – 2022. – № 3. – С. 45-50.
4. Клещева, Н. А., Данилина, Е. К. Чат-бот технология как информационно-коммуникативный ресурс педагогической поддержки процесса формирования метапредметных навыков / Н. А. Клещева, Е. К. Данилина // Вестник Воронежского государственного университета. – 2022. – № 4. – С. 72-79.

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕАЛИЗАЦИИ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ВЗАИМОДЕЙСТВИЯ С ИНТЕЛЛЕКТУАЛЬНЫМИ СИСТЕМАМИ

Мигалевич С. А., Скалозуб К. А., Тимошек Е. С.
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {migalevich, k.skalozub, e.timoshek}@bsuir.by

В данной работе рассматривается проблема, касающаяся современного программного обеспечения – постоянный рост сложности пользовательских интерфейсов, прямо пропорциональный увеличению функциональных возможностей систем, что приводит к затруднениям при освоении системы новыми пользователями. Предлагается концепция целеориентированного естественно-языкового интерфейса, позволяющего пользователю формулировать свои задачи на естественном языке вместо навигации по сложным иерархическим меню. Описывается возможная архитектура такой системы, основанная на комбинации современных методов обработки естественного языка и машинного обучения.

Введение

Пользовательские интерфейсы программного обеспечения прошли путь от командной строки к графическим интерфейсам, что сделало возможным их использованием людьми, которые не обладают какими-то спецефическими навыками. Все это внесло свою лепту в распространение использования вычислительной техники, которая на данный момент используется во всех сферах человеческой жизни.

В то же время, по мере того как программное обеспечение обрастало все более мощным и разнообразным функционалом для решения в том числе и крайне спецефических задач, которые актуальны для очень узкого круга пользователей, его практическая полезность для рядового пользователя при решении простых, повседневных задач зачастую не растет, а наоборот снижается, что связано с усложнением пользовательского интерфейса. Неподготовленный пользователь, сталкиваясь с такой системой, просто не знает, что ему с ней делать.

I. Проблемы современных пользовательских интерфейсов

Большинство пользователей при использовании различных систем ставят перед собой простую и конкретную цель, например, сделать так, чтобы в Microsoft Word первая страница не нумеровалась, нумерация начиналась не с единицы, а с четверки, или чтобы слова автоматически переносились на следующую строку. Но для их достижения пользователю приходится обращаться к поисковым системам, видеоурокам, форумам технической поддержки и так далее. И это при том, что описанные выше задачи являются типовыми и распространенными, а сам пользователь уже не первый раз сталкивается с подобной задачей.

О том, как качественно пользовательский интерфейс справляется со своей функцией, свидетельствует количество запросов в поисковых системах по выполнению описанных выше задач, которые уже стали привычной частью цифрового быта для большинства пользователей Microsoft Word, который является одним их самых распространенных текстовых редакторов. Как итог, пользовательский опыт при решении типовых задач все чаще может быть описан как разгадывание головоломки, которую непреднамерянно создали сами разработчики, а время, которое пользователь затрачивает на нахождение инструкций по решению своей проблемы, может многократно превышать то время, которое необходимо было потратить на то, чтобы претворить эту инструкцию в жизнь.

II. Методы решения вознишкей проблемы

Традиционно для того, чтобы помочь пользователю справиться с освоением некоторого программного средства, существует документация, справка, кнопка помощи и так далее. Однако, если у пользователя возникла необходимость этим всем воспользоваться, то это уже свидетельствует о том, что пользовательский интерфейс со своей задачей не справился, не говоря уже о том, чтобы пользователю пришлось искать информацию по решению задачи в интернете.

Большинство сложных систем, предоставляющих широкий спектр интсрументов для решения вознишких задач пользователя, имееют поиск необходимого пункта в меню, предлагающий ввести пользователю название нужного ему пункта меню. Проблема такого подхода заключается в том, что для получения какой бы то ни было пользы от данного решения, пользователь должен точно знать как называется нужная ему функция или, как минимум, он должен догады-

ваться о том, как она может называться, что в большинстве случаев не может помочь пользователю с решением новой для себя задачи. Например, поиск по запросу «сделать первую страницу без номера» не найдет функцию «особый колонтитул для первой страницы», хотя именно она решает поставленную задачу. Данное решение способно лишь сэкономить время уже опытному пользователю.

Помочь с решением данной проблемы может подход, основанный на семантическом понимании пользовательского запроса. Вместо поиска по ключевым словам предлагается использовать модели машинного обучения, способные улавливать смысловое сходство между разговорными формулировками пользователя и названиями функций интерфейса и их описанием. Современные языковые модели, такие как BERT, преобразуют и пользовательский запрос, и все элементы интерфейса в векторные представления – числовые последовательности, кодирующие их смысл. Ключевым показателем становится косинусное сходство между векторами, вычисляемое через скалярное произведение и позволяющее количественно измерить смысловую близость между разнородными формулировками.

Техническая реализация предполагает создание расширенной базы знаний, где каждый элемент интерфейса описывается не только официальным названием, но и десятками вариантов описания на естественном языке, включая типичные ошибки и разговорные выражения. Нейросетевая модель дообучается на этой базе, приобретая способность понимать, что «убрать красные подчеркивания» означает «отключить проверку орфографии», а «сделать, чтобы везде было одинаково» может соответствовать функции «применить стиль». [1]

Помимо базового подхода с использованием векторных представлений, существуют и другие методы организации семантического поиска в пользовательских интерфейсах. Одним из таких методов является комбинация языковых моделей с графами знаний. В этом случае строится семантическая сеть, где узлы представляют функции интерфейса, а ребра отражают смысловые связи между ними. Когда пользователь формулирует запрос, система не только вычисляет векторное сходство, но и анализирует пути в этом графе, находя функционально связанные элементы, которые могут решить задачу более комплексно. [2]

Интересный подход демонстрируют гибридные системы, сочетающие несколько типов языковых моделей. Крупные модели, такие как GPT, могут генерировать возможные альтернативные формулировки запроса, в то время как более легкие модели, такие как SentenceBERT, эффективно вычисляют семантическое сходство. Это позволяет охватить более широкий спектр возможных способов описания одной и той же функции без чрезмерных вычислительных затрат. Например, запрос «сделать шрифт жирным» может быть перефразирован до «увеличить толщину букв», «сделать текст более заметным», «выделить шрифт» – и все эти варианты можно одновременно проверить на соответствие элементам интерфейса. [3]

Многоязычные языковые модели открывают дополнительные возможности для локализации интерфейсов. Пользователь может формулировать запрос на родном языке, а система сможет находить соответствующие функции в интерфейсе, который может быть реализован на другом языке, что сможет заметно упростить использование программного средства людьми, чей родной язык не является английским. Это может быть особенно актуально для международных компаний, где сотрудники чаще всего используют программное обеспечение с английским интерфейсом, но формулировать запросы смогут на родном языке. [4]

III. Заключение

Подводя итоги, одним из главных преимуществ естественно-языкового интерфейса является значительная экономия времени и сил пользователя. Вместо долгого блуждания по меню в попытках вспомнить, как и где он решал аналогичную проблему в прошлый раз, поиска инструкций в интернете или просмотра видеоуроков, он сможет описать свою задачу словами на естественном и получить результат. Это не только избавляет от постоянного «разгадывания головоломки», созданной сложным интерфейсом, но и позволяет быстрее обучаться работе в системе, что особенно актуально, когда перед пользователем возникает новая для него задача.

1. Goldberg, Y. Neural Network Methods for Natural Language Processing / Y. Goldberg. – San Rafael : Morgan & Claypool Publishers, 2017. – 287 p.
2. Jurafsky D. Speech and Language Processing [Electronic resource] / D. Jurafsky, J. H. Martin. – 3rd ed. – 2025. – 622 p. – Mode of access: `https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf`. – Date of access: 07.10.2025.
3. Yakhontova, I. M., Netrebin, N. M., & Streletskiy, A. D. (2025). Intelligent question-answering system: integration of knowledge bases and language models to improve the efficiency of information processing. – Modern Innovations, Systems and Technologies, 5(1), 1020-1026 p.
4. Добровольский, Д. С., Олейников, В. С., Рац, Е. С. Разработка интеллектуальной системы для автоматического ответа на отзывы покупателей на маркетплейсах с применением LLM // Научный результат. Информационные технологии. – Т. 9, № 4, 2024. – С. 74-84.

КЛАССИФИКАЦИЯ И ОБЗОР МЕТОДОВ ПАРАМЕТРО-ЭФФЕКТИВНОЙ АДАПТАЦИИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ (LLM)

Нестеренков С. Н., Фурсанов С. А., Камышев С. В.
Кафедра программного обеспечения информационных технологий,
центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {s.nesterenkov, s.fursanov, s.kamyshev}@bsuir.by

В последние годы большие языковые модели (LLM, Large Language Models) стали ключевым инструментом в задачах обработки естественного языка. Их применение в специализированных предметных областях сталкивается с проблемами высокой вычислительной сложности и ограниченной доступности данных для дообучения. В данной работе представлен обзор существующих методов параметро-эффективной адаптации LLM, проводится их классификация и сравнительный анализ. Основное внимание уделено подходам, позволяющим адаптировать модели без изменения всех весов, что делает их более практичными для узкоспециализированных задач.

Введение

Большие языковые модели, такие как GPT и LLaMA, демонстрируют высокую эффективность в широком спектре задач генерации текста и семантического анализа [1,2]. Тем не менее, применение LLM в узкоспециализированных областях (например, медицина [3], юриспруденция, инженерные тексты, образование) требует адаптации к специфике терминологии и формулировок [4].

Это связано с тем, что полное дообучение (full fine-tuning) не только требует колоссальных вычислительных ресурсов и времени, но и несет в себе риск «катастрофического забывания» – явления, при котором модель, адаптируясь к новой узкой задаче, теряет часть общих знаний, полученных на этапе предварительного обучения на разнородных данных. Это делает модель менее гибкой и универсальной. Полное дообучение модели требует модификации всех параметров, что при сотнях миллиардов весов практически невозможно.

Появление методов параметро-эффективной адаптации (PEFT, Parameter-Efficient Fine-Tuning) позволяет решить эту проблему, изменяя лишь малую часть параметров модели [5]. Данные методы достигают высокой эффективности за счет того, что не перезаписывают исходные знания модели, а добавляют к ним легковесные, задачно-специфичные модули или производят низкоранговые корректировки весов. Такой подход не только радикально снижает вычислительные затраты и объем требуемых данных для дообучения, но и позволяет быстро переключаться между разными задачами, используя одну базовую модель. В следующем разделе представлена классификация и подробный обзор наиболее распространенных методов PEFT.

I. Обзор методов параметро-эффективной адаптации

Существующие подходы к PEFT можно разделить на несколько основных категорий:

- Adapter-tuning. Метод Adapter-tuning предполагает встраивание в архитектуру модели дополнительных небольших слоев – адаптеров, которые обучаются на специализированных данных, в то время как базовые веса трансформера остаются замороженными [5,6]. Такой подход позволяет использовать одну и ту же модель для различных задач, снижает вычислительные затраты и объем необходимых данных. Типичная архитектура адаптера представляет собой компактную последовательность слоев, часто по схеме «bottleneck», которая включает полносвязный слой, сжимающий эмбеддинги до меньшей размерности (down-project), нелинейную функцию активации (например, GeLU) и слой, восстанавливающий исходную размерность (up-project). Такие модули обычно встраиваются после блока внимания (Multi-Head Attention) и после блока прямой связи (Feed-Forward Network) в каждом слое трансформера. Ключевым преимуществом является их модульность: для новой задачи можно обучить и затем хранить лишь небольшой файл с весами адаптера, «подключая» его к замороженной базовой модели по мере необходимости.
- LoRA (Low-Rank Adaptation). Алгоритм LoRA основан на разложении весовых матриц на низкоранговые компоненты, которые подлежат обучению [7]. Основные веса модели остаются неизменными. LoRA позволяет изменять до 0,1–1 % параметров модели, что делает возможным использование LLM с десятками миллиардов весов даже на ограниченных ресурсах.

– Prompt-tuning и Prefix-tuning. Эти методы не модифицируют архитектуру модели, а используют обучаемые векторы-подсказки (prompts), подаваемые на вход модели [8]. Prompt-tuning оптимизирует фиксированные подсказки, а Prefix-tuning – обучаемые последовательности токенов, которые расширяют вход модели. Подходы особенно эффективны для задач генерации текста и диалоговых систем. Prefix-tuning работает на уровне скрытых состояний (hidden states) модели. Он добавляет последовательность непрерывных (не дискретных) векторов («префикс») к активациям каждого слоя трансформера, причем эти векторы обучаются методом обратного распространения ошибки. В отличие от него, Prompt-tuning является более простым методом, который работает только на входном уровне модели, оптимизируя непрерывные векторы-подсказки, конкатенируемые с эмбеддингами входных токенов. Со временем появились и более продвинутые вариации, такие как P-Tuning, который использует LSTM или MLP для генерации более эффективных промптов.

II. Классификация методов

С точки зрения количества изменяемых параметров, PEFT-методы можно классифицировать следующим образом:

– Минимальные изменения (0,1–1 % параметров) – LoRA, Prefix-tuning.
– Средние изменения (1–3 % параметров) – Adapter-tuning, P-tuning.
– Полное дообучение (100 % параметров) – классический fine-tuning, используется редко из-за вычислительных ограничений.

По назначению методы можно разделить на:

– Универсальные адаптеры – применимы к различным задачам без модификации основной модели.
– Задачно-специфические адаптеры – обучаются под конкретный домен, обеспечивая максимальное качество.

III. Сравнительный анализ

Выбор метода PEFT зависит от компромисса между эффективностью, затратами и универсальностью. Prompt-tuning наиболее экономен (менее 0.1% параметров), но эффективен только в крупных моделях (от 20 млрд параметров). LoRA (0.1-1% параметров) не замедляет инференс и идеальна для ограниченных ресурсов, хотя в основном адаптирует только механизмы внимания. Adapter-tuning (1-3% параметров) демонстрирует наибольшую стабильность в сложных многоэтапных задачах, но добавляет небольшую вычислительную нагрузку при inference. Prefix-tuning показывает выдающиеся результаты в генеративных сценариях, однако требует тщательной настройки гиперпараметров.

Важным практическим аспектом является возможность комбинирования методов. Например, адаптеры можно комбинировать с обучением промптов, а низкоранговые матрицы LoRA – с префикс-тюнингом. Такие гибридные подходы открывают новые возможности для балансировки между параметрической эффективностью и качеством адаптации.

Таким образом, LoRA оптимальна для задач классификации при ограниченных ресурсах, Adapter-tuning – для сложных задач, требующих максимального качества, Prefix-tuning – для генерации, а Prompt-tuning – для сверхбольших моделей. Перспективным направлением является разработка адаптивных методов, автоматически подбирающих тип и масштаб адаптации под конкретную задачу.

Заключение

Параметро-эффективные методы адаптации больших языковых моделей открывают новые возможности для внедрения LLM в специализированные предметные области. Дальнейшие исследования должны быть направлены на:

– Разработку гибридных методов, сочетающих преимущества различных подходов;
– Формирование критериев выбора метода в зависимости от задач и ресурсов;
– Создание экспериментальных прототипов адаптированных моделей для разных доменов.

IV. Список литературы

1. Vaswani A., Shazeer N., Parmar N. и др. Attention Is All You Need // Advances in Neural Information Processing Systems. – 2017. – Vol. 30. – P. 5998-6008.
2. Touvron H., Martin L., Stone K. и др. LLaMA: Open and Efficient Foundation Language Models // arXiv preprint arXiv:2302.13971. – 2023.
3. Скалозуб К. А., Нестеренков С. Н., Ярмош А. Д. Градиентные методы оптимизации в компьютерном зрении для медицинской диагностики // Технологии передачи и обработки информации: материалы Междунар. науч.-техн. семинара, Минск, апр. 2025 г. / редкол. : В. Ю. Цветков [и др.]. – Минск : БГУИР, 2025. – С. 34-36.
4. Lee J., Yoon W., Kim S. и др. BioBERT: a pre-trained biomedical language representation model for biomedical text mining // Bioinformatics. – 2020. – Vol. 36, № 4. – P. 1234-1240.
5. Houlsby N., Giurgiu A., Jastrzebski S. и др. Parameter-Efficient Transfer Learning for NLP // Proceedings of ICML. – 2019.
6. PPfeiffer J., Kamath A., Rücklé A. и др. AdapterFusion: Non-Destructive Task Composition for Transfer Learning // Proceedings of EACL. – 2021.
7. Hu E., Shen Y., Wallis P. и др. LoRA: Low-Rank Adaptation of Large Language Models // arXiv preprint arXiv:2106.09685. – 2021.
8. Li X. L., Liang P. Prefix-Tuning: Optimizing Continuous Prompts for Generation // Proceedings of ACL. – 2021.

АВТОМАТИЗАЦИЯ КОРРЕКТИРОВКИ ТЕКСТА ПРИ ПОМОЩИ LLM

Нестеренков С. Н., Осадчий О. Э.
Кафедра информатики, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by, trapper.md@mail.ru

Исследование посвящено созданию автоматизированного сервиса корректировки резюме (CV) с применением технологий искусственного интеллекта и больших языковых моделей. Разработанное решение обеспечивает выполнение ключевых функций: грамматическая проверка, перефразирование текста, перевод с американского на британский английский и перевод на немецкий язык. Особенностью системы является возможность пакетной обработки, что делает её удобной при работе с большими объёмами данных. Практическая значимость исследования заключается в повышении качества подготовки резюме и расширении инструментов автоматизации не только в бизнес, но и образовательной, а также кадровой сфере.

Введение

Развитие современных технологий, особенно в области искусственного интеллекта и машинного обучения, привело к появлению и активному применению больших языковых моделей (LLM). Эти модели находят широкое использование в задачах обработки текста: от чат-ботов до специализированных сервисов для редактирования документов. Особую актуальность приобретает автоматизация корректировки резюме (CV), поскольку ошибки и несоблюдение языковых норм могут существенно снизить конкурентоспособность соискателя на рынке труда. Использование LLM позволяет решать задачи грамматической проверки, перефразирования и перевода текста, что открывает новые возможности для повышения качества подготовки резюме.Кроме того, внедрение таких технологий способствует экономии времени и ресурсов, особенно в корпоративной среде, где обрабатываются большие объёмы документов одновременно. В перспективе развитие LLM и интеграция с другими системами автоматизации могут позволить создавать более персонализированные и интеллектуальные сервисы для улучшения качества работы пользователей.

I. Принципы работы LLM. Self-attention

Большие языковые модели (LLM) основаны на архитектуре трансформеров, ключевым элементом которой является механизм self-attention. Он обеспечивает определение взаимосвязей между всеми словами во входной последовательности, независимо от их удалённости друг от друга. В отличие от рекуррентных сетей, работающих последовательно, self-attention [1] позволяет обрабатывать текст параллельно, выделяя наиболее значимые фрагменты для текущего контекста. Это обеспечивает высокую точность при решении задач обработки естественного языка и делает LLM универсальными для задач проверки, перевода и перефразирования текста [1,2]. Схема механизма внимания представлена на рисунке 1.

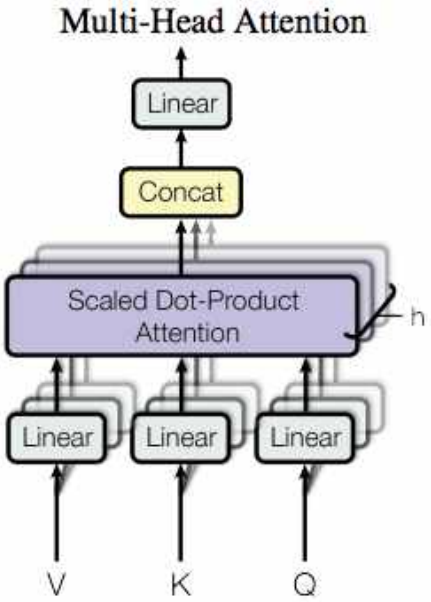


Рис. 1 – Схема механизма внимания

Суть использования данного механизма в содержании сразу нескольких экземпляров механизмов внимания, которые будут принимать входные данные, обрабатывать их, находить взаимосвязи, проставлять коэффициенты важности слов, а дальше, исходя из слов и их коэффициентов определять их важность и смысл исходного входного текста

II. Автоматизация работы с текстом

Использование LLM в задаче корректировки резюме позволяет автоматизировать ключевые этапы обработки текста [3]. К ним относятся выявление и исправление грамматических ошибок, замена стилистически неудачных конструкций, адаптация формулировок под стандарты академического или делового английского языка. Дополнительно реализуется возможность перевода текста между языками и вариантами английского,

что повышает адаптивность резюме для разных рынков труда. Кроме того дополнительной степенью автоматизации является встраивание сервиса корректировки в сервис создания резюме, либо любой другой, что позволяет пользователю не отрываться в другие места для корректировки текста, а работать в одном удобном месте.

III. Пакетная обработка и масштабируемость

Сервис поддерживает не только работу с отдельными предложениями, но и пакетную обработку больших текстовых массивов [4]. Это особенно важно для корпоративных пользователей и кадровых агентств, где требуется обработка множества документов. Масштабируемость достигается за счёт модульной архитектуры и оптимизации вызовов языковых моделей, что снижает время отклика и нагрузку на вычислительные ресурсы.

Кроме того, отдельным ключевым элементом масштабируемости является добавление любой другой языковой модели, в зависимости от потребностей клиента. Такие изменения не будут ломать система, а наоборот смогут плавно в неё встроиться.

IV. Обеспечение конфиденциальности данных

Одним из приоритетов при разработке системы является защита персональной информации. Использование локально развернутых моделей исключает необходимость передачи данных во внешние облачные сервисы, что минимизирует риск утечки конфиденциальных сведений [5]. Дополнительно предусмотрена возможность дообучения модели на закрытых корпоративных датасетах, что повышает качество работы и учитывает специфику конкретных организаций.

V. Скорость работы системы

Производительность разработанного сервиса была протестирована на GPU NVIDIA Tesla T4 с 15 ГБ видеопамяти. Тестирование показало: время обработки предложений варьируется от 0,2 до 1,4 секунды в зависимости от длины текста и сложности структуры. Средняя длина предложений составила 60 символов, при этом среднее время обработки одного предложения на стенде с использованием CUDA составило примерно 0,65 секунды. Результаты тестирования представлены на рисунке 2.

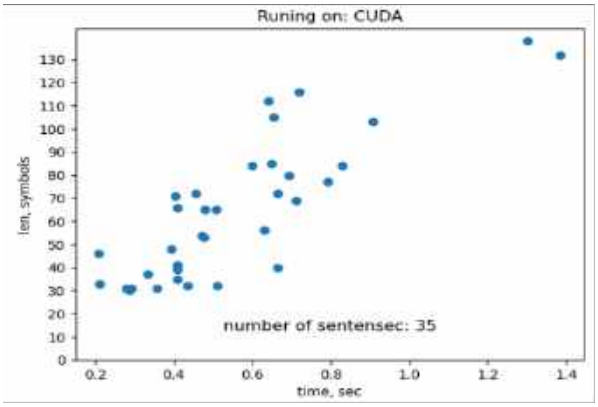


Рис. 2 – Скорость работы системы

VI. Интерфейс пользователя и удобство работы с сервисом

Сервис не имеет интерфейса в принципе, однако благодарю этому, микросервисной архитектуре и открытому API он может быть интегрирован в абсолютно любое корпоративное (частное) приложение [6], а также в корпоративную инфрастуктуру – это увеличивает возможности по переиспользованию сервиса, и его интеграции в экосистему компании. Кроме того, продукт может быть использован отдельно, как облачный сервис [7].

VII. Практическая значимость и перспективы

Предложенный сервис может использоваться индивидуальными соискателями, IT-фирмами, образовательными учреждениями и др.. Его внедрение способствует сокращению времени подготовки резюме, повышению их качества и конкурентоспособности на международном рынке труда. В перспективе возможно расширение языковой базы, интеграция с платформами для трудоустройства и адаптация под другие виды документов (мотивационные письма, академические статьи).

1. Ротман, Д. *Transformers for Natural Language Processing* / D. Rothman. – Packt Publishing, 2021.
2. Голдберг, Й. *Нейросетевые методы в обработке естественного языка* / Й. Голдберг. – О'Рейли, 2017.
3. Батура, Т. В. *Математическая лингвистика и автоматическая обработка текстов* / Т. В. Батура. – РИЦ НГУ, 2016.
4. Рашка, С. *Строим LLM с нуля* / С. Рашка. – Питер, 2025.
5. Акст, Р. *Языковые модели и мир будущего, или Путеводитель по ChatGPT* / Р. Акст. – ЛитРес, 2023.
6. Гоф, Дж., Брайант, Д., Оберн, М. *Проектирование архитектуры API* / Дж. Гоф, Д. Брайант, М. Оберн // 2024.
7. Одиноченко, М. И. Использование облачных сервисов для решения задач, связанных с применением технологии BIG DATA / М. И. Одиноченко, С. Н. Нестеренков // BIG DATA and Advanced Analytics = BIG DATA и анализ высокого уровня : сб. научных статей VIII Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 2022 год) : / редкол. : В. А. Богуш [и др.]. – Минск : Бестпринт, 2022. – С. 252-256.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ ДООБУЧЕНИЯ LLM В УСЛОВИЯХ ОГРАНИЧЕННОСТИ ДЛЯ ДАННЫХ СПЕЦИАЛИЗИРОВАННЫХ ОБЛАСТЕЙ.

Нестерович Г. В., Фурсанов С. А.
Центр информатизации и инновационных разработок, факультет компьютерных систем и сетей,
кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {g.nesterovich, s.fursanov}@bsuir.by

В работе рассматриваются современные подходы к дообучению больших языковых моделей в условиях ограниченности данных в специализированных предметных областях. Проанализированы четыре метода: полное дообучение, параметр-эффективные техники, инструкционное дообучение и Retrieval-Augmented Generation. Для каждого подхода описаны принципы работы, выявлены преимущества и недостатки, а также показаны риски и ограничения, возникающие при дефиците размеченных данных.

Введение

Современные большие языковые модели (LLM) демонстрируют высокую универсальность и способность решать широкий спектр задач, однако при переносе на специализированные области возникает необходимость их дообучения. Проблема осложняется тем, что доступ к большим размеченным корпусам данных в медицине, праве, инженерии и других областях может быть ограничен по ряду причин: высокая стоимость разметки, наличие юридических и этических ограничений, конфиденциальность информации.

В таких условиях актуален вопрос выбора методов, позволяющих эффективно адаптировать модель при минимальных объёмах данных и ограниченных вычислительных ресурсах. Сравнительный анализ различных подходов к дообучению позволяет выявить их сильные и слабые стороны и предложить оптимальные стратегии для практического применения.

I. Полное дообучение

Полное дообучение [1] подразумевает оптимизацию всех параметров предобученной модели на целевом наборе данных с задачами домена. Преимущество этого подхода состоит в максимальной выразительной мощности: модель может глубоко перестроить внутренние представления, что позволяет уловить сложные, нелинейные зависимости и тонкие семантические отличия, характерные для узкой предметной области. В случаях, когда доступны значительные по объёму и качеству экспертные аннотации, полное дообучение часто обеспечивает лучшие результаты по метрикам точности, полноты и соответствия форматам вывода. С другой стороны, при ограниченном объёме данных полный подход демонстрирует несколько серьёзных недостатков. Во-первых, высокой вероятностью выступает переобучение на небольшом и шумном наборе примеров, что ведёт к ухудшению обобщающей способности. Во-вторых, дорогостоящая вычислительная и пространственная нагрузка затрудняет экспериментальную итерацию и масштабирование в продуктивных системах: для каждой доменной версии требуется поддерживать либо отдельную копию модели, либо сложные механизмы переключения слоёв. В-третьих, полное дообучение сопряжено с риском катастрофического забывания общих навыков, что требует применения методов смешивания данных предобучения и доменных примеров, регуляризации, дистилляции или техник непрерывного обучения. Таким образом, полный подход оправдан в тех редких сценариях, когда есть обширные и качественные размеченные наборы и доступны значительные вычислительные ресурсы, но в условиях дефицита данных он требует дополнительных мер по защите от переобучения и высоких затрат.

II. Параметр-эффективное дообучение

Параметр-эффективные методы (PEFT) [2] предлагают обучать небольшой набор дополнительных параметров или низкоранговых поправок, в то время как основная модель остаётся преимущественно замороженной. К типичным техникам относятся адаптеры, LoRA (low-rank adaptation), префикс-тюнинг и промт-тюнинг. Практическая ценность этих подходов заключается в значительном сокращении требуемой памяти и вычислительных затрат: для каждой предметной области достаточно хранить компактный пакет изменений вместо полной копии огромной модели, что облегчает развертывание и управление множеством доменных версий. При малых выборках PEFT демонстрируют повышенную устойчивость к переобучению, так как число обучаемых параметров невелико и риск подгонки под шум снижен. Эти методы хорошо сочетаются с техникой мультизадачного обучения и с передачей адаптеров между смежными областями, облегчают быстрое прототипирование и инкрементальные обновления. В то же время у PEFT есть и ограничения: из-за ограниченного числа новых

параметров теоретически может существовать потолок в том, какие представления можно эффективно скорректировать, и при задачах, требующих фундаментальной перестройки внутренних слоёв, маленькие адаптеры могут оказаться недостаточными. Дополнительно требуется аккуратный выбор мест внедрения адаптеров, ранга в LoRA и гиперпараметров обучения, на что может потребоваться эмпирическая валидация. Наконец, при агрегации множества адаптеров для сложных рабочих процессов возможно появление артефактов согласованности, которые надо отслеживать через метрики и человеческую оценку.

III. Инструкционное дообучение

Инструкционное дообучение [3] направлено на улучшение способности модели правильно интерпретировать и выполнять инструкции пользователя путём обучения на наборах пар «инструкция – желаемый ответ». Для специализированных доменов это особенно полезно, поскольку часто важнее соблюдение регламентов, форматов и последовательности рассуждений, чем просто воспроизведение фактов. При ограниченном количестве примеров аккуратная разработка инструкций, разнообразие формулировок и включение демонстраций форматов ответа могут существенно повысить полезность модели в рабочих сценариях: она начинает лучше соблюдать шаблоны, выдавать ответы требуемой структуры и избегать частых ошибок формата. Положительная сторона инструкционного подхода также в том, что он позволяет быстро корректировать поведение без глубокой перестройки весов модели, что экономит ресурсы и упрощает контроль качества. Минусы включают зависимость от качества самих инструкций: плохо сформулированные или синтетические инструкции способны ввести систематические смещения в поведение, а также не всегда обеспечивают добавление новых фактов или терминов в модельную базу знаний. Инструкционное дообучение часто лучше комбинировать с небольшой выборкой экспертных аннотаций и с элементами человеческой ревизии, чтобы сохранять корректность и избегать индуцированных ошибок.

IV. Retrieval-Augmented Generation

Retrieval-Augmented Generation (RAG) [4] сочетает генеративную модель и модуль извлечения релевантных документов из внешней индексируемой базы знаний, подставляя найденный контекст в качестве дополнительного входа при генерации. Для узких доменов с ограниченной разметкой это крайне практичный путь: вместо того чтобы пытаться «вложить» все факты в веса модели, хранить и поддерживать внешнюю базу, которую можно пополнять неразмеченными источниками, документами и справочниками. Преимущества RAG очевидны: оперативность обновления знаний без дорогостоящей переобучаемости, повышение фактологичности ответов при правильной настройке извлечения и возможность предоставления пользователю ссылок на источники и доказательств. Тем не менее эффективность RAG сильно зависит от качества индексирования, алгоритмов поиска и ранжирования, а также от полноты корпуса: если база неполна или содержит шум, генерация может ссылаться на нерелевантные или противоречивые фрагменты. Дополнительные вызовы включают необходимость верификации и агрегации противоречивых источников, обработку устаревших данных и обеспечение объяснимости ответов. В практических системах RAG часто комбинируют с локальными тонкими дообучениями или контролем качества на уровне ранжирования, применяют контрольно-верификационные пайплайны и меры по калибровке доверия.

Заключение

Сравнение методов показывает, что при ограниченности данных оптимальные решения, как правило, гибридные и зависят от множества факторов: объёма и типа доступных текстов, наличия экспертной разметки, вычислительных и организационных ресурсов, требований к обновляемости знаний и к объяснимости. Полное дообучение остаётся мощным инструментом там, где доступны крупные качественные датасеты и ресурсы для проведения тщательной регуляризации и контроля, однако оно дорого и подвержено переобучению при малых выборках. Параметр-эффективные методы предлагают прагматичный компромисс, снижая затраты на хранение и развертывание и обеспечивая устойчивость при ограниченных примерах, при этом имея свои потолочные ограничения. Инструкционное дообучение повышает управляемость поведения модели и пригодность к форматированным задачам, но требует качественных инструкций и валидации. RAG даёт путь к обновляемости и фактической проверяемости, но требует аккуратной инженерии корпуса и поисковых компонентов.

1. Hands-On Large Language Models / J. Alammar, M. Grootendorst. // O'Reilly –2024
2. Parameter-efficient fine-tuning of large-scale pre-trained language models / N. Ding et al. // Nature Machine Intelligence – 2023 – Vol. 5, P. 220–235
3. The Ultimate Guide to Fine-Tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities / V. Parthasarathy et al. // CeADAR Connect Group – 2024
4. A Simple Guide to Retrieval Augmented Generation / A. Kimothi // Manning – 2025

ПРИМЕНЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РАМКАХ БОРЬБЫ С ДЕЗИНФОРМАЦИЕЙ: СОВРЕМЕННЫЕ МЕТОДЫ И РЕШЕНИЯ

Орлов Д. В., Коростелёв А. Д., Марков А. Н.
Центр информатизации и инновационных разработок, кафедра информатики,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {d.orlov, a.korostelev, a.n.markov}@bsuir.by

Данный доклад посвящен актуальной проблеме распространения дезинформации в цифровом пространстве и анализу роли искусственного интеллекта как инструмента её создания и противодействия. Рассматриваются современные технологии генерации синтетического контента с использованием нейронных сетей, подходы к его выявлению и автоматизированные механизмы противодействия. Отдельное внимание уделено этическим и правовым аспектам использования подобных технологий.

Введение

Цифровые технологии играют важную роль в формировании современного общества, делая социальные сети, новостные платформы и мессенджеры ключевыми каналами распространения информации. Вместе с ростом доступности информации обострилась проблема её достоверности: фейковые новости, манипулятивные тексты, синтетические мультимедийные материалы становятся всё более массовыми [1].

Искусственный интеллект играет в этой сфере двойственную роль. С одной стороны, генеративные модели позволяют создавать контент, который сложно отличить от настоящего, что облегчает распространение ложной информации и манипуляцию общественным мнением. С другой стороны, современные методы анализа текста, изображений, аудио и сетевых взаимодействий предоставляют возможности для автоматического выявления и фильтрации ложного контента.

Таким образом, перед исследователями и разработчиками стоит комплексная задача: с одной стороны, изучить и понять механизмы создания дезинформации с помощью ИИ, с другой – разработать эффективные методы её обнаружения и контрмеры, учитывая технологические, этические и правовые аспекты.

I. Генерация дезинформации с помощью ИИ

Современные генеративные модели позволяют создавать фейковые тексты, изображения, видео и аудио с высокой степенью правдоподобия. Большие языковые модели (LLM) генерируют тексты, имитирующие стиль журналистских статей, научных публикаций или социальных постов, и могут содержать манипулятивные аргументы, что делает их эффективными в распространении недостоверной информации. Генеративные нейросети (GAN, diffusion-модели) создают реалистичные изображения и видеоролики, включая дипфейки, где лица людей и их мимика выглядят естественно и практически неотличимы от настоящих [1].

Технологии синтеза голоса позволяют имитировать интонацию, тембр и стиль речи конкретного человека, что используется для подделки интервью или голосовых сообщений. Доступность облачных сервисов и open-source библиотек снижает порог входа для злоумышленников: создание и распространение синтетического контента стало возможным даже для непрофессионалов. При этом генеративные модели могут комбинировать текст, изображение и аудио, создавая мультимедийные фейки, которые сложно обнаружить традиционными методами. Масштабирование таких систем позволяет быстро распространять дезинформацию, что усиливает её влияние на общественное мнение и затрудняет своевременное выявление.

II. Методы обнаружения дезинформации

Одним из ключевых направлений является использование NLP-подходов (Natural Language PRocessing) для анализа текстового контента. Модели обработки естественного языка классифицируют тексты на категории «правда» или «ложь», выявляют стилистические аномалии и сопоставляют утверждения с источниками. Современные трансформеры, такие как BERT (Bidirectional Encoder Representations from Transformers) и RoBERTa (Robustly Optimized BERT Approach), позволяют извлекать ключевые утверждения и проводить их верификацию с помощью базы фактов. Основное ограничение NLP-подходов заключается в сложности интерпретации сарказма, цитат или контекста, что может приводить к ошибочным классификациям.

Другой эффективный метод – использование графовых моделей и анализ сетевой структуры распространения информации. Анализ сетей помогает обнаруживать закономерности координированных действий, активность ботов и подозрительные группы распространения информации [2]. Графовые нейронные сети помогают

обнаружать аномалии в связях между пользователями и публикациями, всплески репостов и аномальные маршруты контента. Основные проблемы применения графовых моделей включают необходимость доступа к сетевым данным платформ и необходимость учитывать естественные вирусные эффекты, чтобы не путать их с координированной дезинформацией.

Направление визуального анализа фокусируется на выявлении дипфейков и синтетических изображений. Сверточные и временные нейросети ищут артефакты сжатия, несоответствия мимики, освещения и аномалии в частотной области. Мультимодальные проверки, которые сравнивают видео и аудио, повышают точность выявления синтетических материалов. Ограничение этого подхода в том, что генеративные модели становятся всё более реалистичными, что снижает эффективность традиционных признаков детекции [3].

На практике наиболее эффективными являются комплексные системы, которые объединяют текстовый, визуальный и сетевой анализ с проверкой метаданных. Наборы моделей и методы cross-checking позволяют снизить количество ложных срабатываний, повышая точность детекции. Основной недостаток таких систем – высокая сложность реализации и значительные вычислительные ресурсы, необходимые для обработки больших объёмов данных.

III. Контрмеры

Одним из методов противодействия синтетическому контенту является внедрение водяных знаков (watermarking). С помощью невидимых или видимых меток можно отслеживать источник сгенерированного ИИ-контента и идентифицировать его происхождение. Применяются криптографические методы и стеганографические подходы, которые делают метки устойчивыми к базовым изменениям файла. Основное ограничение водяных знаков заключается в том, что они могут быть удалены при конвертации, сжатии или повторном распространении контента [4].

Хранение информации о происхождении контента и его цепочке изменений (provenance) является ещё одним важным инструментом. Метаданные позволяют установить подлинность файла, отслеживать источники и выявлять подделки. Технологии цифровой подписи и хэширования уменьшают возможность фальсификации, однако часто метаданные теряются при скачивании, пересылке или преобразовании контента, что ограничивает их практическую эффективность.

Полуавтоматические системы с участием человека (human-in-the-loop) сочетают автоматическую фильтрацию с экспертной проверкой. Алгоритмы предварительно анализируют контент и выделяют подозрительные материалы, после чего эксперт принимает окончательное решение. Такой подход снижает риск ложных блокировок, улучшает обучение моделей за счёт обратной связи и повышает точность дикции. Главный недостаток – плохая масштабируемость при больших объёмах данных и высокая потребность в человеческих ресурсах [2].

Регуляторы и крупные платформы внедряют правила маркировки синтетического контента, а также программы прозрачности и fact-checking, чтобы систематизировать борьбу с дезинформацией. Такие инициативы помогают выстраивать процессы обнаружения и реагирования на фейки, однако стандарты и подходы сильно различаются между странами и компаниями. Это создает фрагментированную картину регулирования и требует согласования технических и правовых норм для повышения эффективности [4].

IV. Этические вопросы

Автоматические системы детекции несут риск ложных срабатываний и цензуры. Избыточная автоматизация может подавлять свободу слова и использоваться для контроля информации. Необходимы прозрачность алгоритмов, процедуры апелляции и участие независимых аудиторов для минимизации побочных эффектов.

V. Выводы

ИИ не только способствует созданию дезинформации, но и предоставляет инструменты для её обнаружения и борьбы с ней. Перспективным является развитие гибридных систем, совмещающих машинное обучение, watermarking и участие человека. Эффективная борьба с дезинформацией требует внедрения технологических решений, правового регулирования и повышения цифровой грамотности пользователей.

VI. Список литературы

1. Баяк, Е. И. Применение Big Data для анализа социальных сетей / Е. И. Баяк, С. Н. Нестеренков, Д. А. Жалейко // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. IX Междунар. науч.-практ. конф. В 2 ч. Ч. 1 (Республика Беларусь, Минск, 17–18 мая 2023 года) / редкол. : В. А. Богуш [и др.]. – Минск : БГУИР, 2023. – С. 348–350.
2. Половой, А. А. Когнитивные аспекты взаимодействия человека с нейросетевыми системами / А. А. Половой, С. Н. Нестеренков // Информатика : сборник трудов международной молодежной школы «Инженерия – XXI» (Российская Федерация, г. Новороссийск, 15-18 апреля 2025 г.) / редкол : И. В. Чистяков [и др.]. – Новороссийск : БГТУ им. В. Г. Шухова, 2025. – С. 224-226.
3. Mirsky, Y., Lee, W. The Creation and Detection of Deepfakes: A Survey // ACM Computing Surveys. – 2021. – Vol. 54, № 1. – P. 1-41
4. European Commission. Code of Practice on Disinformation [Electronic resource] / European Commission. – Brussels, 2023. – Mode of access: `https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation` . – Date of access: 26.09.2025.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ЦЕН НА РЫНКЕ НЕДВИЖИМОСТИ РЕСПУБЛИКИ БЕЛАРУСЬ

Раптунович Д. И.
Инженерно-экономический факультет, Бизнес-аналитика и цифровой маркетинг,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: i@zmitroc.by

В работе представлен прикладной сравнительный анализ трёх моделей прогнозирования цен на рынке жилой недвижимости Республики Беларусь: множественной линейной регрессии, регрессионного дерева решений и градиентного бустинга. Единая методика включает унифицированный набор признаков, одинаковый протокол валидации и сопоставимые метрики качества (MAE, RMSE, R^2). Показано, что ансамблевые методы обеспечивают наилучший баланс точности и устойчивости при допустимой интерпретируемости факторов. Результаты предназначены для интеграции в интеллектуальную систему поддержки решений.

Введение

Моделирование цен на рынке недвижимости традиционно опирается на подход гедонического ценообразования, предложенный Розеном [1] и развитый в последующих эмпирических исследованиях [2–4]. Однако современные тенденции показывают, что линейные зависимости не всегда адекватно отражают структуру рынка, что требует применения методов машинного обучения и интеллектуального анализа данных [5–7]. Целью данной работы является проведение сравнительного анализа трёх моделей прогнозирования: множественной линейной регрессии, дерева решений и градиентного бустинга, применительно к рынку жилья Республики Беларусь.

I. Постановка задачи и данные

Анализ выполнялся на основе совокупности данных по объектам жилой недвижимости в крупных городах Республики Беларусь (Минск, Брест, Витебск, Гомель, Гродно, Могилёв). Источником информации выступил портал gohome.by [15], предоставляющий архив объявлений за период 2018–2025 гг., а также официальные макроэкономические данные Национального банка и Белстата [13, 14].

В выборку включены объекты вторичного рынка, для которых указаны следующие характеристики:

- местоположение (город, район);
- площадь квартиры, количество комнат и этаж;
- год постройки;
- цена за квадратный метр в белорусских рублях;
- макроэкономические параметры на дату публикации (ставка рефинансирования, курс доллара, средняя зарплата по региону).

Целью анализа является построение и сравнение трёх математических моделей прогнозирования цен предложения на основе идентичного набора признаков. Формально задача представляется как регрессионная:

$$\hat{y}_i = f(X_i) + \varepsilon_i,$$

где $\hat{y}_i$ – прогнозируемая цена, X_i – вектор признаков объекта, ε_i – случайная ошибка.

Перед обучением модели данные прошли процедуру очистки и нормализации. Аномальные значения были удалены, категориальные признаки преобразованы методом one–hot–кодирования, числовые – стандартизированы. Для оценки обобщающей способности использовалась перекрёстная проверка ($k = 5$), обеспечивающая устойчивость результатов при изменении выборки.

Построение репрезентативной базы данных и использование макроэкономических индикаторов позволило учесть влияние факторов инфляции, процентных ставок и региональных различий [10, 11]. Это обеспечивает практическую применимость моделей в рамках анализа реального рынка Беларуси.

II. Модели и протокол обучения

В исследовании сравнивались три модели различной сложности и интерпретируемости. **Модель 1 (M1):** множественная линейная регрессия – классический эконометрический подход, основанный на предположении о линейной зависимости между признаками и ценой [1, 2]. Она служит базовым эталоном для оценки относительной эффективности более сложных алгоритмов.

Модель 2 (M2): дерево решений, реализующее иерархическую сегментацию пространства признаков. Данный метод способен выявлять пороговые и нелинейные зависимости без априорных ограничений на форму функции $f(X)$ [6]. Ограничением является склонность к переобучению при малом числе объектов или избыточной глубине дерева.

Модель 3 (M3): градиентный бустинг – ансамблевый метод, объединяющий множество слабых моделей (деревьев решений) для минимизации функции потерь. Использованы реализации XGBoost и LightGBM, зарекомендовавшие себя в задачах прогнозирования цен и рейтинговых оценок [8, 9].

Процесс обучения включал оптимизацию гиперпараметров методом перебора по сетке (Grid Search) с использованием перекрёстной проверки. В качестве метрик использованы:

1. **MAE (Mean Absolute Error)** – средняя абсолютная ошибка прогноза;
2. **RMSE (Root Mean Squared Error)** – среднеквадратичная ошибка;
3. R^2 – коэффициент детерминации, характеризующий долю объяснённой дисперсии.

Для анализа важности признаков применены SHAP–значения [8], позволяющие интерпретировать влияние отдельных факторов на результат прогноза. Выявлено, что ключевыми переменными, определяющими уровень цен, являются: регион, площадь, тип жилья, курс доллара и средняя заработная плата.

Проведённый протокол обеспечивает сопоставимость моделей и позволяет объективно оценить преимущества алгоритмов машинного обучения по сравнению с классическими эконометрическими методами. Это соответствует современным тенденциям интеграции искусственного интеллекта в задачи рыночного анализа [11, 12].

III. Результаты и анализ

Результаты экспериментов показывают, что M3 обеспечивает наилучшую точность прогноза (снижение RMSE на 30–40 % по сравнению с M1), при этом сохраняет устойчивость при изменениях выборки [7, 9]. Множественная регрессия остаётся интерпретируемой, но не улавливает нелинейных эффектов, что ранее отмечалось в эконометрических исследованиях рынка недвижимости [3, 4]. Для анализа вклада факторов использованы SHAP–значения [8], выявившие ключевые признаки: регион, тип жилья, площадь, курс валюты и среднюю заработную плату.

IV. Выводы и практическая значимость

Алгоритмы ансамблевого типа демонстрируют наилучший баланс между точностью и устойчивостью. Результаты подтверждают применимость современных методов машинного обучения к задачам прогнозирования цен в условиях белорусского рынка [10–12]. Разработанные модели могут быть интегрированы в интеллектуальные системы анализа и поддержки принятия решений в сфере недвижимости.

V. Список литературы

1. Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition / S. Rosen // Journal of Political Economy. – 1974. – Vol. 82, No. 1. – P. 34–55.
2. Malpezzi, S. Hedonic Pricing Models: A Selective and Applied Review / S. Malpezzi // Housing Economics and Public Policy. – Oxford: Blackwell Publishing, 2003. – P. 67–89.
3. Kok, N., Jennen, M., Sloof, R. The Impact of Energy Labels and Occupancy on Office Building Prices / N. Kok, M. Jennen, R. Sloof // Energy Policy. – 2012. – Vol. 47. – P. 489–497.
4. Shiller, R. Irrational Exuberance. – Princeton: Princeton University Press, 2015. – 392 p.
5. Wang, F., Li, M. Forecasting Housing Prices with Machine Learning Algorithms: Evidence from China / F. Wang, M. Li // Applied Economics. – 2021. – Vol. 53(10). – P. 1173–1186.
6. Li, X., Song, Y., Liu, H. A Comparison of Machine Learning Algorithms for Predicting Housing Prices in Urban China / X. Li [et al.] // Computers, Environment and Urban Systems. – 2020. – Vol. 83. – 101514.
7. Zhou, C., Luo, Y., Zhang, Y. Real Estate Price Prediction Using Machine Learning Models / C. Zhou [et al.] // Journal of Advanced Computational Intelligence and Intelligent Informatics. – 2019. – Vol. 23, No. 5. – P. 791–798.
8. Hastie, T., Tibshirani, R., Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. – 2nd ed. – New York: Springer, 2009. – 745 p.
9. Park, S., Bae, S. Application of Random Forest and Gradient Boosting in Real Estate Price Prediction / S. Park, S. Bae // Expert Systems with Applications. – 2021. – Vol. 165. – P. 113923.
10. Glaeser, E., Nathanson, C., Spiegel, O. Housing Market Dynamics and Artificial Intelligence / E. Glaeser [et al.] // NBER Working Paper Series. – 2020. – No. 27658. – 28 p.
11. Russell, S., Norvig, P. Artificial Intelligence: A Modern Approach. – 4th ed. – Pearson, 2021. – 1162 p.
12. Goodfellow, I., Bengio, Y., Courville, A. Deep Learning. – Cambridge, MA: MIT Press, 2016. – 775 p.
13. Национальный банк Республики Беларусь. Основные показатели денежно–кредитной политики и ставки рефинансирования (2010–2025) [Электронный ресурс]. – Режим доступа: `https://www.nbrb.by/statistics/monetarypolicy`. – Дата доступа: 20.10.2025.
14. Белстат. Средняя заработная плата и динамика цен на жильё по регионам Республики Беларусь (2015–2025) [Электронный ресурс]. – Режим доступа: `https://www.belstat.gov.by`. – Дата доступа: 20.10.2025.
15. Портал gohome.by. Рынок недвижимости Беларуси: аналитика и тенденции (2018–2025) [Электронный ресурс]. – Режим доступа: `https://www.gohome.by`. – Дата доступа: 20.10.2025.

АВТОМАТИЗАЦИЯ РАБОТЫ С БАЗАМИ ЗНАНИЙ НА ОСНОВЕ ГИБРИДНОГО ПОДХОДА

Сальников Д. А., Ерофеев А. С.
Кафедра интеллектуальных информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: d.salnikov@bsuir.by, artem.erofeev2703@gmail.com

В работе рассмотрен гибридный подход, сочетающий в себе использование больших языковых моделей, классических RAG-систем и технологию OSTIS для автоматизации работы с базами знаний.

I. Введение

В контексте возрастающего применения больших языковых моделей актуальной задачей является изучение их фундаментальных ограничений, ключевыми из которых выступают галлюцинации, накопление ошибок и низкая управляемость. Галлюцинации, проявляющиеся в генерации правдоподобной, но фактически недостоверной информации, обусловлены статистической природой моделей и отсутствием механизмов верификации. Это зачастую приводит к кумулятивному эффекту накопления ошибок, когда незначительная неточность нарастает в авторегрессионном процессе, нарушая контекстную согласованность текста. Указанные проблемы, в свою очередь, тесно связаны с низкой управляемостью языковых моделей, то есть трудностью точного контроля содержания и стиля, эффективность которого критически зависит от качества пользовательских запросов и специализированной настройки. Комплексное преодоление этих взаимосвязанных проблем определяет возможность надежного применения больших языковых моделей в ответственных сценариях.

II. Проблематика и предлагаемый подход

Современные большие языковые модели, такие как DeepSeek, Grok и ChatGPT, часто оснащаются функцией веб-поиска, позволяющей компенсировать ограничения тренировочных данных – их неполноту или неактуальность. Однако, данный подход имеет существенные недостатки. Во-первых, достоверность источников в сети не гарантирована: интернет содержит значительный объем недостоверной информации, включая тексты, сгенерированные моделями, упомянутыми ранее. Во-вторых, модель может неточно сформулировать поисковый запрос или выбрать нерелевантные ключевые слова. В-третьих, самостоятельный анализ и синтез информации из множества источников требует значительных вычислительных ресурсов, особенно для моделей с большим числом параметров. Эффективным подходом к смягчению данной проблемы является использование архитектуры RAG, которая ограничивает генерацию ответа строго релевантными документами, извлечёнными из внешнего источника знаний. В такой парадигме модель выполняет задачи синтеза и формулирования ответа на основе предоставленного контекста, что позволяет существенно снизить уровень галлюцинаций. Важно отметить, что применение RAG не требует использования крупных моделей и сохранения истории диалога, что открывает возможности для развёртывания более эффективных и компактных систем, демонстрирующих сопоставимую или повышенную точность в задачах, требующих фактической достоверности. Однако и этот метод имеет ограничение, связанное с приблизительным характером векторного поиска и отсутствием гарантий точности при извлечении информации [1].

Предлагаемый подход является гибридным, комбинирует преимущества классических RAG-систем, больших языковых моделей и технологии OSTIS. В архитектуре классической RAG-системы векторная база данных была заменена SC-машиной – эмулятором семантического компьютера, системой управления графовыми базами знаний. За обработку текста и запросов пользователя отвечают семантический фрагментатор текста и локальная большая языковая модель с малым количеством параметров (Gemma от Google), взаимодействие с базой знаний осуществляется посредством API SC-машины, сам пользователь взаимодействует с системой посредством графического интерфейса. Приведенная модель комбинирует преимущества вышеописанных подходов и стремится исправить их недостатки. Большая языковая модель «понимает», что пользователь имеет в виду и о чем говорится в представленном ей фрагменте текста, система автоматически составляет базу знаний на основе произвольного текстового документа без предварительной настройки, генерирует ответ для пользователя на естественном языке. База знаний хранит контекст с учетом семантики, предоставляет инструментарий для его поиска. Возможно внедрение различных команд для управления системой [2].

III. Программная реализация

В рамках реализации прототипа системы ключевыми стали следующие пункты:

- семантическое разбиение произвольных текстовых документов на более мелкие куски (фрагменты);
- автоматическая обработка фрагментов: извлечение тем, конкретных и абстрактных сущностей, процессов;
- автоматическое построение онтологии на основе ранее извлечённых из фрагментов данных с помощью API SC-машины;
- поиск релевантных фрагментов в базе знаний с помощью API SC-машины;
- приём и обработка запроса пользователя на естественном языке;
- генерация ответа с учётом семантики запроса и релевантных фрагментов на естественном языке.

Стоит отметить, что для разбиения текста на фрагменты использована библиотека semchunk, основанная на эффективном и высокоточном алгоритме фрагментации. Она создает фрагменты, которые более семантически значимы, чем обычные токены и рекурсивные фрагменты символов, такие как RecursiveCharacterTextSplitter от LangChain, а также на 85% быстрее, чем его ближайшая альтернатива, semantic-text-splitter. Также, составлена инструкция для большой языковой модели. В ней от модели требовалось всегда отвечать в строгом формате JSON, выделять из представленных фрагментов текста темы, конкретные сущности (имена собственные, научные термины, определения и так далее), абстрактные сущности (метафоры и нематериальные сущности), процессы [3].

IV. Заключение

В результате тестирования по методу чёрного ящика, продемонстрированного на рисунке 1, полная обработка текста длиной 15 тысяч символов заняла около 11 минут, длиной 5 тысяч символов — 2–3 минуты. При локальной работе как SC-машины, так и большой языковой модели на непроизводительном ноутбуке это является удовлетворительным результатом. Точность ответа зависит от формулировки вопроса, может сыграть существенную роль фактор случайности. Это общеизвестные проблемы, присущие большим языковым моделям. Чтобы попытаться улучшить результат, можно использовать большие языковые модели с большим количеством параметров и более производительное аппаратное обеспечение.

V. Список литературы

1. Trust, attitudes and use of artificial intelligence: A global study 2025 [Electronic resource]. – Mode of access: `https://kpmg.com/xx/en/our-insights/ai-and-technology/trust-attitudes-and-use-of-ai.html`. – Date of access: 12.10.2025.
2. GitHub – ostis-ai/sc-machine: Software implementation of semantic memory and its APIs [Electronic resource]. – Mode of access: `https://github.com/ostis-ai/sc-machine`. – Date of access: 13.10.2025.
3. semchunk · PyPI [Electronic resource]. – Mode of access: `https://pypi.org/project/semchunk/`. – Date of access: 14.10.2025.

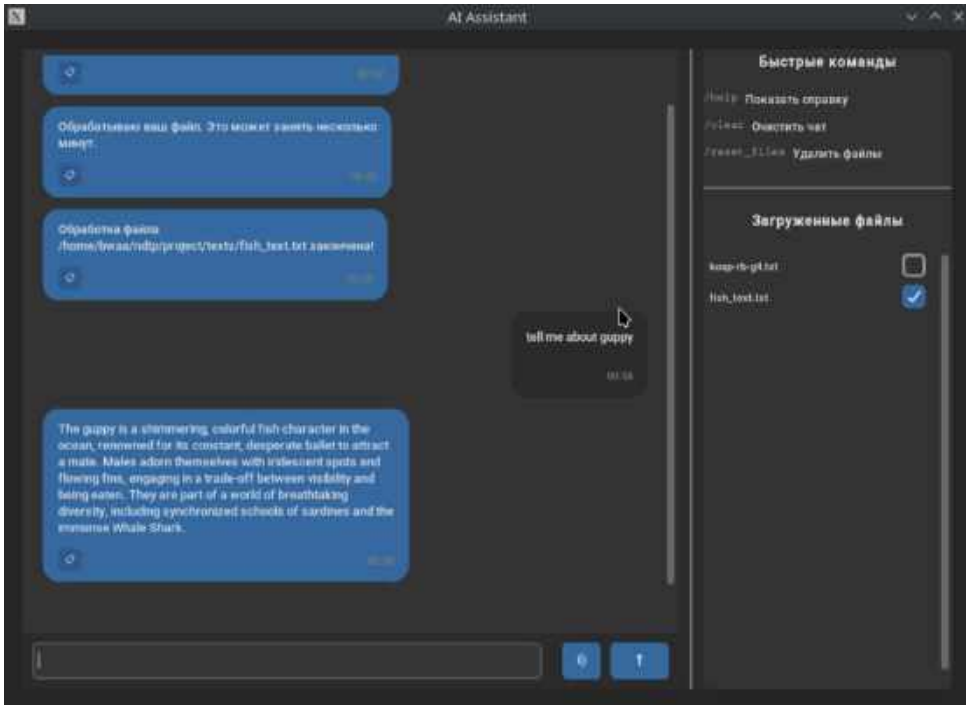


Рис. 1 – Тестирование системы. Обработка загруженного файла и ответ на пользовательский запрос.

АСПЕКТЫ ЗАЩИТЫ ПЕРСОНАЛЬНЫХ ДАННЫХ В ИНТЕЛЛЕКТУАЛЬНЫХ МЕДИЦИНСКИХ СИСТЕМАХ

Сальников Д. А., Крищенович В. А.
Кафедра интеллектуальных информационных технологий
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {d.salnikov, krish}@bsuir.by,

В работе рассмотрены ключевые аспекты обеспечения безопасности и конфиденциальности персональных данных при использовании интеллектуальных медицинских систем в Республике Беларусь. За основу предлагается взять многоуровневый подход, объединяющий пять ключевых компонентов, сочетающий правовые, технические, организационные и этические требования, который позволит легитимно и безопасно использовать искусственный интеллект в учреждениях здравоохранения Беларуси, раскрывая его потенциал при гарантии защиты прав пациентов. Также подчеркнута особая чувствительность медицинских данных и потенциальные риски их несанкционированного раскрытия.

I. Введение

Современная медицина переживает революцию, связанную с активным внедрением интеллектуальных систем и технологий искусственного интеллекта. Эти технологии способны анализировать огромные массивы медицинской информации, помогая в постановке диагнозов, разработке персонализированных планов лечения и прогнозировании эпидемий. Однако, стремительное распространение таких систем одновременно ведет к беспрецедентному росту объемов собираемых и обрабатываемых персональных данных о здоровье граждан. Эти данные относятся к категории особо чувствительных, поскольку раскрывают чувствительные аспекты жизни человека, и их ненадлежащая защита может привести к дискриминации, шантажу и необратимому ущербу репутации. Именно поэтому обеспечение их конфиденциальности и безопасности становится не просто технической задачей, а ключевым условием доверия между пациентом и системой здравоохранения.

II. Комплексный подход

Комплексный подход, должен гармонично объединять несколько ключевых аспектов защиты персональных данных в интеллектуальных медицинских системах, построенных с учетом законодательства Республики Беларусь, в частности Закона № 99-З от 07.05.2021 «О защите персональных данных» (далее – Закон № 99-З) и заключаются в следующих тезисах [1]:

- правовой и нормативный аспект;
- информационная безопасность;
- организационные и административные меры;
- этический аспект и права субъектов данных;
- специфика для этапов работы с средствами искусственного интеллекта.

Правовой и нормативный аспект составляет основу всей деятельности. Он требует строгого соблюдения Закона № 99-З, который является главным регулятором в этой сфере. Одновременно с этим необходимо учитывать положения смежного законодательства, например:

- Закон РБ «О здравоохранении», который также содержит нормы о врачебной тайне [2].
- Кодекс РБ об административных правонарушениях (ст. 23.16) устанавливает ответственность за нарушение порядка обработки персональных данных.
- Ключевые принципы обработки (Статья 7 Закона № 99-З).
- Правовые основания для обработки (Статья 8 Закона № 99-З). Для обработки персональных данных, в том числе в медицинских ИИ, необходимо одно из оснований, главным из которых является согласие субъекта. Для специальных категорий данных (к которым относятся данные о здоровье) согласие, как правило, является обязательным, за исключением строго определенных случаев (например, обработка в целях медицины, при условии, что данные обрабатываются медработником).
- Обработка должна иметь четкое правовое основание.
- Данные собираются для заранее определенных, заявленных и законных целей (например, постановка диагноза, научное исследование). Не допускается обработка для целей, несовместимых с первоначальными.
- Обеспечение точности, полноты и соответствия данных целям обработки.
- Не допускается обработка избыточных данных по отношению к заявленным целям.

Следующим ключевым аспектом является, информационная безопасность, обеспечиваемая алгоритмами и протоколами для защиты информации на всех этапах. Подразумевается обязательное использование шифрования как передаваемых данных, так и информации, хранящейся на серверах. Большую роль играют методы

псевдонимизации, когда прямые идентификаторы заменяются на метки, и анонимизации, то есть необратимого удаления связи данных с личностью. Не менее важен строгий контроль доступа, основанный на принципе минимальных привилегий, когда каждый пользователь получает права только к тем данным, которые необходимы для его работы. Обязательным является ведение детальных журналов учета всех действий с персональными данными.

Организационные и административные меры касаются создания правильной структуры управления данными. Оператор (медучреждение или разработчик) обязан разработать и внедрить внутренний документ – политику защиты персональных данных. В рамках этой работы назначаются ответственные лица, а все сотрудники, имеющие доступ к информации, проходят обязательное обучение. Перед внедрением новых технологий рекомендуется проводить оценку рисков для прав субъектов данных. Ключевым требованием является обеспечение локализации данных, т. е. накопление, хранения и обработки персональных данных граждан РБ должна осуществляться только на территории страны [3].

Этический аспект и права субъектов данных направлены на построение доверия. Система должна быть прозрачной, а субъекты данных – четко информированы о том, как и зачем используются их сведения. Медицинские интеллектуальные системы должны гарантировать реализацию прав гражданина, включая право на отзыв согласия и право на доступ к своим данным. Критически важным принципом является сохранение окончательного контроля за принятием решений за врачом-человеком, который проверяет и утверждает рекомендации, данные искусственным интеллектом.

III. Специфика работы с искусственным интеллектом

На каждом этапе работы с искусственным интеллектом существуют свои специфические требования. Первостепенно, необходимо получать однозначное согласие субъекта на использование его обезличенных персональных данных. Использование данных без согласия возможно лишь в исключительных случаях, прямо предусмотренных законом (например, для статистических целей).

Непосредственная обработка данных конкретного пациента в работающей системе должна обеспечивать вводимым данным для анализа такой же высокий уровень защиты, как и для основных медицинских записей, поскольку результат работы системы (диагноз, прогноз) также сам является конфиденциальным и подлежит защите.

Таким образом, легитимное использование интеллектуальных медицинских систем в Беларуси требует комплексного подхода. Этот подход объединяет неукоснительное соблюдение законодательства, внедрение передовых технических мер защиты, выстраивание четких организационных процедур и следование этическим принципам. Только так можно раскрыть потенциал технологий, гарантируя неприкосновенность частной жизни и прав пациентов.

IV. Заключение

Внедрение интеллектуальных медицинских систем в Беларуси требует строгого соблюдения Закона № 99-З «О защите персональных данных». Это влечет за собой не только техническое переоснащение, но и глубокую организационную перестройку: разработку политик, обучение персонала, реализацию прав субъектов и обеспечение локального хранения данных. Комплексный подход, объединяющий правовые, технические и организационные меры, является обязательным условием для легитимного и этичного использования методов искусственного интеллекта в здравоохранении РБ.

V. Список литературы

1. О защите персональных данных [Электронный ресурс] : Закон Республики Беларусь, 7 мая 2021 г., № 99-З // Национальный правовой Интернет-портал Республики Беларусь. – Режим доступа: `https://pravo.by/document/?guid=12551&p0=H12100099`. – Дата доступа: 17.10.2025.
2. О здравоохранении [Электронный ресурс] : Закон Республики Беларусь, 18 июня 1993 г., № 2435-XII // Национальный правовой Интернет-портал Республики Беларусь. – Режим доступа: `https://pravo.by/document/?guid=3871&p0=v19302435`. – Дата доступа: 17.10.2025.
3. О государственном информационном ресурсе «Реестр операторов персональных данных [Электронный ресурс] : приказ Оперативно-аналитического центра при Президенте Республики Беларусь от 1 июня 2022 года № 94 // Информационно-правовая система ЭТАЛОН-ONLINE – Все законодательство РБ. – Режим доступа: `https://etalonline.by/document/?regnum=t62205021`. – Дата доступа: 17.10.2025

АЛГОРИТМЫ ИНТЕЛЛЕКТУАЛЬНОГО ПОИСКА ИНФОРМАЦИИ В РАСПРЕДЕЛЕННЫХ БАЗАХ ДАННЫХ С ПРИМЕНЕНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Скалозуб К. А., Ярмош А. Д., Романюк М. В.
Отдел информационных технологий,
центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {k.skalozub, a.iarmosh, romanuk}@bsuir.by

В статье рассматриваются алгоритмы интеллектуального поиска информации в распределенных базах данных с применением методов машинного обучения. Анализируются подходы к организации поиска, включая использование методов классификации, кластеризации и ранжирования результатов с применением нейросетевых моделей. Показана роль машинного обучения в адаптивной обработке запросов, повышении точности и скорости поиска. Приводятся современные направления развития интеллектуальных алгоритмов в области информационного поиска и их применение в распределенных системах.

Введение

С ростом объемов данных и числа распределенных хранилищ становится все более актуальной задача эффективного поиска информации в распределенных системах. Традиционные методы информационного поиска (IR, Information Retrieval) часто оказываются недостаточно эффективными при масштабировании: проблемы распределения данных, сетевые задержки, неоднородность источников и высокая нагрузка усложняют задачу.

Интеграция методов машинного обучения (ML) в алгоритмы поиска позволяет повысить качество ранжирования, предсказать релевантность, адаптировать индексацию под нагрузку и динамику данных. В распределенных базах данных (например, распределенных хранилищах, облачных СУБД, мультиагентных схемах) добавляются еще следующие задачи: как обучать модели, когда данные не централизованы, как минимизировать передачу данных, как справляться с задержками, с несинхронностью и т.д.

Цель статьи – рассмотреть основные классы алгоритмов интеллектуального поиска в распределенных базах данных, показать, как методы ML интегрируются в них, выявить трудности и направления для исследований.

I. Основные подходы алгоритмов интеллектуального поиска с ML

1. Локальные модели и глобальная агрегация. На каждом узле строится локальная модель (например, классификатор, ранжировщик), которая оценивает релевантность документов по запросу. Затем результаты с разных узлов объединяются с учетом дополнительных сигналов (например, доверие узла, статистика качества) и происходит окончательный ранжир [1].
   Плюсы: модель обучается локально, не требуется сбор всех данных в одном месте.
   Минусы: различие распределений между узлами может ухудшать качество, агрегация сложна.
2. Распределенное обучение (Data/Model Parallelism). Методы распределенного машинного обучения, такие как обучение с разделением данных или с разделением модели, применяются для обучения моделей, используемых для поиска/ранжирования. Особенно актуально, когда признаки или представления документов (векторы, embedding) используются в поиске.
   Пример: использование TensorFlow в распределенных системах для обучения нейронных моделей поиска.
3. Обучаемые индексы и итеративное перераспределение данных. Идея: традиционные индексы (B-дерево, хэш-таблицы) заменяются или дополняются ML-моделями, которые прогнозируют положение записи. В распределенном контексте это может сочетаться с динамическим перераспределением данных для балансировки нагрузки и ускорения доступа.
   Пример: работа IRLI (Iterative Repartitioning for Learning to Index – метод итеративного перераспределения данных для оптимальной индексации в распределенной среде).
4. Обучение с активным выбором запросов. Метод активного обучения может уменьшить затраты на разметку и улучшить качество определения релевантности документов, особенно в контексте информационного поиска [2].
5. Объяснимые методы поиска (Explainable IR). Поскольку модели поиска могут быть сложными (нейросети, градиентный бустинг и др.), важна интерпретируемость – чтобы понимать, почему система выдала тот или иной результат.

II. Структура гибридной архитектуры интеллектуального поиска

1. Партиционирование/шардирование данных. Данные (документы) распределяются по узлам по какому-то ключу (например, тематическому, хеш-функции, географическому признаку). Может быть использовано динамическое перераспределение (repartitioning) в зависимости от нагрузки.
2. Индексация и эмбеддинг. На каждом узле формируется локальный индекс (например, инвертированный) и представления документов в виде эмбеддингов (векторов в пространстве признаков) для более интеллектуального поиска сходства [3].
3. Обучение локальных моделей. На каждом узле обучается модель релевантности (например, градиентный бустинг, нейронная сеть) на локальной выборке, учитывая запросы и клики с этого узла.
4. Обмен метаинформацией/синхронизация. Узлы обмениваются некоторыми сводными статистиками, обобщенными моделями или агрегированными параметрами (например, через алгоритмы Federated Learning). Это может улучшать глобальную модель без передачи всех данных.
5. Маршрутизация запросов. Перед выполнением запроса система решает, на какие узлы его отправлять: полный цикл, часть узлов по тематике, узлы с высоким показателем качества и т.д.
6. Агрегация результатов и коррекция ранжирования. Отбираются наиболее релевантные результаты от каждого узла, затем выполняется агрегация с учетом доверия к узлам, статистики качества, глобальных признаков и возможно применения финального ранжировщика [4].
7. Обновление/дообучение. Периодически (или онлайн) происходит обмен сведениями, дообучение моделей, перераспределение данных и обновление индексов, чтобы система адаптировалась к новым данным и паттернам запросов.

III. Проблемы, ограничения и направления исследований

- Коммуникационные издержки: как минимизировать передачу данных между узлами, особенно при обмене признаков или моделей;
- неоднородность распределений: как справляться с ситуацией, когда узлы содержат данные разных тематик/доменов;
- конфиденциальность и приватность: важно сохранять приватность при обмене данными и параметрами моделей;
- временная изменчивость данных: как обеспечивать адаптивность (онлайн-обучение, адаптация к дрейфу данных);
- интерпретируемость: пользователям и администраторам важно понимать, почему был выбран тот или иной результат;
- оценка качества: как оценивать качество распределенного поиска, учитывая сеть, задержки, неполные соответствия между узлами;
- совместимость с существующими СУБД: как встроить ML-алгоритмы в существующие распределенные СУБД, не переписывая все «с нуля».

Заключение

Алгоритмы интеллектуального поиска информации в распределенных базах данных – тема, сочетающая вызовы из области информационного поиска, распределенных вычислений и машинного обучения. Применение ML-методов позволяет повысить качество поиска, адаптироваться к нагрузке и динамике, но требует продуманных архитектур распределенного обучения, обмена информации и агрегирования результатов.

Рекомендации:

- Для старта можно использовать гибридную стратегию: локальные модели и обмен сводной информацией;
- при обучении моделей использовать распределенный или федеративный подход, чтобы избежать централизованной передачи данных;
- исследовать алгоритмы перераспределения данных и обучаемых индексов в распределенных средах;
- внедрять элементы объяснимости и оценивать систему не только с точки зрения точности, но и с точки зрения задержек, коммуникаций и адаптивности.

Список литературы

1. Трофимов, И. Е. Распределенные вычислительные системы для машинного обучения / И. Е. Трофимов // Информационные технологии и вычислительные системы. – 2017. – № 3. – С. 64–65.
2. Аныш, Х. Применение инструментов машинного обучения и интеллектуальный анализ данных в отношении баз данных с небольшим количеством записей / Х. Аныш // Информатика, вычислительная техника и управление. – 2021. – С. 347–359.
3. Raschka, S. Machine Learning with PyTorch and Scikit-Learn / S. Raschka, Y. Liu, V. Mirjalili – Packt Publishing, 2022. – P. 649–660.
4. Макшанов, А. В. Современные технологии интеллектуального анализа данных: учебное пособие для СПО / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь – СПб.: Лань, 2020. – С. 12–19.
5. Львовский, С. М. Набор и вёрстка в системе LaTeX / С. М. Львовский // Издательство: МЦНМО, 2006. – 448 с.

СОПОСТАВЛЕНИЕ ПРОГРАММНЫХ РЕАЛИЗАЦИЙ, АЛГОРИТМЫ И ИНСТРУМЕНТЫ НАСТРОЙКИ НЕЙРОННЫХ СЕТЕЙ

Таранчук В. Б.
Кафедра компьютерных технологий и систем, Белорусский государственный университет
Минск, Республика Беларусь
E-mail: taranchuk@bsu.by

Будут обсуждаться итоги комплексного сравнительного анализа современных подходов к построению, обучению и оптимизации нейронных сетей на примере задачи сглаживания ряда данных (представительный набор данных наблюдений при наличии шумов, помех, выбросов, потерь). Рассмотрены не только теоретические аспекты архитектуры и обучения моделей, но и практические вопросы автоматизации поиска оптимальных параметров с использованием двух различных технологических платформ – Wolfram Mathematica и Python (TensorFlow, Keras).

Введение

Искусственный интеллект становится неотъемлемой частью современного мира – от автоматизации производства и диагностики заболеваний до создания транспортных систем, анализа финансовых рынков. Одним из наиболее заметных прорывов становятся искусственные нейронные сети (ИНС), которые уже не воспринимаются, как абстрактная технология из научных публикаций, а активно используются в повседневной жизни.

Параллельно с ростом популярности нейронных сетей быстро эволюционируют программные инструменты для их разработки, модификации и тестирования. Специалистам доступны универсальные вычислительные системы компьютерной математики, например, система компьютерной алгебры Wolfram Mathematica (WM) [1], гибкие фреймворки, в частности, на языке Python ([2]): TensorFlow, Keras, PyTorch, а также специализированные библиотеки для автоматизации подбора гиперпараметров – Optuna, Ray Tune и другие. Выбор подходящей экосистемы становится самостоятельной задачей, от которой зависит успех внедрения интеллектуальных технологий в промышленность, медицину, транспорт, сферу цифровых сервисов.

I. Инструменты Mathematica для построения и настройки ИНС.

WM предоставляет множество инструментов для удобного построения и настройки нейронных сетей без необходимости вручную программировать каждую деталь. В состав Mathematica входит модуль Neural Networks, который позволяет разработчикам использовать простой и понятный синтаксис для быстрого создания многослойных нейросетевых моделей практически любой сложности.

Конструкция нейронной сети в Mathematica строится декларативно, путём последовательного описания слоёв. Важной особенностью WM Neural Networks является наличие широкого набора базовых и специализированных слоёв. Пользователь может создавать последовательные и разветвлённые композиции слоёв, комбинируя: LinearLayer – для выполнения линейных преобразований; PoolingLayer – для уменьшения размерности и повышения устойчивости к шуму; ElementwiseLayer и BatchNormalizationLayer – для реализации поэлементных и нормализующих операций; ConvolutionLayer – для выделения пространственных признаков в изображениях; ReLU, Tanh, Softmax и другие функции активации – для внедрения нелинейности и корректного формирования выходного сигнала.

Система WM поддерживает различные форматы входных и выходных данных. Благодаря встроенным кодировщикам и декодерам, сеть может работать не только с числовыми векторами, но и с изображениями, текстами, аудиосигналами и их комбинациями. Дополнительным преимуществом Neural Networks WM является наличие большого числа предобученных моделей и специализированного Wolfram Neural Net Repository, где представлены архитектуры для решения самых разных задач. Пользователь может импортировать готовую сеть, адаптировать её под собственные данные, визуализировать структуру и ход работы, а также анализировать метрики качества интерактивно.

После того как архитектура нейронной сети определена и подготовлены данные, следующим шагом становится непосредственно обучение модели. В WM для этого используется функция NetTrain – позволяет запустить процесс обучения, указав обучающий набор данных и нужные параметры оптимизации. В NetTrain пользователь задаёт количество эпох (число полных проходов по всем обучающим данным), размер мини-пакета данных (batch size), а также выбирает устройство, на котором будет выполняться обучение (например, центральный процессор или графическая видеокарта). Чтобы контролировать качество обучения и избежать проблемы переобучения, можно указать дополнительный набор данных (валидационный) для оценки про-

межуточных результатов. Замечательной особенностью Neural Networks WM является встроенная визуализация процесса обучения. В режиме реального времени WM показывает, как изменяется ошибка, как растёт точность модели по мере её обучения. Пользователь может остановить обучение, скорректировать архитектуру нейросети или изменить гиперпараметры.

Настройка и оптимизация геперпараметров в WM. Одним из важнейших этапов в обучении нейронных сетей является настройка гиперпараметров, от которых напрямую зависит точность модели, скорость сходимости и общая устойчивость к переобучению. Гиперпараметры, в отличие от параметров модели, не обучаются напрямую во время градиентного спуска, а определяются до начала процесса обучения. В WM реализация настройки гиперпараметров осуществляется с использованием встроенных возможностей функции NetTrain, а также при помощи процедурного программирования и визуализации. WM предоставляет пользователю прямой доступ к наиболее значимым гиперпараметрам в виде опций, передаваемых в функцию NetTrain. В частности, к числу часто настраиваемых параметров относятся: скорость обучения (LearningRate), тип оптимизатора (Method), размер минипакета (BatchSize), количество эпох (MaxTrainingRounds), а также состав и доля валидационного набора (ValidationSet). Mathematica предоставляет пользователю богатый инструментарий для самостоятельного построения процедурного поиска оптимальных значений. Для симуляции методов Grid Search и Random Search могут использоваться такие конструкции языка, как Table, RandomReal, Map, NestList и другие итеративные средства. Для оценок насколько успешно обучилась нейронная сеть, можно использовать предоставляемые специальные метрики качества. В WM для этих целей предусмотрена функция NetMeasurements, используя которую можно легко вычислить основные важные показатели, включая: точность, среднеквадратичная ошибка, кросс-энтропия.

Также Mathematica предоставляет полезную функцию NetInformation. С её помощью можно получить подробную информацию об активной модели: сколько в ней параметров, какой её размер, какую точность она показывает на обучающих и валидационных данных. Конечно же, в контексте простых моделей возможно использование стандартных WM функций оптимизации (FindMinimum, NMinimize, MachineLearningModelParameterOptimization).

Важно и следует напомнить, что пользователи WM могут комбинировать с внешними средами (в частности, Python) через интерфейс ExternalEvaluate, WolframScript или REST API, можно подключать сторонние оптимизаторы и расширять функциональность системы.

II. Python – о средствах настройки и оценки ИНС

Язык программирования Python занимает ведущие позиции благодаря своей универсальности, лаконичному синтаксису и огромному количеству специализированных библиотек. Среди наиболее популярных и востребованных библиотек для работы с нейронными сетями на языке Python выделяют TensorFlow, Keras, PyTorch и Theano. В данной работе внимание уделено библиотекам TensorFlow и Keras, так как их инструменты позволяют быстро создавать нейронные сети разного уровня сложности, оперативно проводить эксперименты и прототипирование, эффективно применять готовые решения в задачах обработки и анализа больших объёмов данных.

На примерах рассматриваемой в работе задачи для повышения эффективности и качества моделей в TensorFlow реализованы инструменты автоматического поиска оптимальных архитектур и гиперпараметров, в том числе реализована интеграция с модулями Keras Tuner и Optuna. Отмечено, как пользователь может экспериментировать с различными конфигурациями сети, автоматически подбирать размер слоёв, функции активации, оптимизаторы и другие параметры без необходимости вручную запускать десятки экспериментов. Примером поясняется выбор гиперпараметров с использованием Optuna, которая основана на методах байесовской оптимизации и реализует улучшенный алгоритм поиска – Tree-structured Parzen Estimator (TPE), позволяющий эффективно исследовать пространство гиперпараметров.

Заключение

Сопоставление результатов решения задачи устранения шумов [3], вычислительных экспериментов проведенных в двух рассмотренных средах показали, что всегда точность решения конкретной задачи зависит не только от выбранной платформы, но и от характеристик искажений (шума) исходных данных, а также от метода оптимизации гиперпараметров. Вместе с тем, при правильной настройке обе исследуемые платформы (Wolfram Mathematica и Python с библиотеками TensorFlow, Keras и Optuna) способны стабильно показывать хорошие результаты, хотя, не полностью совпадающие.

III. Список литературы

1. Neural Networks in the Wolfram Language [Electronic resource] / Wolfram Documentation. – Mode of access: `https://reference.wolfram.com/language/tutorial/NeuralNetworksOverview.html/`. – Date of access: 29.09.2025.
2. Шолле, Ф. Глубокое обучение на Python. 2-е издание / Франсуа Шолле – СПб. : Питер, 2023. – 1041 с.
3. Таранчук, В. Б. Средства и примеры интеллектуальной обработки данных для геологических моделей / В. Б. Таранчук // Проблемы физики, математики и техники. – 2019. – № 3 (40). – С. 117–122.

ПРОГНОЗИРОВАНИЕ РИСКОВ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Тимошенко В. А., Бегляк Е. В., Мигалевич С. А.
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {v.timoshenko, katarina, migalevich}@bsuir.by

Доклад посвящен анализу современных методов прогнозирования рисков с использованием искусственного интеллекта. Рассматриваются вероятностные модели оценки рисков, нейросетевые методы генерации сценариев, алгоритмы машинного обучения для выявления закономерностей и оценка точности прогнозов. Приводятся примеры практического применения технологий ИИ для минимизации финансовых, производственных и технических рисков в различных сферах деятельности.

Введение

Современные организации функционируют в условиях высокой неопределенности, характеризующейся динамичными изменениями внешней среды, технологическим прогрессом и глобальными вызовами. Кризисы финансовых рынков, перебои в цепях поставок, стихийные бедствия и кибератаки подчеркивают необходимость своевременной оценки рисков и разработки сценариев развития событий [1].

Искусственный интеллект (ИИ) открывает новые возможности для анализа больших объемов данных, выявления скрытых закономерностей и автоматизации процесса оценки вероятности различных событий. Использование ИИ позволяет количественно оценивать вероятность наступления неблагоприятных исходов, строить сценарии их развития, анализировать сложные взаимозависимости между факторами и прогнозировать последствия [2].

I. Вероятностные методы прогнозирования

Вероятностные методы являются фундаментом прогнозирования рисков. Суть подхода заключается в описании риска как случайной величины с заданным распределением вероятностей. Пусть событие риска R характеризуется случайной величиной X, отражающей возможный ущерб. Тогда вероятность наступления критического события определяется выражением:

$$P(R) = \mathbb{P}(X \geq X_{\text{critical}}),$$

где X_{critical} – пороговое значение ущерба, при превышении которого событие считается значимым [3].

Ожидаемый ущерб (Expected Loss, EL) рассчитывается как сумма произведений вероятностей и соответствующих величин ущерба:

$$EL = \sum_{i=1}^{n} P(R_i) \cdot L_i,$$

где L_i – потенциальный ущерб при реализации риска R_i, $P(R_i)$ – вероятность наступления данного риска. Эти модели позволяют ранжировать риски по значимости и формировать приоритетные меры по их минимизации.

Для повышения точности прогнозов применяются методы Байесовской статистики, которые позволяют обновлять оценки вероятностей по мере поступления новой информации. Временные ряды и модели Markov Chain учитывают динамику факторов риска, обеспечивая прогноз на несколько шагов вперед и моделирование сценариев развития событий [4].

Дополнительно используется метод Монте-Карло, позволяющий моделировать тысячи возможных сценариев и оценивать распределение потерь с учетом редких, но потенциально катастрофических событий [5]. Такой подход применяется в страховании, финансовом секторе и инженерии безопасности.

II. Нейросетевые методы и генеративные модели

Современные методы машинного обучения применяются для прогнозирования сложных, многомерных рисков. Генеративные модели, такие как вариационные автоэнкодеры (VAE) и генеративные состязательные сети (GAN), создают новые сценарии на основе исторических данных и случайного шума:

$$\hat{X}_{t+1} = G_\theta(\hat{X}_t, Z_t),$$

где $\hat{X}_t$ – сгенерированные сценарии на шаге t, G_θ – генеративная функция с параметрами θ, Z_t – случайный шум [6].

Генеративные модели позволяют создавать стресс-сценарии, выходящие за рамки наблюдаемых данных, что особенно важно при анализе «черных лебедей» – редких и трудно прогнозируемых событий.

Рекуррентные нейронные сети (RNN) и модели с долгой краткосрочной памятью (LSTM) эффективны для анализа временных рядов, прогнозирования финансовых показателей, выявления сбоев в производственных процессах и мониторинга эксплуатационных рисков. Их преимущество заключается в умении учитывать долгосрочные

зависимости, которые часто играют ключевую роль в системах с накопительным эффектом.

Гибридные методы, объединяющие нейросети с экспертной оценкой и вероятностными моделями, повышают точность прогнозов и устойчивость системы к шуму и неполноте данных. Использование ансамблей моделей снижает влияние случайных ошибок и позволяет строить надежные прогнозы даже при ограниченном объеме данных.

Отдельным направлением является применение графовых нейронных сетей (GNN) для анализа взаимосвязей между элементами сложных систем, например, в цепях поставок или киберфизических системах. Такой подход позволяет моделировать каскадные эффекты и оценивать системные риски.

III. Оценка точности и надежности моделей

Оценка точности прогнозов является ключевым элементом. Стандартные метрики включают среднеквадратичную ошибку (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - X_i)^2},$$

а также среднюю абсолютную ошибку (MAE), коэффициент детерминации R^2 и вероятностные интервалы.

Методы кросс-валидации, бутстрепинга и тестирования на отложенной выборке помогают оценивать устойчивость моделей, предотвращать переобучение и гарантировать адекватность прогнозов новым данным. Также важно учитывать неопределенность прогнозов, чтобы управленцы могли принимать решения с вероятностным диапазоном возможных исходов.

IV. Практическое применение

ИИ для прогнозирования рисков используется в финансовой сфере, производстве, энергетике, здравоохранении, логистике и управлении проектами. Автоматизация анализа позволяет:

- выявлять потенциально опасные сценарии и тенденции;
- оценивать влияние отдельных факторов на вероятность ущерба;
- планировать ресурсное обеспечение для снижения вероятности негативных событий;
- создавать комплексные отчеты для руководства и органов контроля;
- адаптировать модели под специфику конкретной организации.

Примеры применения:

В производстве ИИ прогнозирует технические сбои оборудования, выявляет паттерны износа и прогнозирует аварийные ситуации; в финансовой отрасли модели помогают оценивать кредитные, рыночные и операционные риски; в здравоохранении прогнозирование рисков позволяет выявлять кризисные ситуации, планировать ресурсы и оптимизировать процессы оказания помощи [7]; в логистике и управлении цепями поставок ИИ прогнозирует перебои, сбои поставок и повышает устойчивость процессов.

Заключение

Использование ИИ для прогнозирования рисков обеспечивает точную оценку вероятностей, выявление скрытых закономерностей и автоматизацию анализа сложных сценариев. Наиболее перспективны гибридные модели, совмещающие вероятностные методы, генеративные модели и экспертную проверку. Практическая реализация требует качественных данных, прозрачности алгоритмов, регулярного контроля точности прогнозов и непрерывного совершенствования моделей.

Комплексный подход к прогнозированию рисков с использованием ИИ позволяет организациям принимать обоснованные решения, минимизировать потенциальные убытки и адаптироваться к изменяющимся условиям внешней среды, повышает устойчивость и конкурентоспособность.

1. Jorion, P. *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd ed. McGraw-Hill, 2006. ISBN 978-0071464950. – Дата доступа: 28.09.2025.
2. Goodfellow, I., Pouget-Abadie, J., et al. Generative Adversarial Nets [Electronic resource] // Advances in Neural Information Processing Systems. – 2014. – Mode of access: `https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`. – Date of access: 29.09.2025.
3. Hochreiter, S., Schmidhuber, J. Long Short-Term Memory [Electronic resource] // Neural Computation. – 1997. – Mode of access: `https://direct.mit.edu/neco/article/9/8/1735/6109/Long-Short-Term-Memory`. – Date of access: 29.09.2025.
4. Risk.net. Using AI to predict operational risk [Electronic resource] / Risk.net. – London, 2022. – Mode of access: `https://www.risk.net/operational-risk/7802746/using-ai-to-predict-operational-risk`. –Date of access: 30.09.2025.
5. Shreve, S. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, 2004. ISBN 0387401016. – Date of access: 30.09.2025.
6. Battiston, S., Farmer, J. D., et al. Financial networks and systemic risk [Electronic resource] // Science. – 2016. – Vol. 351, no. 6274. – P. 818-819. – Mode of access: `https://www.science.org/doi/10.1126/science.aad0299`. – Date of access: 01.10.2025.
7. Miotto, R., Wang, F., et al. Deep learning for healthcare: review, opportunities and challenges [Electronic resource] // Briefings in Bioinformatics. – 2018. – Vol. 19, no. 6. – P. 1236–1246. – Mode of access: `https://academic.oup.com/bib/article/19/6/1236/4102709`. – Date of access: 01.10.2025.

АРХИТЕКТУРЫ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ АНАЛИЗА МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Ходжиметов Э. Ш., Захарьев В. А.
Кафедра интеллектуальных информационных технологий, Кафедра систем управления
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {khodjimetovelyor05012003pr}@gmail.com

Представлен подход к анализу медицинских изображений на основе моделей глубокого обучения и современных архитектур компьютерного зрения. Использование сверточных нейронных сетей и трансформерных моделей позволяет повысить точность диагностики патологий на рентгеновских, МРТ и КТ-снимках. Описаны программные средства обработки изображений и результаты тестирования на открытых медицинских датасетах.

Введение

Анализ медицинских изображений является одной из ключевых задач современной медицины, так как позволяет своевременно выявлять заболевания и контролировать их развитие. Традиционные методы диагностики, основанные на визуальном анализе снимков врачом-специалистом, подвержены субъективности и ограничены физическими возможностями человека. Развитие технологий машинного обучения и компьютерного зрения позволило автоматизировать процесс обработки изображений и повысить точность диагностических решений. Использование нейронных сетей обеспечивает выделение скрытых информативных признаков, недоступных для классических алгоритмов фильтрации и сегментации [1].

I. Архитектуры и методы анализа

Современные методы автоматизированной интерпретации медицинских изображений сочетают достижения компьютерного зрения, глубокого обучения и медицинской визуализации. За последнее десятилетие произошёл переход от классических методов с ручным проектированием признаков к полностью обучаемым моделям на основе глубоких нейронных сетей – прежде всего свёрточных (CNN) и трансформеров (ViT).

CNN остаются основой благодаря эффективному извлечению локальных пространственно-инвариантных признаков, что идеально для обнаружения локальных аномалий (очагов, затемнений, кальцификатов). Ключевые архитектуры:

- ResNet – за счёт остаточных связей позволяют обучать очень глубокие сети, необходимые для детекции малоконтрастных (subtle) патологий;
- DenseNet – плотное соединение слоёв улучшает поток градиента и переиспользование признаков, повышая чувствительность к мелким паттернам (например, ранним интерстициальным изменениям на CXR);
- EfficientNet – compound scaling сбалансированно увеличивает глубину, ширину и разрешение, что критично для работы с высокодетализированными рентгенами, КТ и МРТ.

Трансформерные модели (ViT, Swin-Transformer) дополняют CNN, обеспечивая глобальный анализ изображения и лучшее распознавание сложных структур, таких как опухоли на МРТ [2].

Обработка изображений включает следующие этапы:

- нормализация и предобработка снимков;
- выделение признаков с помощью предобученных моделей;
- классификация или сегментация патологических областей.

В данной работе использована предобученная модель EfficientNet с адаптацией на медицинский датасет CheXpert, содержащий рентгеновские снимки органов грудной клетки. Наиболее перспективным направлением дальнейшего развития представляется использование ансамбля специализированных моделей, где каждая решает свою подзадачу: одна отвечает за высокоточную локализацию грубых аномалий, другая – за обнаружение малоконтрастных изменений, третья – за глобальный контекст и дифференциальную диагностику. Такое разделение ответственности позволяет повысить общую надёжность и интерпретируемость системы, а также эффективно бороться с проблемой domain shift. Принцип взаимодействия компонентов внутри общего ансамбля и последовательность этапов обработки представлены на блок-схеме алгоритма (рисунок 1).

II. Результаты экспериментов

Оценка эффективности предлагаемых методов была проведена на двух широко используемых в научном сообществе наборах данных грудных рентгенограмм – CheXpert и ChestX-ray14. Набор CheXpert включает пять ключевых целевых классов: ателектаз, кардиомегалия, консолидация, отёк лёгких и плевральный выпот. Набор ChestX-ray14, в свою очередь, содержит более ши-

рокий спектр диагнозов, среди которых для целей сопоставления в настоящей работе выделяются пневмоторакс и пневмония – патологии, традиционно представляющие особый интерес для задач первичного автоматизированного триажа.

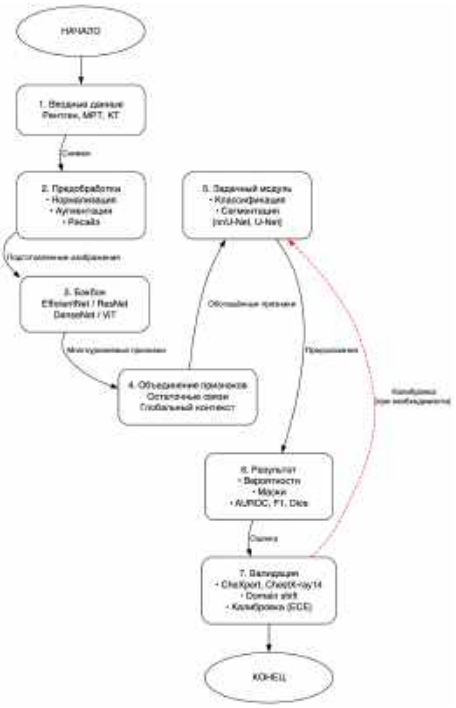


Рис. 1 – Алгоритм анализа медицинских изображений на основе ансамбля моделй

Основной метрикой оценки выбран AUROC – наиболее устойчивый показатель при дисбалансе классов и разных порогах. Дополнительно используется macro-F1, учитывающий чувствительность и точность по всем классам без весов. Это позволяет одновременно оценивать дискриминационную способность модели и её устойчивость к перекосу данных.

Согласно опубликованным результатам модели CheXzero (zero-shot/self-supervised) на тестовом наборе CheXpert, значения AUROC составляют: плевральный выпот – 0.958 (95 ДИ: 0.953–0.963), отёк – 0.961 (0.946–0.974), консолидация – 0.871 (0.851–0.888), кардиомегалия – 0.898 (0.894–0.903), ателектаз – 0.798 (0.780–0.817). Среднее значение AUROC по этим пяти классам – 0.897.

По метрике macro-F1 различия между моделью и радиологами статистически незначимы: 0.606 (95 ДИ: 0.571–0.638) против 0.619 (0.585–0.642).

Для набора ChestX-ray14 приведены результаты базовой модели ResNet-50 из оригинальной работы: пневмоторакс – 0.7891, пневмония – 0.6333. Среднее значение AUROC по этим двум классам – 0.711.

Полученные результаты показывают, что для визуально выраженных патологий (например, выпот, отёк) AUROC достигает ≈ 0.96, тогда как для менее однозначных находок (ателектаз, пневмония) значения заметно ниже. При внешней валидации на данных из других медицинских учреждений (domain shift) возможное снижение показателей требует проведения внешних тестов и оценки калибровки (например, по ECE или с использованием диаграмм надёжности)[6].

III. Выводы

Модели глубокого обучения продемонстрировали высокую эффективность в анализе медицинских изображений. Однако их клиническая применимость определяется не только точностью на внутренних наборах данных, но и устойчивостью к сдвигу домена, корректной калибровкой и полнотой отчётности.

Модели глубокого обучения показывают высокую эффективность в анализе медицинских изображений, но их клиническая применимость зависит не только от точности на внутренних датасетах, а от устойчивости к domain shift, правильной калибровки и полноты отчётности. Для сегментации надёжным стандартом остаются U-архитектуры и автонастраиваемый nnU-Net; перспективу межцентровой переносимости дают foundation-модели (MedSAM и аналоги).

Практическое внедрение требует внешней валидации, соблюдения TRIPOD+AI и EU AI Act (для высокорисковых систем). Рекомендуется использовать nnU-Net/MONAI как воспроизводимую основу, интегрировать трансформерные и foundation-модели, оценивать не только AUROC/Dice, но и калибровку (ECE, диаграммы надёжности), проводить внешние тесты, внедрять мониторинг drift и повторную калибровку, а также обеспечивать полную документацию и отчётность.

Список литературы

1. Ганичев П. А., Тихомирова А. А., Дохов М. А. Перспективы использования искусственного интеллекта в радиологии: краткий обзор // Визуализация в медицине. – 2023. – Т. 4, № 4. – С. 7-14.
2. Куликов А. А., Каширина И. Л., Савкина Е. Ф. Сегментация объёмных образований печени на мультифазных КТ-изображениях с использованием фреймворка nnU-Net // Моделирование, оптимизация и информационные технологии. – 2025. – Т. 13, № 1. – DOI: 10.26102/2310-6018/2025.48.1.040.
3. Лейхтер С. В. Обзор методов регистрации изображений на основе нежёстких преобразований // Математические структуры и моделирование. – 2024. – № 1(69). – С. 88–93. – DOI: 10.24147/2222-8772.2024.1.88-93.
4. Семантическая сегментация изображений с использованием методов глубокого обучения: учебные материалы. – ННГУ им. Н. И. Лобачевского, 2020. – 46 с.
5. Кочина Е. В., Швайко П. М., Лихолап О. Б. Создание и печать 3D-модели фрагмента лицевого скелета при помощи программы 3D Slicer // (материалы, описание методики; русскоязычная публикация).

НЕЙРОСЕТЕВЫЕ МОДЕЛИ СЕМАНТИЧЕСКОГО ПОИСКА В ПРЕДМЕТНО-ОРИЕНТИРОВАННЫХ КОЛЛЕКЦИЯХ ТЕКСТОВ

Чернявский К. Э., Марцинкевич В. А.
Отдел информационных технологий, отдел сетевых технологий, отдел информационных технологий,
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {k.cherniavskij, vlad}@bsuir.by

В работе рассматриваются нейросетевые подходы к семантическому поиску информации в предметно-ориентированных текстовых коллекциях. Анализируются архитектуры BERT и GPT, демонстрирующие способность учитывать контекст и смысловые связи в сложных запросах. Показано, что трансформеры обеспечивают более высокую релевантность результатов по сравнению с классическими методами. Приведены примеры интеграции моделей в специализированные поисковые системы, а также результаты экспериментального сравнения на корпусе научных публикаций.

Введение

Современные предметно-ориентированные коллекции текстов – научные архивы, отраслевые базы данных, специализированные библиотеки – требуют инструментов поиска, способных учитывать сложную терминологию и контекст. Классические методы, такие как TF-IDF и BM25, основаны на частотном анализе слов и не интерпретируют смысловые связи, что ограничивает их применимость в узкоспециализированных областях. Особенно это проявляется при работе с научными публикациями, где запросы могут включать многозначные термины, синонимы и сложные грамматические конструкции.

С развитием нейросетевых технологий появились модели, способные учитывать контекст и семантику текста. Архитектуры BERT и GPT, основанные на трансформерах, демонстрируют высокую эффективность в задачах интеллектуального поиска.

Цель работы – исследовать применение трансформеров в предметно-ориентированных коллекциях, сравнить их с классическими подходами и оценить качество поиска на реальных данных.

I. Классические методы поиска

До появления нейросетевых моделей основными инструментами поиска информации были статистические методы, основанные на частотном анализе слов. Одним из наиболее распространённых подходов является TF-IDF (Term Frequency-Inverse Document Frequency), который оценивает значимость термина в документе относительно его распространённости в коллекции. Несмотря на простоту и эффективность в ряде задач, TF-IDF не учитывает порядок слов, грамматические связи и контекст, что ограничивает его применимость при обработке сложных запросов.

Другим популярным методом является BM25 – вероятностная модель, основанная на расширении TF-IDF. Она учитывает длину документа и насыщенность терминами, что позволяет более гибко ранжировать результаты. Однако, как и TF-IDF, BM25 работает на уровне отдельных слов и не способен интерпретировать смысловые связи между ними.

Классические методы широко применяются в поисковых системах, таких как Lucene и Elasticsearch, благодаря своей скорости и простоте реализации. Тем не менее, они демонстрируют ограниченную эффективность при работе с многозначными терминами, синонимами и контекстнозависимыми запросами. Например, запрос «поиск информации с помощью нейросети» может не найти документы, содержащие фразы «глубокое обучение для извлечения данных» или «семантический анализ текста».

Несмотря на историческую значимость и широкое распространение, классические методы поиска уступают современным нейросетевым подходам в задачах, требующих понимания контекста и семантики.

II. Нейросетевые архитектуры

Ограничения классических методов стимулировали развитие моделей, способных учитывать контекст и семантику. Одним из наиболее значимых достижений в этой области стало появление трансформерных архитектур, таких как BERT (Bidirectional Encoder Representations from Transformers) и GPT (Generative Pre-trained Transformer).

Модель BERT обучается на задаче восстановления пропущенных слов и предсказания следующего предложения, что позволяет ей формировать двунаправленные контекстуальные представления текста. Значение каждого слова определяется не только его окружением слева, но и справа, что критически важно для понимания сложных языковых конструкций. GPT, в свою очередь, использует авторегрессионный подход,

при котором каждое следующее слово предсказывается на основе предыдущих, что делает модель особенно эффективной в генерации текста и диалоговых системах.

Обе архитектуры основаны на механизме внимания, который позволяет модели фокусироваться на наиболее значимых частях входного текста. Это обеспечивает высокую точность при обработке длинных и насыщенных контекстом запросов. Кроме того, трансформеры масштабируются на большие объемы данных и могут быть дообучены на специализированных корпусах, что делает их адаптируемыми к конкретным предметным областям.

Нейросетевые модели не только превосходят классические методы по качеству поиска, но и открывают возможности для семантического ранжирования, кластеризации документов, извлечения ответов и построения интеллектуальных агентов.

III. Интеграция в специализированные поисковые системы

Интеграция нейросетевых моделей в поисковые системы стала важным этапом в развитии технологий извлечения информации. Такие модели позволяют не только анализировать текстовые данные, но и интерпретировать смысл пользовательских запросов.

Одним из первых примеров масштабного внедрения трансформеров стала модель BERT, интегрированная в Google Search. Она позволила улучшить обработку длинных и сложных запросов, повысив точность ранжирования и релевантность выдачи. В отличие от классических алгоритмов, BERT учитывает контекст слов, грамматические зависимости и скрытые смысловые связи, что делает результаты поиска более осмысленными.

В научных системах, таких как Semantic Scholar и Microsoft Academic, нейросетевые модели используются для семантического поиска, автоматического аннотирования статей и построения тематических связей между публикациями. Это позволяет исследователям находить релевантные работы даже при использовании узкоспециализированной терминологии или нестандартных формулировок.

IV. Экспериментальное исследование

Для оценки эффективности нейросетевых моделей в задачах интеллектуального поиска информации было проведено экспериментальное сравнение трансформерной архитектуры BERT с классическим методом TF-IDF. В качестве тестового корпуса использовалась коллекция научных статей из области компьютерных наук, включающая более 5000 документов на русском и английском языках, отобранных из открытых научных источников. Запросы формировались вручную и охватывали как общие, так и специализированные темы, включая многозначные термины и контекстно-зависимые выражения.

Каждый запрос обрабатывался двумя системами: одна использовала TF-IDF с косинусным сходством, другая – предобученную на научных текстах модель BERT. Релевантность результатов оценивалась вручную экспертами по шкале от 0 до 3, где 3 означало полное соответствие запросу. Также рассчитывались стандартные метрики: точность, полнота и F1-мера.

Результаты показали явное преимущество нейросетевого подхода. Средняя точность BERT составила 0.87 против 0.62 у TF-IDF, полнота – 0.81 против 0.58, F1-мера – 0.84 против 0.60. Особенно заметно преимущество трансформеров при обработке запросов с вариативной терминологией и сложной синтаксической структурой. BERT демонстрировал устойчивость к вариативности формулировок и лучше справлялся с многозначностью.

Заключение

Рассмотренные в работе нейросетевые модели, основанные на трансформерных архитектурах, демонстрируют высокую эффективность в задачах интеллектуального поиска информации. В отличие от классических методов, они способны учитывать контекст, грамматические зависимости и скрытые смысловые связи, что особенно важно при обработке сложных и многозначных запросов.

Экспериментальное сравнение BERT и TF-IDF на корпусе научных публикаций показало значительное преимущество нейросетевого подхода по ключевым метрикам качества поиска. Это подтверждает целесообразность внедрения трансформеров в современные поисковые системы.

1. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // arXiv:1810.04805, 2018.
2. Brown T., Mann B., Ryder N. и др. Language Models are Few-Shot Learners // arXiv:2005.14165, 2020.
3. Lu Y., He D., Ouyang Y. и др. Neural Document Ranking with BERT // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2019.
4. Manning C. D., Raghavan P., Schütze H. Introduction to Information Retrieval. – Cambridge: Cambridge University Press, 2008. – 482 p.
5. Vaswani A., Shazeer N., Parmar N. и др. Attention Is All You Need // arXiv:1706.03762, 2017.

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ПЕРСОНАЛИЗАЦИИ В МОБИЛЬНЫХ ПРИЛОЖЕНИЯХ НА ОСНОВЕ АНАЛИЗА ПОЛЬЗОВАТЕЛЬСКИХ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Шоть К. С.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by, ksyshashot2005@gmail.com

В данной работе рассматриваются современные подходы, используемые в мобильных приложениях, а также методы, основанные на анализе пользовательских данных с применением машинного обучения.

Введение

В современном мире наблюдается мобильное доминирование – всё больше задач люди выполняют с помощью смартфонов. Конкурентоспособность приложения во многом определяется качеством пользовательского опыта (User Experience, UX) и полнотой реализованных функций. При этом простые шаблонные интерфейсы и минимальный набор возможностей уже не обеспечивают достаточной вовлечённости.

Персонализация, выражающаяся в адаптации интерфейса, рекомендациях контента, уведомлениях и настройке функций под тип пользователя, становится ключевым инструментом коммерческой эффективности и удержания аудитории. Использование методов машинного обучения (Machine Learning, ML) и искусственного интеллекта (Artificial Intelligence, AI) позволяет повысить релевантность контента, увеличить время взаимодействия и снизить отток пользователей

I. Типы данных для персонализации

Для создания профиля пользователя приложению необходимы различные типы данных. На их основе методы машинного обучения, представленные в следующей секции, «учатся». Это позволяет формировать индивидуальные рекомендации и адаптировать функционал. Выделяются следующие категории:

- Поведенческие данные: отражают взаимодействие с приложением – частоту использования, длительность сессий, последовательность действий.
- Демографические данные: характеризуют аудиторию по возрасту, полу, языку, региону.
- Контекстные данные: уточняют условия использования – время, местоположение, качество соединения, операционную систему и тип устройства.
- История предпочтений: отражает накопленный опыт – оценки, отзывы и реакции на рекомендованный контент.

Такое деление формирует основу для создания пользовательского профиля и признаков, используемых моделями машинного обучения.

II. Методы машинного обучения для персонализации приложений

Методы машинного обучения автоматически анализируют большие объемы пользовательских данных, что позволяет определить закономерности в пользовательских предпочтениях и поведении. Ниже рассмотрены ключевые направления машинного обучения, которые используются в современных системах.

1. Кластеризация пользователей. Метод неконтролируемого обучения, основанный на объединении пользователей в группы (кластеры) по схожим характеристикам или поведению. Основные алгоритмы:
   - K-means – делит пользователей на заданное число групп, минимизируя различия внутри каждой из них.
   - Иерархическая кластеризация – формирует древовидную структуру, отражающую взаимосвязи между кластерами.
   - DBSCAN – выделяет кластеры по плотности данных, что позволяет отсекать выбросы и находить группы произвольной формы.
2. Классификация предпочтений. Метод контролируемого обучения, при котором модель, обученная на размеченных данных, прогнозирует действия пользователя. Основные алгоритмы:
   - Логистическая регрессия – решает бинарные задачи, например, прогноз отклика на рекомендованный контент.
   - k-ближайших соседей (k-NN) – определяет категорию пользователя на основе схожести с известными примерами.
   - Машины опорных векторов (SVM) – строят нелинейные границы между группами, повышая точность анализа.

3. Рекомендательные системы. Ключевой элемент персонализации, обеспечивающий адаптивную подачу контента. Подходы к построению:
   - Коллаборативная фильтрация – анализирует сходства между пользователями или элементами контента, предполагая, что схожие пользователи совершают схожие выборы.
   - Контентно-ориентированные методы – предлагают контент, похожий на тот, с которым пользователь взаимодействовал ранее.
   - Гибридные модели – объединяют оба подхода, устраняя их недостатки, например проблему «холодного старта» при отсутствии данных о новом пользователе.
4. Контекстно-зависимые модели. Используют не только профиль пользователя, но и внешние условия взаимодействия: геолокацию, устройство, качество соединения и др. Основные подходы:
   - Нейронные сети – учитывают динамические параметры и формируют рекомендации в реальном времени.
   - Адаптивные алгоритмы – выбирают рекомендации в зависимости от текущего контекста.

III. Интеграция персонализации в мобильное приложение

Интеграция персонализации направлена на улучшение индивидуального пользовательского опыта. Процесс включает три основных этапа:

1. Сбор данных. Фиксируются профильные, поведенческие и контекстные данные пользователя для дальнейшего анализа.
2. Обработка. Данные очищаются, агрегируются и анализируются с применением моделей машинного обучения, формируя профиль пользователя с ключевыми признаками.
3. Применение. Интерфейс и функционал адаптируются под полученный профиль, обеспечивая персональные рекомендации и уведомления.

IV. Преимущества интеграции персонализации в мобильное приложение

Использование персонализации, особенно на основе методов машинного обучения, является ключевым фактором удержания пользователей и повышения эффективности приложения. Основные преимущества:

- Повышение вовлечённости: пользователю чаще предлагается контент, соответствующий его интересам.
- Рост продаж: анализ пользовательских действий позволяет выявлять закономерности и предлагать сопутствующие товары или услуги.
- Оптимизация пользовательского пути: приложение адаптирует фильтры и рекомендации на основе повторяющихся действий.
- Расширенная аналитика: персонализация облегчает сегментацию аудитории и повышает эффективность маркетинговых кампаний.

V. Недостатки, вызовы и ограничения

Интеграция персонализации сопровождается техническими, этическими и организационными трудностями:

1. Конфиденциальность данных. Необходимость сбора большого объёма информации повышает риск утечек и нарушений. Решение — анонимизация, шифрование и федеративное обучение.
2. Проблема «холодного старта». Недостаток данных о новых пользователях снижает точность рекомендаций. В качестве решения используются опросы и анализ базовых демографических сведений.
3. Ресурсоёмкость. Модели машинного обучения требуют значительных вычислительных ресурсов и синхронизации данных при минимальных задержках.
4. Этические риски. Избыточная персонализация может ограничивать разнообразие контента и перегружать уведомлениями. Решение – возможность настройки пользователем уровня персонализации.

VI. Заключение

Применение технологий персонализации в мобильных приложениях на основе анализа пользовательских данных и машинного обучения представляет собой мощный инструмент повышения качества пользовательского опыта и коммерческой эффективности. В данной работе рассмотрена методологическая база: какие данные используются, какие алгоритмы применимы, а также какие преимущества и недостатки имеет само применение персонализации.

VII. Список литературы

1. Personalized Machine Learning / J. McAuley // Cambridge University Press – 2022.
2. Practical Recommender Systems / K. Falk. // Manning Publications – 2019.
3. Алгоритмы анализа пользовательских предпочтений и генерации рекомендаций с использованием методов машинного обучения / Н. А. Александров, С. Н. Нестеренков // Информационные технологии и системы 2024 (ИТС 2024) = Information Technologies and Systems 2024 (ITS 2024) : материалы междунар. науч. конф., Минск, 20 нояб. / редкол. : Л. Ю. Шилин [и др.]. — Минск : БГУИР, 2024. -– С. 139–140.
4. Создание приложений машинного обучения. От идеи к продукту / Э. Амейзен // СПб.: «Инфра-М» – 2021
5. Recommender Systems Handbook (3rd ed.) / F. Ricci, L. Rokach, B. Shapira // Springer – 2022.

АНАЛИЗ ВЕКТОРОВ АТАК НА СИСТЕМЫ СОЦИАЛЬНОГО ВОССТАНОВЛЕНИЯ КЛЮЧЕЙ В СИСТЕМАХ СУВЕРЕННОЙ ИДЕНТИФИКАЦИИ (SSI) И РОЛЬ ПОРОГОВОЙ КРИПТОГРАФИИ В ИХ МИТИГАЦИИ

Шушкевич Д. В.
Факультет компьютерных систем и сетей,
Кафедра программного обеспечения информационных технологий,
Белорусский Государственный Университет Информатики и Радиоэлектроники
Минск, Республика Беларусь
E-mail: dzmitry.shushkevich@gmail.com

В работе анализируются системы социального восстановления ключей (SKR) в контексте суверенной идентификации (SSI). Систематизированы векторы атак (сговор опекунов, социальная инженерия, атака Сивиллы), показывающие уязвимость наивных реализаций SKR, основанных на социальном доверии. В качестве основного механизма митигации предложено использование пороговых схем подписи (TSS). Сравнительный анализ моделей SKR на смарт-контрактах и с усилением TSS демонстрирует криптографическую устойчивость последних к атакам сговора и краже ключа. Сформулированы направления для дальнейших исследований.

Введение

Парадигма суверенной идентификации (Self-Sovereign Identity, SSI), основанная на децентрализованных идентификаторах (DID) и верифицируемых учетных данных (VC), возвращает пользователям контроль над цифровой личностью.[1, 2, 3] Однако, ключевой проблемой массового внедрения SSI остается управление криптографическими ключами: в децентрализованных системах их утеря равносильна безвозвратной потере личности и активов, что критично для нетехнических пользователей.[4, 5]

Для решения этой проблемы предложены системы социального восстановления ключей (Social Key Recovery, SKR). SKR позволяет пользователю задействовать социальную сеть «опекунов» (guardians) для восстановления доступа, [6, 7] что снижает порог входа. Однако реализации SKR на базе смарт-контрактов вводят новые уязвимости, основанные на социальном доверии.[8]

I. Основы SSI и социального восстановления

Парадигма суверенной идентификации

Концепция SSI возвращает пользователям контроль над идентификационными данными, [1, 2] опираясь на технологические столпы: децентрализованные идентификаторы (DID) – уникальные URI, разрешаемые в JSON-документ с ключами, [3, 7] и верифицируемые учетные данные (VC) – криптографически подписанный эмитентом набор утверждений.[3, 2]

Механизм социального восстановления ключей

Технически, кошелек с социальным восстановлением (SKR) функционирует как смарт-контракт, развернутый в блокчейне. В контракте запрограммированы ключевые роли: Ключ для подписи (Signing Key) для регулярных операций и Опекуны (Guardians) – набор доверенных адресов (друзья, устройства), кворум (например, 3 из 5) которых может коллективно одобрить смену ключа. При утере ключа владелец инициирует процедуру, опекуны подписывают транзакцию, и смарт-контракт заменяет старый ключ на новый, предоставленный владельцем.

II. Анализ векторов атак

Угрозы для SKR можно разделить на атаки социального и технического уровней.

Атаки на социальном уровне

- Сговор опекунов. Кворум опекунов может объединиться для замены ключа владельца на свой собственный, похитив его цифровую личность.[3, 4] Эта уязвимость усугубляется «проблемой рационального актора»: в системе, где сговор максимизирует выгоду участников, он является предсказуемым результатом.
- Социальная инженерия. Злоумышленник может обманом убедить владельца или опекунов инициировать процедуру восстановления на вредоносный ключ.
- Атака Сивиллы (Sybil Attack). Атакующий убеждает пользователя назначить опекунами несколько псевдонимов, контролируемых одним лицом. Получив контроль над кворумом, он может единолично захватить аккаунт.

К этим угрозам добавляются технические уязвимости, такие как ошибки в коде смарт-контракта или компрометация устройств опекунов, что нельзя сбрасывать со счетов.[9]

III. Пороговая криптография как

механизм митигации

В отличие от схем типа SSS, требующих уязвимой реконструкции ключа в одной точке, пороговые схемы подписи (TSS) являются более совершенным решением. TSS — это протокол многосторонних вычислений (MPC), в котором полный секретный ключ никогда не собирается в одном месте. Участники (владелец и опекуны) совместно генерируют единый публичный ключ и индивидуальные доли секрета. Для авторизации операции пороговое число участников совместно создает единую цифровую подпись, не раскрывая свои доли. Таким образом, распределяется сама способность подписывать, а не части ключа, что полностью устраняет единую точку отказа.

IV. Сравнительный анализ моделей

В модели TSS право владения привязано к единому публичному ключу, а управляющие транзакции требуют пороговой подписи. Сравнение моделей SKR на смарт-контрактах и с TSS представлено в таблице 1.

Таблица 1 – Сравнение моделей SKR по устойчивости к атакам

Атака	SKR (смарт-контракт)	SKR + TSS	Эффективность TSS
Сговор опекунов	Уязвима	Устойчива	Сговор позволяет подписать транзакцию, но не украсть ключ. Атака криптографически невозможна.
Социальная инженерия	Уязвима	Частично митигирована	Требуется обмануть пороговое число t опекунов, а не одного владельца. Сложность атаки возрастает.
Атака Сивиллы	Уязвима	Частично митигирована	Атакующий может авторизовать транзакции, но не может украсть долгосрочный ключ.
Уязвимость реконструкции	Уязвима	Устойчива	Приватный ключ никогда не восстанавливается в одном месте, что устраняет единую точку отказа.

Заключение

Проведенный анализ показывает, что системы социального восстановления ключей являются важным шагом к повышению удобства использования SSI для массовой аудитории. Однако их безопасность не может основываться на хрупкой модели социального доверия. Атаки на социальном уровне, в частности сговор опекунов, мотивированный как злым умыслом, так и рациональным эгоизмом, представляют собой критическую уязвимость в наивных реализациях SKR.

Внедрение пороговой криптографии, и в частности пороговых схем подписи, обеспечивает необходимую криптографическую основу для усиления этих систем. Распределяя возможность подписи без централизации или реконструкции ключа, TSS превращает SKR из системы, основанной на отзывном человеческом доверии, в систему с надежной, верифицируемой криптографической безопасностью.

Дальнейшая работа в этой области должна быть сосредоточена на разработке формальных, доказуемо безопасных моделей для кошельков на базе TSS, исследовании механизмов выбора опекунов, устойчивых к атаке Сивиллы (возможно, с использованием анализа социальных графов), и анализе производительности протоколов DKG и распределенной подписи на потребительских устройствах.

Список литературы

1. Allen, C. The Path to Self-Sovereign Identity // Life With Alacrity. – 2016. [Electronic resource]. – Mode of access: `http://www.lifewithalacrity.com/2016/04/the-path-to-self-sovereign-identity.html`.
2. Müller, P. A Survey on Essential Components of a Self-Sovereign Identity / P. Müller, C. P. Mainka, S. Kruse, T. Wollnitza, J. Pohlmann // arXiv preprint arXiv:1807.06346. – 2018.[6]
3. Buterin, V. Why we need wide adoption of social recovery wallets [Electronic resource]. – 2021. – Mode of access: `https://vitalik.ca/general/2021/01/11/recovery.html`.
4. Korir, D. Smart Contract-Based Social Recovery Wallet Management Scheme for Digital Assets / D. Korir, H. Kim, A. Yousaf // 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC). – 2023. – P. 0506–0512.[3]
5. Douceur, J. R. The Sybil Attack // Peer-to-Peer Systems. – Springer, 2002. – P. 251–260.
6. Halpern, J. Y. Rational secret sharing and multiparty computation / J. Y. Halpern, V. Teague // Proceedings of the thirty-sixth annual ACM symposium on Theory of computing. – 2004. – P. 623–632.
7. W3C. Decentralized Identifiers (DIDs) v1.0 [Electronic resource] // W3C Recommendation. – 2022. – Mode of access: `https://www.w3.org/TR/did-core/`.
8. Барсукевич, С. Н. Смарт-контракты и их роль в обеспечении доверия в финансовых транзакциях / С. Н. Барсукевич, С. Н. Нестеренков, П. С. Жуковец // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей X Международной научно-практической конференции – Минск : БГУИР, 13 марта 2024 г. ; – Ч. 1. – С. 309–315.

ANALYSIS OF THE ARCHITECTURE DESIGN APPROACH BASED ON REQUIREMENTS

Adzinets D., Sychev Y., Alooeff E.
Belarusian State University of Informatics and Radioelectronics, Kaspi Bank JSC, ATEK
Minsk, Belarus, Almaty, Kazakhstan, Wroclaw, Poland
E-mail: adzinets@bsuir.by, sigmadel10@gmail.com, alooeff@atek.dev

We present a requirement-driven architecture for .NET production applications: three layers (API, Core, Infrastructure) with dependency inversion and low coupling. Over one year, unit-test coverage rose from 7% to 78%, data-source switch time fell from 4.2 days to 1 day, .NET migration time fell from 1 week to 1 day, and null/mapping incidents decreased from 23 per quarter to 3 per quarter. This separation simplifies testing, speeds migrations, and lets teams replace data stores without changing business logic.

Introduction

During the software lifecycle there is a design phase. It is common to use reusable, proven solution to a recurring problem at the system level, addressing concerns related to the overall structure, component interactions, and quality attributes of the system. This is also known as software architecture patterns and they typically affect system-level concerns. Modern software applications must not only meet business requirements but also be easy to adapt to changes and remain reliable.

In real production systems, applications often fall into two extremes: the lack of any architecture, which leads to "spaghetti code" and huge methods, or the other extreme — excessive use of patterns and layers without real need. This article considers the principle of selecting and designing architecture based on key requirements.

Therefore, practical principles for modern production applications are prioritized over theoretical ideas applied without understanding.

I. Methods

Based on the team's experience maintaining different projects, the following requirements for new application architecture were defined:

- Simplicity of component testing;
- Easy migration to new platform versions;
- Simplicity of maintenance and code understanding;
- Ability to replace data sources without changing business logic;
- Avoiding libraries that heavily affect architecture.

II. Justification of Architecture Choice

Testability was one of the main priorities during design. As Microsoft recommendations show, for easy unit testing, a central layer that does not depend on the infrastructure or other layers is needed [1]. Therefore, it was decided to use a Core layer for business logic and interfaces for external communication; and an Infrastructure layer for implementations.

Since the Core layer is isolated from external dependencies, business rules can be tested independently, without running a database, web server, or other infrastructure elements. For integration tests, this also brings advantages: instead of real infrastructure components, you can connect alternative implementations (fake services, in-memory databases, etc.), which makes testing complex scenarios much easier [2] [3].

This approach also solves another requirement – the ability to replace data sources without changing business logic (fig. 1).

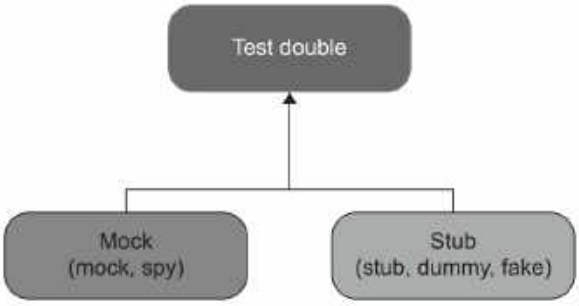


Figure 1 – All variations of test doubles can be categorized into two types: mocks and stubs.

III. Project Structure

The architecture follows the Dependency Inversion principle: the infrastructure depends on abstractions defined in the application core [1].

The project is divided into three main layers:

- API – controllers, input validation, serialization;
- Core – domain models, use cases, repository contracts;
- Infrastructure – data access implementations, mapping.

Table 1 – Application Layers, Responsibilities, and Dependencies

Layer	Responsibility	Dependencies
API	controllers, input validation, serialization	Core, Infrastructure (for DI only)
Core	domain models, use cases, repository contracts	—
Infrastructure	implementation of data access, mapping	Core

Note: the API project references Infrastructure only for dependency injection configuration; business logic (Core) knows nothing about Infrastructure types.

As the table shows, business logic is fully concentrated in the Core layer, which does not depend on external details. The Core layer contains the domain model (entities, use cases) and abstractions for operations that need implementation on the infrastructure level – for example, repository interfaces for data access. These interfaces define interaction contracts, and concrete implementations are located in the Infrastructure layer. This achieves the Dependency Inversion Principle: instead of business logic depending on the database or data framework, infrastructure details depend on abstractions defined in Core [4] [5].

IV. Results

During one year of using the proposed architecture in production projects, quantitative and qualitative metrics confirmed its effectiveness. The comparison before and after implementation is shown below.

Table 2 – Before/After Metrics (average over 1 year)

Metric	Before	After
Business logic unit test coverage	7%	78%
Average time to change data source	4.2 days	1 day
Time for .NET version migration	1 week	1 day
Number of null/mapping incidents	23 per quarter	3 per quarter

V. Interpretation of Results

The increase of test coverage by more than 11 times became possible due to the isolation of business logic in Core and the removal of tight integration with infrastructure components. This improved confidence in changes and simplified refactoring.

Reducing the time required to switch data sources confirms the independence from specific storage implementations: to migrate to another storage, it is enough to implement a new repository interface without touching business logic.

Faster platform migration is explained by minimal dependency on platform APIs and high modularity. Infrastructure updates do not require rewriting Core logic.

The decrease in mapping and null-related errors is related to clear separation of responsibilities and centralized data transformation handling in the Infrastructure layer.

VI. Conclusion

Applying a requirement-based architecture helps to create flexible, testable, and reliable systems. Practical results show a significant increase in quality and a reduction in operational costs.

1. Common web application architectures. Microsoft Learn *[Electronic resource]*. Available from: `https://learn.microsoft.com/en-us/dotnet/architecture/modern-web-apps-azure/common-web-application-architectures` (accessed 24 Oct 2025).
2. Martin, R. C. *Clean Architecture: A Craftsman's Guide to Software Structure and Design*. Pearson, 2017. 432 p.
3. Fowler, M. *Patterns of Enterprise Application Architecture*. Addison-Wesley Professional, 2002. 560 p.
4. Khorikov, V. *Unit Testing Principles, Practices, and Patterns*. Manning Publications, 2020. 304 p.
5. Lock, A. *ASP.NET Core in Action*. Third Edition. Manning Publications, 2023. 984 p.

LLM-BASED INTELLIGENT SYSTEM FOR QUALITY CONTROL OF STUDENT ASSIGNMENTS

Dmitry Pakutnik
Department of Intelligent Information Technologies,
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
E-mail: d.pakutnik@bsuir.by

This work explores the potential of large language models for automated quality control of student assignments through a human-in-the-loop approach. The proposed LLM-based intelligent system uses the generative abilities of language models to interact with experts and refine evaluation criteria dynamically. This adaptive process enhances grading accuracy, consistency, and interpretability, bringing automated assessment closer to human-level quality.

Introduction

The rapid development of large language models has transformed many areas of education, including automated assessment and feedback generation. Traditional grading systems often rely on predefined rules or static algorithms, which limits their ability to evaluate open-ended and concept-based student responses. Such approaches frequently lack flexibility, interpretability, and alignment with pedagogical standards.

To overcome these limitations, this research explores the integration of large language models with expert feedback within a human-in-the-loop framework [1]. The approach seeks to create an adaptive mechanism for refining evaluation criteria and ensuring transparent, consistent grading decisions. By combining the reasoning capacity of language models with human expertise, the study aims to enhance both the automation and pedagogical quality of assessment, contributing to the advancement of intelligent educational technologies.

I. Related work

Automated assessment technologies have progressed from early rule-based and feature-engineered systems to modern transformer architectures capable of semantic understanding. Models such as BERT and RoBERTa improved the accuracy of short-answer grading but still rely on predefined rubrics and lack adaptability when facing diverse student responses [2]. These limitations restrict their application in real educational environments where interpretability and consistency are essential.

The emergence of large language models has significantly expanded the possibilities of automated evaluation. LLMs demonstrate advanced reasoning and contextual comprehension, enabling more flexible grading that considers both the content and intent of student responses [3]. However, fully automated systems often misinterpret ambiguous instructions or domain-specific criteria, resulting in inconsistent evaluations.

To address these challenges, recent research explores **human-in-the-loop (HITL)** approaches that integrate expert feedback into the grading process [4]. Studies such as *GradeHITL* show that human-guided refinement of grading rubrics increases reliability, transparency, and alignment with pedagogical goals. These developments highlight the potential of combining LLM-based reasoning with expert guidance to create adaptive and interpretable assessment frameworks — an idea further explored in the following section.

II. Proposed system

The proposed LLM-Based Intelligent System for Quality Control of Student Assignments extends the human-in-the-loop framework to educational quality control by integrating LLM-based semantic reasoning, expert knowledge, and reinforcement learning for adaptive rubric optimization. This design enables both high-level automation and continuous refinement of grading criteria based on expert feedback.

The automated evaluation process is formulated as a classification function:

$$\hat{y} = F_\theta(G \parallel a),$$

where G is the grading rubric, a is the student's answer, θ are the LLM parameters, and $\parallel$ denotes text concatenation. The system assigns each answer a to one of the categorical score classes $\{c_i \mid i = 1, \ldots, C\}$, where C represents the total number of categories.

Unlike traditional fine-tuning of model parameters θ, which is computationally expensive and requires large labeled datasets, this approach focuses on **rubric optimization** through human feedback. The rubric G is iteratively refined into a more accurate and interpretable version G^*, defined as:

$$G^* = \underset{G}{argmax}\, \mathbb{E}_{a \in A}[\mathrm{Eval}(F_\theta(G \parallel a), y)]$$

where *Eval* measures the agreement between model predictions and expert-assigned scores. An overview of the framework is illustrated in Figure 1.

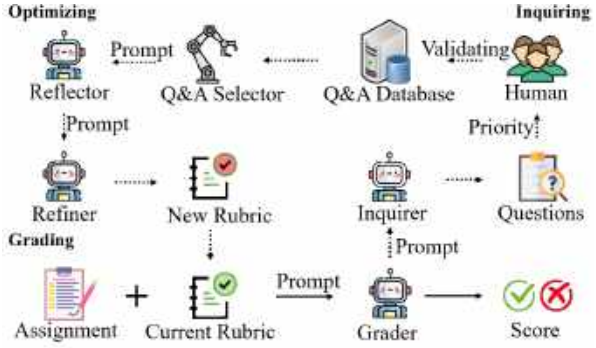


Figure 1 – Automated Grading with Human-in-the-Loop

Model Architecture The system comprises three interacting components that form a continuous feedback loop:

1. **Grader** – the LLM-based core evaluator responsible for analyzing student responses according to the current rubric. It employs *chain-of-thought (CoT)* reasoning to produce both intermediate explanations and final grades. This improves transparency and identifies ambiguities or inconsistencies within the rubric.
2. **Inquirer** – a generative component that detects rubric ambiguities or low-confidence grading cases and formulates targeted clarification questions. These questions are ranked using an uncertainty metric and filtered through a reinforcement learning policy π_θ that optimizes the selection of the most informative expert interactions according to the reward function:

$$r = R(F(G,x,h) \mid x) = \text{Eval}(\hat{y},y)$$

where h is a question–answer pair from the expert database. A higher reward r indicates improved grading accuracy and rubric consistency. The reinforcement learning workflow used for Q&A selection and feedback optimization is illustrated in Figure 2.

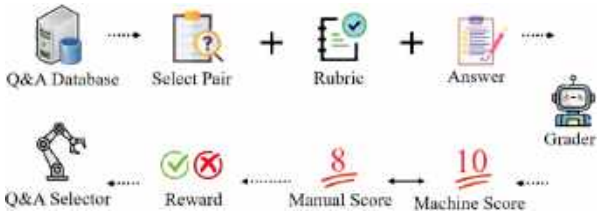


Figure 2 – Illustration of the RL-training process for Q&A selection and feedback optimization

3. **Optimizer** – refines the rubric by incorporating validated human–AI interactions. It operates as a multi-agent mechanism consisting of the *Retriever*, *Reflector*, and *Refiner*, which iteratively enhance the rubric G_t through gradient-like updates:

$$G_t = G_{t-1} + \Delta G$$

where $\Delta G = f(\text{Expert Q\&A}, A_{\text{error}})$

This iterative optimization process ensures that each refinement step integrates expert insights into the grading logic, improving rubric clarity and adaptability.

Through the interaction of these components, the system dynamically aligns computational evaluation with expert judgment. The result is a grading framework that combines automation and human oversight to achieve consistent, interpretable, and pedagogically sound quality control of student assignments.

III. Conclusion

This work outlines an approach to integrating large language models with expert feedback to enhance the quality and consistency of automated assessment. The proposed human-in-the-loop mechanism is intended to support dynamic refinement of grading rubrics, improving transparency and alignment with pedagogical standards [5].

By continuously incorporating expert insights, such a framework could evolve from a static evaluator into an adaptive learning system capable of resolving ambiguities and maintaining reliable performance across diverse educational contexts. The approach contributes to advancing automated assessment toward greater interpretability, fairness, and instructional value.

IV. References

1. Chu Y., Li H., Yang K., Shomer H., Liu H., Copur-Gencturk Y., Tang J. A LLM-powered automatic grading framework with human-level guidelines optimization [Electronic resource] // arXiv. – 2024. – arXiv:2410.02165. – Mode of access: `https://arxiv.org/abs/2410.02165`. – Date of access: 23.10.2025
2. Devlin J. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Electronic resource] // arXiv, 2018. – Mode of access: `https://arxiv.org/abs/1810.04805`. – Date of access: 23.10.2025
3. Xie W., et al. Grade like a human: Rethinking automated assessment with large language models [Electronic resource] // arXiv. – 2024. – arXiv:2405.19694. – Mode of access: `https://arxiv.org/abs/2405.19694`. – Date of access: 23.10.2025
4. Wu X., Xiao L., Sun Y., Zhang J., Ma T., He L. A survey of human-in-the-loop for machine learning // Future Generation Computer Systems. – 2022. – Vol. 135. – P. 364–381
5. an L., Sha L., Zhao L., Li Y., Martinez-Maldonado R., Chen G., Li X., Jin Y., Gašević D. Practical and ethical challenges of large language models in education: a systematic scoping review // British Journal of Educational Technology. – 2024. – Vol. 55, № 1. – P. 90–112. – Mode of access: `https://bera-journals.onlinelibrary.wiley.com/journal/14678535`. – Date of access: 23.10.2025

ZERO-SHOT AUTOMATIC ASSIGNMENT GRADING WITH LARGE LANGUAGE MODELS

Dmitry Pakutnik
Department of Intelligent Information Technologies,
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
E-mail: d.pakutnik@bsuir.by

This work proposes a conceptual framework for zero-shot automatic assignment grading using large language models. The approach employs prompt-based reasoning to generate both grades and explanatory feedback, aiming to improve scalability, transparency, and pedagogical value in higher education assessment.

Introduction

Assessment plays a key role in higher education, helping to ensure that learning outcomes are measurable and consistent with academic expectations. However, grading continues to be time-consuming and inconsistent, despite decades of automation efforts. Most existing systems rely on predefined answers or large labeled datasets, limiting their flexibility across disciplines and languages.

Recent progress in large language models has introduced new possibilities for reasoning-based evaluation [1]. This paper proposes a conceptual framework for zero-shot automatic assignment grading, where language models assess student work through prompt-guided reasoning rather than training data. The approach aims to enhance scalability, transparency, and pedagogical feedback while reducing the burden on educators [2].

I. Background and Problem Definition

Before turning to newer language-based methods, it is helpful to briefly describe the main types of automated grading systems used today. Traditional automated grading systems are typically of two kinds:

1. **Rule-based evaluators**, which depend on rigid templates or pattern matching;
2. **Supervised learning models**, which require extensive annotated datasets to learn from human grading examples.

Both approaches are constrained by domain specificity and lack the flexibility to adapt to new courses or assignment types without substantial reconfiguration. They also tend to focus on correctness alone, neglecting the explanatory and formative aspects of feedback that support student learning.

Manual grading, on the other hand, remains the gold standard for interpretive accuracy but is slow, inconsistent, and subject to human bias. As class sizes increase and curricula diversify, maintaining both speed and fairness becomes unsustainable. Therefore, an adaptive, general-purpose evaluation system capable of reasoning about open-ended student responses is a critical need in higher education.

II. Conceptual Framework

The proposed **zero-shot LLM grading framework** is built upon the principle of linguistic reasoning rather than statistical training [3]. It positions the language model as an autonomous evaluator capable of interpreting assignment prompts, understanding the logic of student responses, and generating both evaluative judgments and explanatory feedback. The foundation of this approach lies in *prompt engineering*, where well-structured textual instructions define the assessment context, criteria, and expected response format.

Unlike traditional systems that depend on predefined answers or domain-specific models, the zero-shot framework allows for flexibility across disciplines and assignment types. The grading logic is expressed entirely in natural language, enabling the model to apply general reasoning skills to diverse academic inputs. This adaptability makes the framework suitable for subjects involving quantitative problem-solving, conceptual explanations, or written argumentation.

As illustrated in Fig. 1, the framework operates through four core stages:

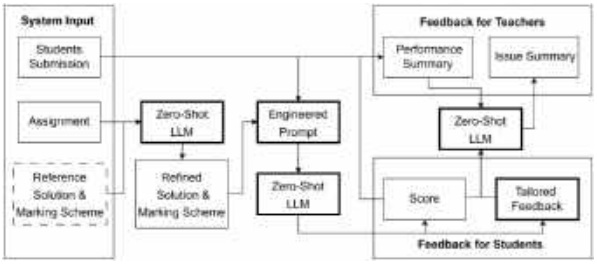
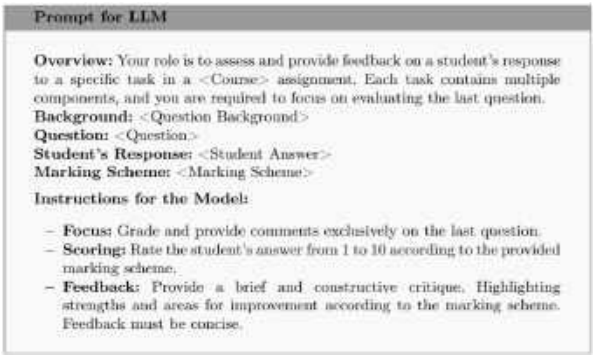


Figure 1 – Overview of the Zero-Shot LLM-Based Automated Assignment Grading System

1. **Input Assembly**
   The system consolidates essential components of the evaluation task – the assignment question, the student's response, and, when available, a grading rubric or reference criteria. These inputs provide the contextual basis for assessing both the content and quality of the submission.
2. **Prompt Construction**
   A structured prompt is formulated to guide the model's reasoning. The prompt defines the evaluator's role, outlines the assessment

dimensions–such as correctness, completeness, reasoning, and clarity–and specifies the required output: a numerical grade accompanied by a concise explanation. A conceptual example of an evaluation prompt is provided in Fig. 2.


Prompt for LLM

Overview: Your role is to assess and provide feedback on a student's response to a specific task in a <Course> assignment. Each task contains multiple components, and you are required to focus on evaluating the last question.
Background: <Question Background>
Question: <Question>
Student's Response: <Student Answer>
Marking Scheme: <Marking Scheme>

Instructions for the Model:

- **Focus:** Grade and provide comments exclusively on the last question.
- **Scoring:** Rate the student's answer from 1 to 10 according to the provided marking scheme.
- **Feedback:** Provide a brief and constructive critique. Highlighting strengths and areas for improvement according to the marking scheme. Feedback must be concise.

Figure 2 – Evaluation prompt for the Automated Assignment Grading system

3. **Evaluation and Feedback Generation**
   The language model processes the prompt and produces an integrated response consisting of a score and a written justification. The explanation highlights the logic behind the assessment, thereby functioning as immediate formative feedback for the learner.
4. **Consistency Moderation**
   To maintain reliability, multiple evaluations may be aggregated or cross-validated. Statistical normalization and human moderation can be applied to ensure alignment with institutional grading standards and to prevent systematic deviation.

Through these stages, the framework transforms the assessment process into an interaction driven by understanding and reasoning rather than data memorization. The use of natural-language prompts allows the evaluator's expectations to be communicated directly to the model, enabling adaptable and interpretable grading. The resulting system combines computational efficiency with pedagogical depth, ensuring that every assigned score is supported by a transparent and instructive rationale.

III. Advantages of the Zero-Shot Approach

The zero-shot approach introduces a new paradigm in automatic assignment grading by relying on linguistic reasoning rather than task-specific training or large labeled datasets. Its flexibility across domains allows the same framework to evaluate diverse subjects, from humanities essays to technical problem-solving, making it adaptable to a wide range of courses and assignment types.

Another key benefit is its reduced resource dependence. Traditional automated grading systems require extensive datasets and retraining, whereas the zero-shot framework operates using only task descriptions and rubrics embedded in prompts, making it more accessible for institutions with limited computational resources. It also ensures transparency and interpretability, as each evaluation includes a textual explanation outlining the reasoning behind the assigned grade, supporting both academic integrity and formative learning.

Additionally, the framework offers rapid adaptation and pedagogical support. By modifying prompts, the system can adjust to new assignments or updated rubrics without retraining, while the explanatory feedback guides students in identifying conceptual gaps and improving reasoning skills.

IV. Conclusion

The zero-shot automatic assignment grading framework represents a promising approach to modernizing assessment in higher education. By leveraging the reasoning and generative capabilities of large language models through structured prompts, it offers scalable, adaptable, and interpretable evaluation across diverse subjects and assignment types. The integration of explanatory feedback alongside numeric scores enhances the educational value of grading, supporting student learning and reflection while maintaining transparency and academic integrity.

This approach has the potential to transform grading from a time-consuming administrative task into a dynamic, formative process. With careful design, prompt optimization, and human oversight, zero-shot grading frameworks can improve consistency, reduce resource demands, and provide meaningful guidance to students, paving the way for more efficient and pedagogically effective assessment in universities.

1. Brown T., Mann B., Ryder N., Subbiah M., Kaplan J. D., Dhariwal P., Neelakantan A., Shyam P., Sastry G., Askell A., et al. Language models are few-shot learners // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 1877–1901
2. Chiang C. H., Chen W. C., Kuan C. Y., Yang C., Lee H. Y. Large language model as an assignment evaluator: Insights, feedback, and challenges in a 1000+ student course [Electronic resource] // arXiv preprint arXiv: 2407.05216. – 2024. – Mode of access: `https://arxiv.org/abs/2407.05216`. – Date of access: 24.10.2025
3. Wei J., Bosma M., Zhao V. Y., Guu K., Yu A. W., Lester B., Du N., Dai A. M., Le Q. V. Finetuned language models are zero-shot learners [Electronic resource] // arXiv preprint arXiv:2109.01652. – 2021. – Mode of access: `https://arxiv.org/abs/2109.01652`. – Date of access: 24.10.2025

NETWORK TRAFFIC ANALYSIS BASED ON DEEP LEARNING ALGORITHMS

Xia Enduo, Fan Linda, He Hongyan
Department of Radiophysics and Computer Technologies, Belarusian State University
Minsk, Republic of Belarus
E-mail: enduoxia@gmail.com, fld420605545@gmail.com, hehy9527@gmail.com

We propose a unified anomaly detection framework based on CNN-BiLSTM-Attention to address the low accuracy and high false positive rates of existing traffic-based network anomaly detection methods. The framework represents each network session as a 2D grayscale image, uses 2D-CNN for spatial feature extraction, BiLSTM for long-range temporal dependency capture, and an attention mechanism to highlight the most discriminative features. Experiments on the USTC-TFC-2016 and CICDDoS 2019 datasets show that this framework significantly outperforms state-of-the-art methods in accuracy, detection rate, and false positive rate, demonstrating its effectiveness and generalization for network traffic classification.

Introduction

With the rapid development of the Internet and information technology, cyberattacks pose a significant threat to network and user information security. Traditional solutions relying on past experience or localized packet information often result in false detections and false positives. To address these issues, this paper proposes a hierarchical model combining 2D-CNN with Bi-LSTM-Attention. The USTC-TFC-2016 and CICDDoS 2019 datasets were normalized to 16,000 bytes and reshaped into 40×40 grayscale images. The model uses 2D-CNN to extract spatial features from each packet image and Bi-LSTM-Attention to model long-term dependencies across the session. Experimental results show that the proposed model significantly outperforms traditional methods, with an inference latency of less than 0.8 ms per flow on an RTX 2080Ti graphics card. This confirms that reshaping packet data into 2D images and mining local patterns enhances the accuracy and robustness of real-time malicious traffic detection.

I. Related Work

Traditional ML methods like Random Forest [4], SVM [6], DT [2], and KNN [5] are widely used for intrusion detection but struggle with large-scale anomalies. Deep learning, with algorithms like CNN [1], and BiLSTM [3], offers better feature extraction. BiLSTM reduces false alarms, and the attention mechanism is gaining popularity in this field. This paper proposes a CNN-BiLSTM-Attention intrusion detection model. It combines 2D-CNN and BiLSTM to extract spatiotemporal features and uses the attention mechanism to filter significant features, thereby enhancing classification performance. Experiments on two public datasets show that the proposed model achieves superior classification results.

II. CNN-BiLSTM-Attention Model

Figure 1 shows the framework of the CNN-BiLSTM-Attention algorithm. We first segment the original data stream into conversation streams, convert each packet in the conversation stream into a grayscale image, and then concatenate them to obtain a grayscale image of the conversation stream. Next, for the conversation stream image, we use a 2D-CNN to split the conversation into a sequence of packets and perform linear embedding. An LSTM is used to extract the temporal relationship between each packet in the conversation stream. The input of each hidden layer of the LSTM serves as the final embedding of the packet sequence in the conversation. The conversation stream embedding is fed into a multi-head attention layer to obtain a feature map for each session. A CNN is then used to extract the final features of the conversation stream from this feature map. Finally, a fully connected layer is used to identify data flow.

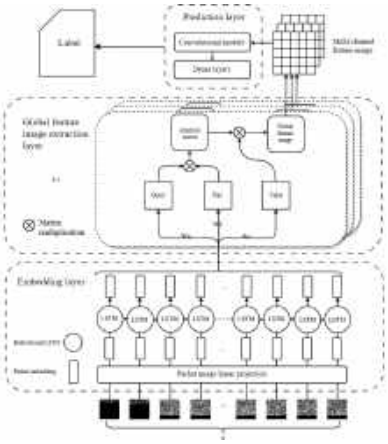


Figure 1 – Overall of CBAM

During the preprocessing phase, we focus on the payload of application layer packets and ignore other layers.The network flow is divided into sessions S based on five-tuples.

$$S = \{P_1, P_2, \ldots, P_n\} \tag{1}$$

The packet size is uniformly set to 1600 bytes, with any missing bytes padded with 0x00. The first t packets are taken, and any missing bytes are padded to t. Each packet is converted to a 40x40 grayscale

image, generating a sequence of packet images that are then merged to form the session flow image S.For example, Figure 2. In the embedding layer, a 2D-CNN is used to decompose the session image into a sequence of packet images, generating an initial embedding for each packet. Specifically, a two-dimensional convolution operation is performed on each packet image D_i to create a linear embedding:

$$E_i = W_{\text{conv}} * D_i + b_{\text{conv}} \tag{2}$$

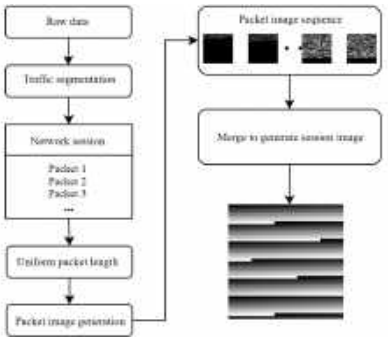


Figure 2 – Data preprocessing of CBAM

For each linear embedding E_i, a bidirectional LSTM layer is used to obtain contextual information:

$$H_i = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{3}$$

Finally, all H_i are assembled into the final output representation X of the session, where $d = 2 \times \text{lstm}_d$.he global feature map extraction layer introduces a self-attention mechanism to extract global features from the conversation flow. First, a linear transformation is performed:

$$Q,K,V = XW_q, XW_k, XW_v \tag{4}$$

Then the self-attention weight matrix A is calculated:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

The global feature map M is obtained:

$$M = A \cdot V \tag{6}$$

Furthermore, a multi-head self-attention mechanism is used to project the query, key, and value matrices into h subspaces, generating multiple feature maps Z_i:

$$Z_i = \text{softmax}\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right)V_i \tag{7}$$

These feature maps are then concatenated into a three-dimensional tensor Z. At the prediction layer, a convolutional neural network is used to extract features from the feature map Z. First, a convolution operation is performed:

$$F' = \sigma(W * Z + b) \tag{8}$$

Then global average pooling is performed:

$$v = \frac{1}{t' \cdot m'} \sum_{i=1}^{t'} \sum_{j=1}^{m'} F'_{i,j} \tag{9}$$

Finally, classification is performed:

$$\hat{y} = \text{softmax}(W' \cdot v + b') \tag{10}$$

III. Experimental analysis

To validate the generalization performance of our experimental model, this study employed a widely used public network traffic dataset. In our experiments, we primarily evaluated our experimental model and compared it with several advanced flow classification methods. These methods employ different feature extraction and classification techniques. As shown in Figure 3, our experimental model outperforms other models across all metrics, demonstrating its strong generalization capabilities across various encryption scenarios.

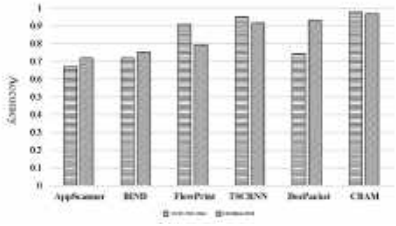


Figure 3 – Accuracy of different models

IV. Conclusion

The CNN-BiLSTM-Attention model uses self-attention to extract global network flow information and LSTM to embed packets, addressing the temporal limitations of attention mechanisms. This design enhances the model's ability to capture packet relationships and temporal information, yielding more stable and comprehensive features and improving classification accuracy. Future work will focus on increasing CNN-BiLSTM-Attention's classification speed and efficiency, adding more network flow statistical features, and exploring model interpretability.

V. References

1. Chen G., Su J., Abnormal Traffic Detection Algorithm Based on Deep Neural Network[J]. Netinfo Security, 2019, 19(6): 68–75. DOI: 10.3969/j.issn.1671-1122.2019.06.0068-08.
2. Li Qiang, Yan Cheng-hua, Zhu Yao. Analysis and Detection of Network Traffic Anomaly Based on Decision Tree // Computer Engineering. – 2012. – Vol. 38, No. 5. – P. 92–95. DOI: 10.3969/j.issn.1000-3428.2012.05.027.
3. Sun H., Wang J., Wang P., An Y. L., et al. Network Intrusion Detection Method Based on Attention-BiTCN[J]. Netinfo Security, 2024, 24(2): 309–318.
4. Lin Weining, Chen Mingzhi, Zhan Yunqing, Liu Chuanbao. Research on an Intrusion Detection Algorithm Based on PCA and Random Forest Classification // Information Security. – 2017. – No. 11. – P. 50–55. DOI: 10.3969/j.issn.1671-1122.2017.11.005.
5. PNess S., Eswarakrishnan V., Sridharan H., Shinde V., Janapareddy N.V.P., Dhanawat V. Anomaly Detection in Network Traffic Using Advanced Machine Learning Techniques [J]. IEEE Access, 2025, 13: 16133–16149. DOI: 10.1109/ACCESS.2025.3526988.
6. Moore A. W., Zieve D. Internet Traffic Classification using Bayesian Analysis Techniques[J]. Published in 2005.

Секция «Проектирование встраиваемых систем»

Section «Embedded Systems Design»

ИССЛЕДОВАНИЕ ВРЕМЕНИ ВЫПОЛНЕНИЯ ЧАСТИЧНОЙ РЕКОНФИГУРАЦИИ FPGA XILINX

Босько Н. В., Луцик Ю. А.
Кафедра электронных вычислительных машин,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: nadezhdabosko@gmail.com

В данной работе представлены результаты исследования времени, затрачиваемого на выполнение частичной реконфигурации на FPGA Xilinx. Представлены результаты измерения времени реконфигурации для ряда реконфигурируемых модулей с различными характеристиками. Проанализирована зависимость данного времени от размера используемых частичных битстримов, а также размера битстримов от размера заданной реконфигурируемой области.

Введение

Способность FPGA к реконфигурации – изменению реализуемой ресурсами кристалла FPGA логической схемы – является их определяющим свойством. Под функцией частичной реконфигурации понимается возможность изменения отдельной части реализуемой на FPGA логической схемы, в то время как остальная, статическая её часть остается неизменной и продолжает работать.

Для использования частичной реконфигурации предварительно задаются области кристалла, доступные для частичной реконфигурации (реконфигурируемые области) и разрабатываются все доступные для загрузки в них реконфигурируемые модули. Для выполнения частичной реконфигурации используются частичные битстримы (файлы для конфигурации FPGA), каждому из которых соответствует определенный реконфигурируемый модуль.

При разработке динамически реконфигурируемых устройств на FPGA важно иметь представление о времени, затрачиваемом на выполнение частичной реконфигурации, и влияющих на него факторах. Это позволяет принять наиболее оптимальные решения о характеристиках предполагаемых к реконфигурации модулей.

I. Описание исследования

Исследование было проведено с использованием FPGA Xilinx XC7A100TFGG484-2. В терминологии Xilinx технология динамической частичной реконфигурации имеет название Dynamic Function eXchange (DFX) [1]. В [1] указано, что время частичной реконфигурации напрямую зависит от размера используемого частичного битстрима. В данном исследовании была проанализирована указанная зависимость.

Частичная реконфигурация выполнялась с помощью IP-ядра DFX Controller [2], подключенного к ICAP (Internal Configuration Access Port). Используемая в проекте тактовая частота составляла 100 МГц, что является максимальной тактовой частотой для ICAP [1]. Частичные битстримы для всех реконфигурируемых модулей были предварительно записаны во внешнюю DDR3-память, доступную IP-ядру DFX Controller.

Выполнение частичной реконфигурации инициировалось путем подачи на DFX Controller сигналов-триггеров, соответствующих необходимым для загрузки модулям. Данные сигналы контролировались с помощью VIO (Virtual Input-Output). Измерение времени частичной реконфигурации производилось с помощью пользовательского IP-ядра, сохраняющего количество тактов, на протяжении которых был активен заданный сигнал. В данном случае измерение проводилось с использованием сигналов vsm_<name>_rm_-decouple от DFX Controller, которые переходят в состояние «1» по началу реконфигурации и возвращаются в состояние «0» по ее завершению [2]. Стоит отметить, что в проведенных экспериментах было использовано сжатие частичных битстримов.

На кристалле были заданы три реконфигурируемые области (РО) разных размеров, содержащие указанное в таблице I количество основных ресурсов.

Таблица 1 – Количество ресурсов в исследуемых РО

РО	Slice LUTs	Slice regs	Slice	DSP
A	6856	13712	1714	16
B	1152	2592	288	0
C	160	320	40	0

Для РО «А» было создано четыре реконфигурируемых модуля (РМ): A1-A4, для РО «В» и РО «С» – по два: B1-B2 и C1-C2. Используемые для проведения исследования модули реализуют различные арифметические операции. В таблице 2 указаны используемые данными модулями ресурсы, а также размеры в байтах частичных битстримов, сгенерированных для каждого из модулей.

II. Результаты исследования

Было измерено количество тактов, затрачиваемое на выполнение частичной реконфигурации для каждого из модулей. Измерение для

каждого модуля было проведено 3 раза, после чего было определено среднее значение. Результаты измерений, переведенные в наносекунды, представлены в таблице 3 (средние значения были округлены).

По результатам измерений можно увидеть линейную зависимость между размером частичного битстрима реконфигурируемого модуля и временем, требуемом для его загрузки. График данной зависимости представлен на рисунке 1. Точками отмечены измеренные значения.

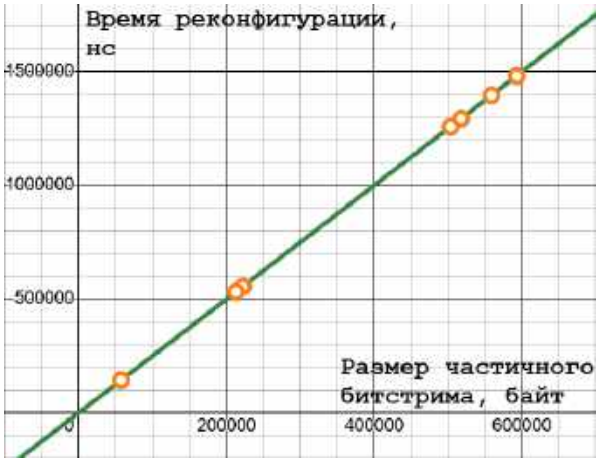


Рис. 1 – График зависимости времени реконфигурации от размера битстрима

Реконфигурация начинается тогда, когда начинается передача частичного битстрима на ICAP, и завершается, когда весь битстрим был передан. Соответственно, время реконфигурации увеличивается линейно при доставке на ICAP больших объемов данных. Примерная формула зависимости: $y = 2{,}50001x + 666{,}2277$.

С помощью данной зависимости можно получить представление о времени, требуемом на реконфигурацию модуля, размер битстрима которого известен. На время реконфигурации также могут оказывать влияние другие факторы, такие как, например, задержки при чтении данных из внешней памяти IP-ядром DFX Controller.

III. Размер битстрима

Так как время частичной реконфигурации напрямую зависит от размера используемого частичного битстрима, следует обратить внимание также на факторы, от которых зависит последний. Можно заметить, что реконфигурируемые модули области «A» используют разные объемы ресурсов. Однако при этом размеры частичных битстримов модулей примерно схожи в пределах одной РО, и значительно различаются между разными РО.

Из этого можно сделать вывод, что размер частичного битстрима модуля напрямую зависит от размера реконфигурируемой области, в которой он используется, даже если в действительности реконфигурируемый модуль занимает только малую часть данной реконфигурируемой области.

IV. Выводы

Из результатов проведенного исследования можно сделать вывод, что при планировании реконфигурируемых областей и разработке реконфигурируемых модулей являются нежелательными ситуации, в которых в РО используются модули, занимающие лишь малую часть их ресурсов. При этом для сокращения затрачиваемого на частичную реконфигурацию времени размеры реконфигурируемых областей следует минимизировать.

1. AMD. Vivado Design Suite User Guide: Dynamic Function eXchange [Electronic resource]. – Mode of access: `https://docs.amd.com/r/en-US/ug909-vivado-partial-reconfiguration/`. – Date of access: 23.10.2025.
2. AMD. Dynamic Function eXchange Controller Product Guide [Electronic resource]. – Mode of access: `https://docs.amd.com/v/u/en-US/pg374-dfx-controller/`. – Date of access: 23.10.2025.

ГЕНЕРАЦИЯ СЛУЧАЙНЫХ ЧИСЕЛ НА ОСНОВЕ СХЕМЫ ФНФ КОНФИГУРИРУЕМОГО КОЛЬЦЕВОГО ОСЦИЛЛЯТОРА

Бурко Л. А., Иванюк А. А.
Кафедра информатики, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: burkoliana@gmail.com, ivaniuk@bsuir.by

В статье исследуется поведение физически неклонируемых функций (ФНФ), реализованных на основе конфигурируемого кольцевого осциллятора (ККО). Такие структуры позволяют эффективно решать задачи аппаратной идентификации и генерации случайных чисел при минимальных затратах вычислительных и схемотехнических ресурсов. Представленные результаты показывают, что поведение ФНФ на базе ККО обладает выраженной стохастической природой, что подтверждает их пригодность для использования в системах, требующих высокой степени случайности и устойчивости к клонированию.

Введение

Современные информационные системы предъявляют всё более высокие требования к безопасности, аутентификации и защите данных. Одним из перспективных направлений в этой области является использование физически неклонируемых функций (ФНФ) [1], основанных на естественных вариациях физических параметров микросхем. Особый интерес представляют ФНФ, реализованные на основе конфигурируемых кольцевых осцилляторов (ККО), так как они имеют простую аппаратную реализацию, обладают высокой скоростью работы и способны формировать устойчивые уникальные отклики. Такие структуры позволяют создавать уникальные аппаратные идентификаторы, которые невозможно воспроизвести даже при полном копировании технологии изготовления устройства. Так же, помимо задач идентификации [2], ФНФ на базе ККО могут служить источником аппаратной случайности, что делает их эффективными для применения в генераторах случайных чисел и криптографических протоколах.

I. Описание эксперимента

Схема эксперимента (рисунок 1) более подробно уже была описана в предыдущем исследовании[2].

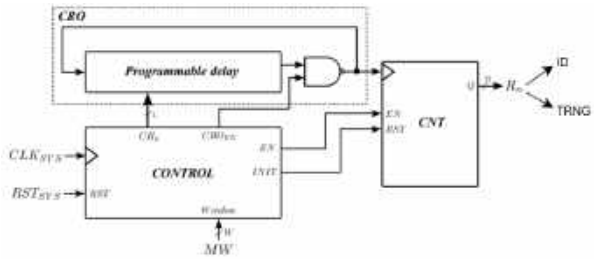


Рис. 1 – Схема эксперимента

Для сбора данных в эксперименте использовалась плата быстрого прототипирования Digilent ZYBO Z7-10 (Xilinx Zynq-7000). ФНФ ККО формирует уникальный ответ R_n, характеризующийся различиями в последовательности выходных битов. При многократном количестве измерений счетчика, для всех значений можно выделить стабильную и нестабильную группу. На рисунке 2 отображены битовые вероятности поразрядно: черным цветом – $P_1 = 1$, белым – $P_1 = 0$, серым – $0{,}45 < P_1 < 0{,},55$, где P_1 – вероятность появления 1 на определенном разряде.

Рис. 2 – Битовые вероятности при различных CH_n

Стабильную часть можно использовать в качестве уникальных идентификаторов (значения красного цвета можно стабилизировать по мажоритарному принципу). Как показано на рисунке 2, при одинаковой прошивке ПЛИС, но при различных CH_n, получаются каждый раз различные идентификаторы. Схожесть имеют CH_0 и CH_f, но идентификатор для CH_f на 1 бит длиннее, тем самым видно, что все идентификаторы различны.

II. Уникальность ответов ФНФ

Нестабильная часть (биты серого цвета на рисунке 2) обладает свойствами равномерного распределения. Помимо этого, было выдвинуто предположение, что значения со счетчика являются уникальными и неповторимыми. Для проверки этого предположения были собраны 4 сета данных после перезапусков платы. Битовая последовательность собирается путем конкатенации значений счетчика, где $P_1 \approx 0{,}5$. Для графического отображения был использован Random Walk

Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
https://www.bsuir.by/ru/kaf-informatiki/yadro

2D (RW2) тест, для которого битовая последовательность делится на пары битов, и в зависимости от этого строится траектория движения («00» – влево, «01» – вверх, «10» – вправо, «11» – вниз). Результаты теста RW2 представлены на рис. 3. Видно, что все 4 траектории различны.

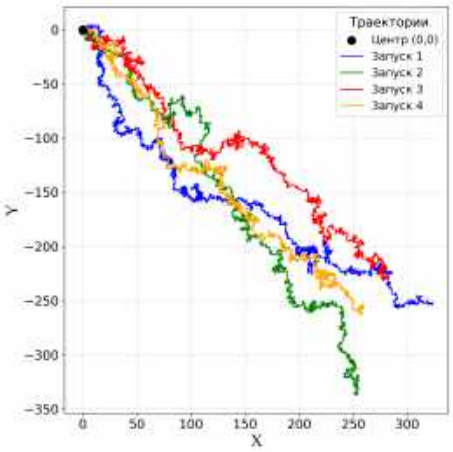


Рис. 3 – Траектории RW2 теста для различных перезапусков

III. Равномерность данных от ФНФ

Как видно из RW2 теста, преобладают подпоследовательности «11» и «10». Для визуальной проверки равномерности данных можно построить графический тест распределения точек на плоскости, для которого битовая последовательность делится на числа a_0,a_1,a_2 ... по 8 бит (количество бит может быть любое) и далее отображаются точки с координатами (a_0,a_1),(a_1,a_2),.... На рисунке 4 (слева) видно, что для исходных данных со счетчика виден особый паттерн, но в дальнейшем, с примененением постобработки, от паттерна можно избавиться (рис. 4 (справа)). При проверке последовательностей на Health Tests из пакета NIST[3], прошел только Rank-тест. Тем самым, можно сделать вывод, что исходные данные от ФНФ ККО не могут являться хорошим источником случайности.

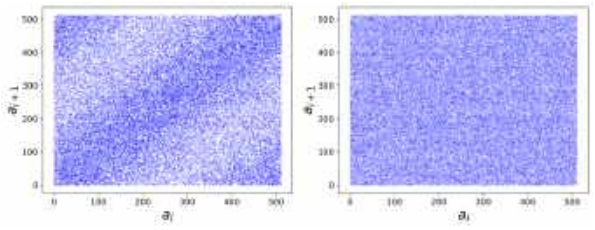


Рис. 4 – Результаты теста на распределение значений на плоскости

Разряды значений с вероятностью близкой к 0,5 (на рис.2 это разряды [7-0]) могут быть хорошей основой для источника случайных чисел. Для этого можно применить технологию постобработки, построенную на принципе регистра сдвига с линейной обратной связью.

Постобработка данных выполнялась с использованием схемы MISR (Multiple-Input Signature Register, многовходового регистра сигнатуры). Данная схема предназначена для объединения нескольких входных потоков данных (в рассматриваемом случае это разряды [7:0]) с целью формирования итоговой сигнатуры, являющейся свернутым представлением совокупного результата всех предшествующих операций. В качестве порождающего полинома использовался полином «1A7», соответствующий выражению $\varphi(x) = 1+x+x^2+x^3+x^6+x^8$. Как видно из RW2 теста на рисунке 5, данные после постобработки обладают достаточной визуальной случайностью.

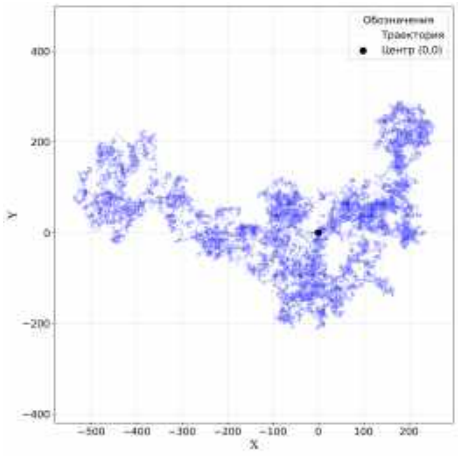


Рис. 5 – RW2 тест для значений с постобработкой

После постобработки прошли 12 из 13 тестов из пакета NIST и все 148 неперекрывающийся шаблонов. Не прошел тест Universal. Предполагается, что для его прохождения необходим больший объем данных.

IV. Вывод

В результате исследования было показано, что ФНФ ККО, при минимальном использовании аппаратных ресурсов, является хорошим источником идентификации, и при применении постобработки, может выступать надежным, уникальным источником случайных чисел, что подтверждают NIST-тесты и графическое отображение.

V. Список литературы

1. Иванюк, А. А. Исследование физически неклонируемой функции конфигурируемого кольцевого осциллятора / А. А. Иванюк // Информатика. – 2025. – Т. 22, № 1. – С. 73–89. – DOI: 10.37661/1816-0301-2025-22-1-73-89.
2. Unclonable Identification and True Random Number Generation Based on CRO PUF Pattern Recognition and Information Processing (PRIP'2025), Ivaniuk A, Burko L., Proceedings of the 17th International Conference, 16–18 Sept. 2025, Minsk, Belarus. – Minsk : UIIP NASB, 2025. – 496 p.
3. NIST 800-90B, Recommendation for the Entropy Sources Used for Random Bit Generation / `https://doi.org/10.6028/NIST.SP.800-90B`

ВРЕМЕННАЯ МОДЕЛЬ RS-ЗАЩЁЛКИ, РЕАЛИЗОВАННОЙ КАК ИСТОЧНИК СЛУЧАЙНОСТИ НА ПЛИС ТИПА FPGA

Кайкы М. Н., Иванюк А. А.
Кафедра информатики, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: kaikymykhailo@gmail.com, ivaniuk@bsuir.by

Исследуется метастабильность RS-защёлок с различными значениями технологических задержек на ПЛИС типа FPGA как источник энтропии для генераторов случайных чисел. Показывается зависимость вероятности перехода RS-защёлки в метастабильное состояние от введенного коэффициента симметрии.

Введение

Разработка генераторов истинно случайных чисел (ГИСЧ) остается актуальной задачей в области криптографии и защиты информации. Одним из перспективных физических источников энтропии для ГИСЧ является метастабильное состояние цифровых элементов, в частности, RS-защелок, где бистабильная система хаотическим образом сходится к одному из двух устойчивых состояний [1]. Однако использование данного явления сопряжено с фундаментальной проблемой: метастабильность изначально рассматривалась в цифровой схемотехнике как нежелательное и аномальное явление, что обусловило недостаточность существующих функциональных и временных моделей для ее точного прогнозирования и анализа. Стандартные детерминистические модели на языках описания аппаратуры (HDL) не способны адекватно описать вероятностную природу и временные параметры метастабильных процессов. В связи с этим, необходимость разработки комплексной методики функционального моделирования метастабильности в RS-защелках становится важной.

I. Реализация RS-защёлки на технологии FPGA

На рисунке 1 изображена структурная схема разработанного аппаратного обеспечения для проведения исследования. Для реализации структур типа RS-защёлка на базе примитивов FPGA от компании Xilinx семейства Zynq7000 используется два блока типа 6-ти входовой таблицы истинности (генераторов переключательной функции), связанных друг с другом при помощи обратной связи с образованием комбинационной петли и выполняющих функцию "И-НЕ"(*NAND2*). При этом два LUT6 абсолютно идентичны друг другу функционально, количество защёлок равнялось $N = 1024$. Для формирования управляющих сигналов на входы Set/Reset защёлок был разработан управляющий автомат, при этом стоит отметить, что управляющих регистров в схеме всего два – один для формирования сигнала Set, второй для формирования сигнала Reset, что позволяет сформировать 2048 технологически уникальных задержек распространения управляющих сигналов до каждого из экземпляров RS-защёлок (δ_{13}, δ_{23}).

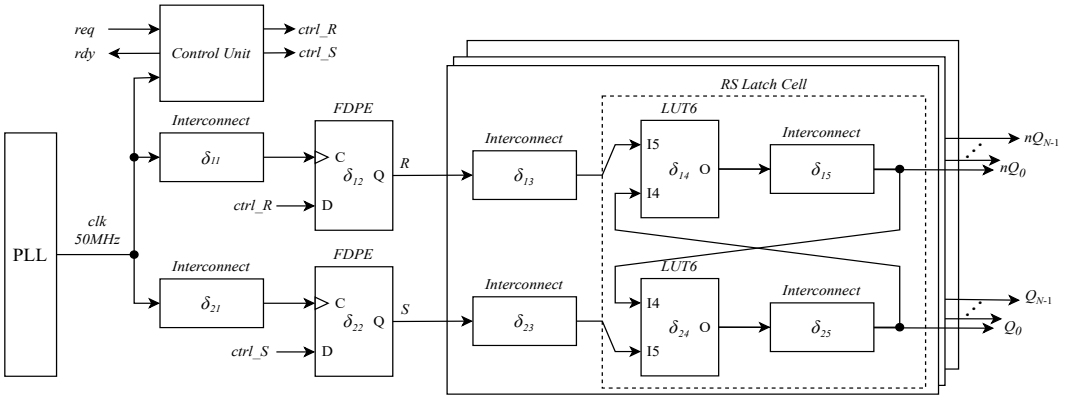


Рис. 1 – Структура стенда для исследования RS-защёлок с задержками на кристалле ПЛИС

Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
https://www.bsuir.by/ru/kaf-informatiki/yadro

II. Временное моделирование RS-защёлок

В ходе проведения моделирования разрабатываемой схемы после процесса имплементации на кристалле FPGA в среде моделирования QuestaSim были получены значения уникальных задержек компонентов схемы ($\delta_{11}, \delta_{12}, \delta_{13}, \delta_{14}, \delta_{15}, \delta_{21}, \delta_{22}, \delta_{32}, \delta_{24}, \delta_{25}$). В процессе моделирования на все 1024 защёлки подавалась последовательность входных воздействий, которая приводила защёлки в последовательность состояний (*Reset*, *Store*, *Forbidden*, *Store*), которое способно привести RS-защёлку в метастабильное состояние. В результате моделирования было получено 1024 различных и уникальных значения счётчиков C, которые фиксировали количество нарастающих и спадающих фронтов на выходе Q для всех 1024 защёлок и были реализованы в тестовом окружении, обладая нулевыми задержками переключения. При $C = \{0, 2\}$ – мы считаем что RS-защёлка стабильно перешла из состояния *Forbidden* в состояние *Store* и сохранила состояние 0 или 1, а при $C > 2$ наблюдается осцилляция на выходе защёлки Q. Всего было зафиксировано 76,85% стабильных защёлок и 23,15% нестабильных. При этом стоит отметить, что при фиксированном времени моделирования, не все значения C были идентичны друг другу, что говорит о разном периоде осцилляции на выходах RS-защёлок. Также было обнаружено что некоторые нестабильные RS-защёлки переходили в устойчивое состояние через некоторое время моделирования, соответственно осцилляция на выходе являлась затухающей, такое же поведение защёлок приведено при КМОП реализации в работе [2].

III. Влияние симметрии в RS-защёлке на её стабильность

Введем коэффициент симметрии (K), который будем рассчитывать по формуле 1.

$$K = \frac{\delta_{upper} - \delta_{lower}}{max(\delta_{upper}, \delta_{lower})}, \quad (1)$$

где $\delta_{upper} = \delta_{11} + \delta_{12} + \delta_{13} + \delta_{14} + \delta_{15}$,
$\delta_{lower} = \delta_{21} + \delta_{22} + \delta_{23} + \delta_{24} + \delta_{25}$.

Результат рассчёта коэффициентов K для всех 1024 RS-защёлок изображен на рисунке 2, при этом синим цветом отображены защёлки, ответы которых были стабильны, а красным – нестабильные защёлки с осцилляцией на выходе Q. Как видно из риснука 2, при коэффициенте K стремящемся к 0 количество осциллирующих (нестабильных) RS-защёлок увеличивается, при этом нестабильные защёлки фиксировались на промежутке $K = [0,121; -0,151]$, полученная зависимость позволяет сделать вывод: способность RS-защёлки переходить в состояние осцилляции на выходах Q, nQ – зависит от степени симметрии при построении RS-защёлки и схемы её управления.

IV. Заключение

В ходе исследования был введен коэффициент симметрии в RS-защёлках, разработанных по технологии FPGA, было обнаружено что при коэффициенте симметрии стремящегося к 0 – вероятность перехода в состояние метастабильности RS-защёлки повышается. Дальнейшие работы направлены на исследование изменения коэффициента симметрии от технологических вариаций и эмпирическое доказательство полученных результатов на реальных экземплярах FPGA.

V. Список литературы

1. Sunar, B. A Provably Secure True Random Number Generator with Built-In Tolerance to Active Attacks / B. Sunar, W. Martin, D. Stinson // Computers, IEEE Transactions on. – 2007. – Vol. 56. – P. 109–119.
2. Kacprzak, T. Analysis of Oscillatory Metastable Operation of an RS Flip-Flop / T. Kacprzak // Solid-State Circuits, IEEE Journal of. – 1988. – Vol. 23. – P. 260–266.

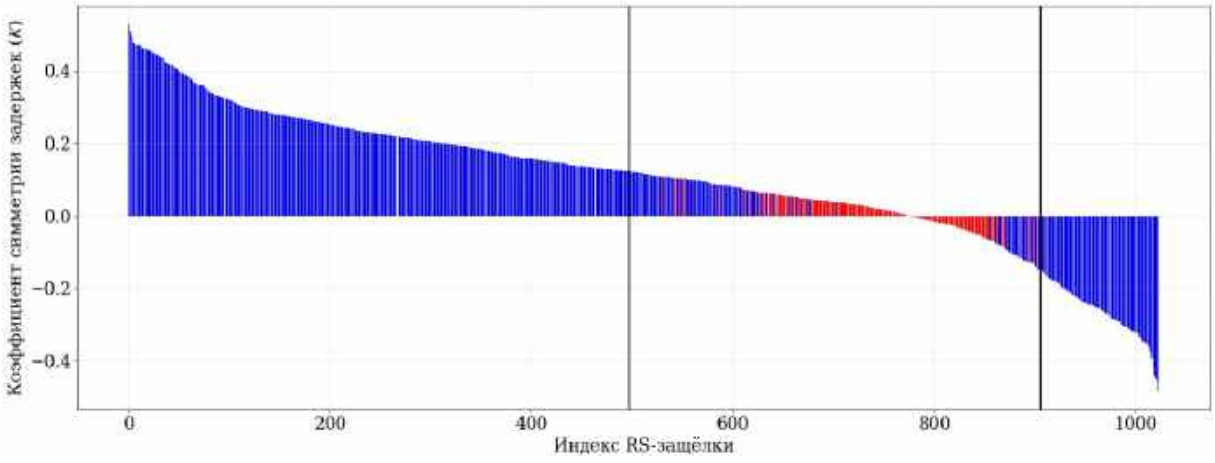


Рис. 2 – Зависимость стабильности RS-защёлок от коэффицинта симметрии

ИССЛЕДОВАНИЕ ЭФФЕКТОВ КВАНТОВАНИЯ ПАРАМЕТРОВ НЕЙРОННОЙ СЕТИ НА ОСНОВЕ ОБУЧАЕМОГО РАЗДЕЛИМОГО ПРЕОБРАЗОВАНИЯ

Кривальцевич Е. А., Вашкевич М. И.
Кафедра электронных вычислительных средств,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: krivalcevi4.egor@gmail.com, vashkevich@bsuir.by

Представлена реализация нейронной сети на основе обучаемого двумерного разделимого преобразования для распознавания рукописных цифр на базе платформы Zynq-7000. Выполнен анализ влияния точности представления параметров нейронной сети на её производительность, а также на требуемые для реализации аппаратные ресурсы ПЛИС.

Введение

Нейронные сети (НС) являются базовыми компонентами многих технических систем. Обычно для обучения и исполнения НС используются графические процессоры, однако эти устройства не могут использоваться в контексте встраиваемых систем поскольку имеют большое энергопотребление. В настоящее время программируемые логические интегральные схемы (ПЛИС) типа FPGA рассматриваются, как перспективные вычислительные платформы для реализации НС, которые способны с одной стороны обеспечить высокую производительность и приемлемое для встраиваемых систем энергопотребление [1]. Одной из особенностей ПЛИС является ограниченный объем памяти для хранения параметров НС, по этой причине важной научной задачей является разработка архитектур НС, которые способны обеспечить высокую производительность при использовании малого числа параметров.

Традиционно ускорители НС на FPGA строятся с использованием архитектуры на основе многослойного перцептрона или сверточной нейронной сети [1]. Недавно также был предложен вариант архитектуры НС на основе обучаемого двумерного разделимого преобразования (ОДРП), особенностью которой является малое число параметров и высокая точность [2]. В настоящей работе проводится исследование влияния квантования параметров НС на основе ОДРП на точность распознавания и аппаратные затраты FPGA.

I. Разработка НС на основе ОДРП

В работе рассматривается задача распознавания изображений рукописных цифр из набора данных MNIST. Для проведения эксперимента была выбрана НС LST-1 [2], состоящая из слоя ОДРП (на рис. 1 оно обозначено как LST – *learned separable transform*) и полносвязного слоя с выходной функцией активации softmax.

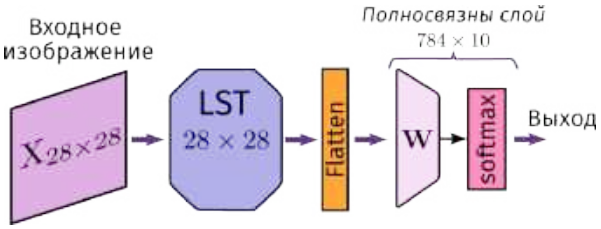


Рис. 1 – Структура LST-1 [2]

ОДРП принимает на вход изображение $\mathbf{X}$ размера $d_{in} \times d_{in}$ и формирует выходное изображение $\mathbf{Y}$ размера $d_{out} \times d_{out}$ путем обработки по строкам и столбцам двумя различными полносвязными слоями:

$$\mathbf{Y} = \mathrm{LST}(\mathbf{X}) = \sigma(\mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{X}^T + \mathbf{b}_1) + \mathbf{b}_2) \quad (1)$$

где $\sigma(\cdot)$ – функция активации, $\mathbf{W}_1$,$\mathbf{W}_2$ – матрицы весовых коэффициентов линейных преобразований полносвязных слоев для обработки строк и столбцов соответственно, $\mathbf{b}_1$,$\mathbf{b}_2$ – смещения весовых коэффициентов.

Для реализации НС была выбрана отладочная плата ZYBO на базе ПЛИС семейства Zynq-7000. НС была реализована в виде IP-блока с интерфейсом AXI-Stream, по которому от ARM процессора на вход НС подавались изображения, ответы НС также передавались по интерфейсу AXI-Stream.

II. Тестирование НС

НС LST-1 показала эффективность применения двумерного разделимого преобразования при решении задач распознавания рукописных цифр. Основное достоинство сети данного типа заключается в достижении высокой точности распознавания, при значительно меньшем количестве параметров в сравнении с другими полносвязными архитектурами [2].

При аппаратной реализации НС выполняются вычисления в формате с фиксированной

Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
https://www.bsuir.by/ru/kaf-informatiki/yadro

запятой. В процессе квантования весовых коэффициентов($\mathbf{W}_1$,$\mathbf{W}_2$,$\mathbf{W}_o$) и смещений ($\mathbf{b}_1$,$\mathbf{b}_2$,$\mathbf{b}_o$) происходит снижение точности распознавания и аппаратных затрат.

В эксперименте разрядность дробной части коэффициентов НС изменялась от 2 до 11 бит. Для каждой разрядности производилась подача на НС всех 10 тыс. тестовых изображений базы MNIST. На основании отчетов после имплементации синтезатора Vivado 2024 было выбрано 2 основных измеряемых параметра ПЛИС – LUT и Flip-Flop. Полученные результаты экспериментов представлены на рис. 2, где на одном графике совмещены точность распознавания и количество использованных элементов LUT и FF в зависимости от разрядности коэффициентов НС.

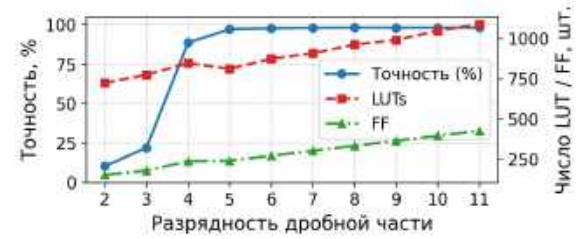


Рис. 2 – Точность и аппаратные затраты НС

Из рис. 2 видно, что точность более 97% достигается начиная с 5 бит в дробной части.

На рис. 3 представлены распределения весов для каждого слоя модели LST-1 и рассчитанное расстояние Кульбака-Лейблера между квантованными и исходными весами.

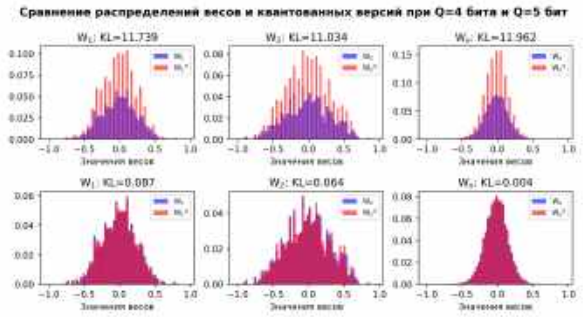


Рис. 3 – Распределения весов модели LST-1 для 4 и 5 бит в дробной части

Из рис. 3 видно, что распределение весов во всех слоях имеет форму, близкую к нормальному распределению, с центром вблизи нуля. После перехода от 5 бит к 4 битам расстояние между распределениями увеличивается и модель начинает хуже работать, её точность падает, что подтверждается экспериментальными результатами (рис. 2).

Для того, чтобы отобразить входные данные и выход ОДРП на двумерной плоскости, сохраняя при этом локальные взаимосвязи между объектами, используется метод t-SNE [3]. Точки, расположенные близко друг к другу, соответствуют изображениям, имеющем схожие признаки.

На рис. 4 представлен тестовый набор MNIST. Можно заметить, что классы цифр частично разделяются, однако между многими группами наблюдается значительное перекрытие. Это связано с тем, что исходные данные представлены в виде значений пикселей, и различия между цифрами ещё не выделены в виде устойчивых признаков.

Рис. 4 – t-SNE визуализация для тестовых изображений MNIST

После прохождения данных через ОДРП (рис. 5) видно, что классы, соответствующие различным цифрам, становятся значительно более компактными и отчётливо разделёнными. Стоит отметить, что цифры 9, 5 и 7 также имеют значительное перекрытие.

Рис. 5 – t-SNE визуализация для тестовых изображений MNIST после применения ОДРП

В соответствии с полученными результатами можно сделать вывод, что наиболее оптимальной разрядность будет 5 бит, которая позволяет с высокой вероятностью правильно распознать цифры на изображении и не использовать избыточные аппаратные ресурсы FPGA.

III. Заключение

В работе исследовано влияние разрядности весовых коэффициентов НС на точность распознавания и аппаратные затраты FPGA, а также обоснована причина наблюдаемых эффектов квантования.

IV. Список Литературы

1. Xiao R., Shi J., Zhang C. FPGA implementation of CNN for handwritten digit recognition // 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). – P. 1128-1133.
2. Vashkevich M., Krivalcevich E. Compact and Efficient Neural Networks for Image Recognition Based on Learned 2D Separable Transform // 2025 27th International Conference on Digital Signal Processing and its Applications (DSPA). – P. 1-6.
3. Van der Maaten L., Hinton G. Visualizing Data using t-SNE // Journal of Machine Learning Research. – 2008. – Vol. 9. – P. 2579–2605.

IP-КОМПОНЕНТ КОНВЕЙЕРНОГО УСКОРИТЕЛЯ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ СНК СЕРИИ ZYNQ-7000

Осипов А. С., Петровский Н. А.
Кафедра электронных вычислительных средств[1],
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: axelosip3345@gmail.com, nick.petrovsky@bsuir.by

В работе представлена аппаратная реализация IP-ядра ускорителя вычисления нейронных сетей на базе ПЛИС типа FPGA серсии Zynq-7000. Вычислитель реализован в виде конвейрного процессорного ядра с микропрограммным управлением, поддерживаются операции перемножения матриц и свертки с заданными параметрами. Проведён синтез для кристалла макетной платы Zybo-Z7 сверточных архитектура на основе вещественно значимых и кватернионовых нейронных сетей.

Введение

Свёрточные нейронные сети (CNN) являются фундаментальным компонентом глубокого обучения и лежит в основе многих мультимедийных приложений. Однако высокая вычислительная сложность требует энергоэффективной для носимых устройств. Проектирование ускорителей для FPGA [1] позволяет обеспечить низкую латентность системы, так и оценить энергоэффективность в другом технологическом базисе.

I. Архитектура сопроцессора

Для реализации были выбраны основные типы слоёв: полносвязный, сверточный, max-pooling и flatten для упрощения адресации. Помимо этого предусмотрены отдельные стадии для приема и передачи данных по интерфейсам AXI-Stream.

Для хранения весов используется внешняя DDR-память FPGA из-за большого объёма данных. Ввиду синхронным операциям чтения и записи, а также с целью сокращения критического пути, применена двухстадийная конвейерная архитектура. Первая стадия посвящена формированию управляющих сигналов и запроса на чтение памяти, вторая – вычисление операции MAC (Multiply and Accumulate).

Структурная схема конвейера с внутренними модулями приведена на рисунке 1. Устройство управления отвечает за координацию внешних интерфейсов: формирование адресов и установку управляющих сигналов вычислительному модулю. Для предотвращения конфликтов сигналы задерживаются на один такт при передаче между стадиями.

Вычислительный модуль помимо операции MAC выполняет функцию $max(x_{max},x)$ для слоя *pool-max* и применяет функцию активации *leaky_-relu* с фиксированным наклоном 2^{-N} перед сохранением данных в RAM. Память RAM, организованная двухстранично на основе внутрикристального BRAM, хранит входные и промежуточные данные слоя. По завершении вычислений происходит логическая перестановка страниц, а выходные данные с предыдущего слоя становятся входными для следующего. Из-за ограничений BRAM чтение и запись в один такт невозможны, что требует завершить транзакцию записи перед перестановкой страниц. Для этого вводится дополнительная задержка конвейера на такт работы.

Дополнительно модуль включает вспомогательные блоки (*preprocessor* и *conditioner*) для подготовки данных и синхронизации интерфейсов.

II. Устройство управления

Для настройки и контроля сопроцессора используются Control and Status регистры (CSRs). Регистр CS обеспечивает программный сброс, указывает причину исключения и номер текущего этапа.

Группа регистров `config0-15` задаёт структуру нейронной сети: этап (fullycon, conv2d, poolmax, flatten, приём или выдача данных) и его параметры, размерность и объём входных данных. Согласованность с предыдущим слоем определяются автоматически, что позволяет сократить размерность регистров. Некорректная конфигурация сети вызывает исключение и приостанавливает работу сопроцессора до сброса.

III. Интерфейсы IP-компонента

IP-компонент включает один интерфейс AXI-Lite для регистров конфигурации, два интерфейса AXI-Stream для потоковой передачи входных и выходных данных, а также интерфейс доступа к внешней DDR-памяти на основе механизма request–acknowledge.

Стандартный протокол request–acknowledge требует установки запроса и ожидания ответа на последующем такте, что требует два такта работы на чтение данных. Для снижения латентности ме-

[1]Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
`https://www.bsuir.by/ru/kaf-informatiki/yadro`

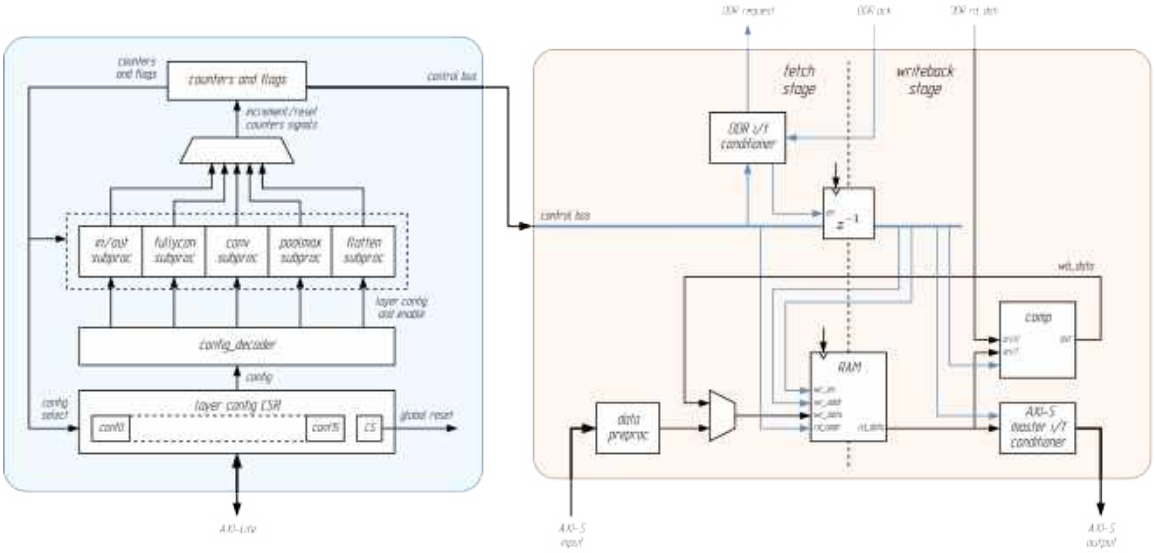


Рис. 1 – Структура сопроцессора

ханизм был модифицирован: сопроцессор может выдать следующий адрес без ожидания ответа от DDR. При своевременном отклике конвейер работает непрерывно, при задержке со стороны DDR – приостанавливается, повторно используя предыдущий адрес. Данный подход обеспечивает непрерывность работы конвейера, однако может возникнуть критический путь при прохождении сигнала acknowledge к IP и обратно.

IV. Результаты синтеза аппаратуры и заключение

В таблице IV приведены результаты синтеза двух конфигураций нейронных сетей: вещественнозначимых (real-valued) и с применением алгебры кватерниов (quaternionic) [2]. Макетирование выполнено для макетной платы Zynq Z7-10. Полученный критический путь проходит через BRAM→умножитель→аккумулятор→BRAM. Для нейронной сети на рисунке 2 расчёт латентности представлен в таблице IV; обработка одного изображения 32 × 32 занимает 77мс при $f_{CLK} = 128$ МГц.

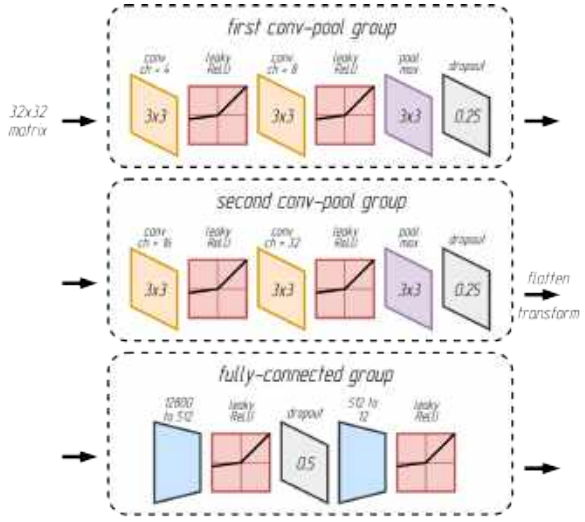


Рис. 2 – Пример структуры нейронной сети

Таблица 1 – Результаты синтеза сопроцессора для кристалла xc7z010clg400-1

Результаты синтеза	
Модель CNN	Quaternionic (4×Q16.13)
Кол-во config CSR	16
RAM size	8 × 32 × 32 words, 64-bit word wide
LUT	2222 / 17600 (12.6%)
FF	1269 / 35200 (3.6%)
BRAM	32 / 60 (53.3%)
DSP	48 / 80 (60%)
f_{CLK}	106 МГц.
Модель CNN	Real-valued (Q16.13)
Кол-во config CSR	16
RAM size	32 × 32 × 32 words, 16-bit word wide
LUT	1289 / 17600 (6.9%)
FF	1236 / 35200 (3.5%)
BRAM	32 / 60 (53.3%)
DSP	1 / 80 (1.3%)
f_{CLK}	128 МГц.

Таблица 2 – Латентность модели Real-valued

Этап	**Latency**	**Этап**	**Latency**
input 32x32	1 024	conv:3x3 ch:4	36 000
conv:3x3 ch:8	232 064	pool:3x3	48 672
conv:3x3 ch:16	672 768	conv:3x3 ch:32	2 245 760
pool:3x3	115 200	flatten	12 800
fc out:512	6 554 112	fc out:12	6 156
output	12	interm.	11
Итого			9 892 179

1. Kastner, R. Parallel Programming for FPGAs [Electronic resource] / R. Kastner, J. Matai, S. Neuendorffer // arXiv e-prints. – 2018. –Mode of access: https://arxiv.org/abs/1805.03648. – Date of access: 24.10.2025.
2. Osipov, A., Petrovsky, N. FPGA Implementation of Quaternionic Fully Connected Neural Network for Image Classification // Pattern Recognition and Information Processing (PRIP'2025): Proc. of the 17th Int. Conf., 16–18 Sept. 2025, Minsk, Belarus. – Minsk : UIIP NASB, 2025. – P. 230-234.

ИССЛЕДОВАНИЕ ХАРАКТЕРИСТИК МЕТОДА ПОСТОБРАБОТКИ СЛУЧАЙНЫХ ДАННЫХ НА БАЗЕ МНОГОВХОДОВОГО СИГНАТУРНОГО АНАЛИЗАТОРА

Петровский Д. А., Иванюк А. А.
Кафедра информатики, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: petrovsky.dmitr@gmail.com, ivaniuk@bsuir.by

В работе исследуются характеристики метода постобработки случайных данных на базе многовходового сигнатурного анализатора. Целью работы является оценка эффективности предложенной модификации метода. В результате эксперимента получено, что внесённые изменения позволяют избавиться от дублирования символов в выходной последовательности без увеличения числа источников энтропии. Это позволило повысить значение энтропии по Шеннону выходной последовательности случайных бит на 5 %.

Введение

Случайные числа играют ключевую роль в криптографии, статистическом моделировании, игровой индустрии и в других областях. Они делятся на псевдослучайные числа (ПСЧ), получаемые с помощью детерминированных алгоритмов, и истинно случайные числа (ИСЧ), в основе генерации которых лежит измерение характеристик неконтролируемых физических процессов, происходящих внутри источника энтропии (ИЭ).

Существует множество стандартов и рекомендаций, регулирующих проектирование генераторов истинно случайных чисел (ГИСЧ). В качестве примера можно привести документы, разработанные Национальным институтом стандартов и технологий США (*NIST*) [1]. Согласно рекомендациям независимый однобитный ИЭ должен удовлетворять условию «полной энтропии»[2] и иметь блок тестирования (БТ) для обнаружения критических сбоев.

ИЭ, применяемые в ГИСЧ, зачастую обладаю недостаточными статическими характеристиками выходной последовательности бит для их применения. *NIST* рекомендует использовать функции постобработки на базе криптографических примитивов, удовлетворяющих условию «полной энтропии». Однако широкое применение получили методы постобработки на основе регистра сдвига с линейной обратной связью (от англ. *Linear Feedback Shift Register*, *LFSR*). В качестве примера можно привести одновходовой сигнатурный анализатор и моговходовой сигнатурный анализатор (от англ. *Multi-input Signature Analyzer*, *MISA*).

I. Описание метода постобработки

Метод постобработки, использующий *MISA* (рис.1), для синтеза схемы и математического описания опирается на примитивные и неприводимые характеристические полиномы $\phi(x)$ над полем GF(2), где $\mathbf{D}(k)$ – n-разрядный двоичный вектор, являющийся состоянием *MISA* в k-й момент времени, $n = deg(\phi(x))$.

Рис. 1 – Обобщенная схема постобработки на базе *MISA*

В данной работе рассматривается модификация метода, заключающаяся в использовании принципа перехода схемы *MISA* из состояния $\mathbf{D}(k)$ в $\mathbf{D}(k+j)$ за один такт синхронизации, где $j > 1$ – коэффициент сжатия во времени. Схемотехническая реализация данной модификации представлена в работе [2]. Дальше будем обозначать модифицированную схему как *MISA**.

$$\mathbf{B}_i(C,j,g) = \sum_{s=1}^{j} \mathbf{C}^{s\cdot g-i} \times e_1, \quad \mathbf{A}(\mathbf{C},j,g) = \mathbf{C}^{j\cdot g}$$
$$\mathbf{D}(k+1) = A(\mathbf{C},j,g) \times \mathbf{D}(k) \oplus \sum_{i=0}^{g} \mathbf{B}_i(\mathbf{C},j,g) \times y_i(k). \tag{1}$$

Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
https://www.bsuir.by/ru/kaf-informatiki/yadro

[2]Discussion on the Full Entropy Assumption of the SP 800-90 Series
https://csrc.nist.gov/pubs/ir/8427/final

Математическая модель *MISA** описывается выражениями (1), где $\mathbf{C}$ — матрица сдвига и обратных связей *LFSR*, основанная на коэффициентах порождающего полинома $\phi(x)$, $\mathbf{A}(\mathbf{C}, j, g)$ — матрица обратной связи *MISA** с разрядностью входного слова g бит, основанная на матрице $\mathbf{C}$, возведенной в степень $j \cdot g$, что позволяет получить коэффициенты сдвига и обратной связи для коэффициента сжатия j, e_1 – вектор стандартного базиса, $\mathbf{B}_i(\mathbf{C}, j, g)$ — вектор-столбец коэффициентов, определяющий воздействия i-го бита входного слова $y(k)$ на $\mathbf{D}(k)$, получаемый в результате умножения матрицы $\mathbf{C}$, возведенной в степень, на вектор e_1, для получения первой строки матрицы $\mathbf{C}$ в качестве вектора коэффициентов.

II. Описание эксперимента

В эксперименте оценивалась зависимость изменения энтропии по Шеннону выходного потока бит от выбранного порождающего полинома $\phi(x)$, разрядности входного слова g и коэффициента сжатия j. В качестве входного набора данных был использован CIFAR-10 [3]. Схема эксперимента представлена на рис. 3.

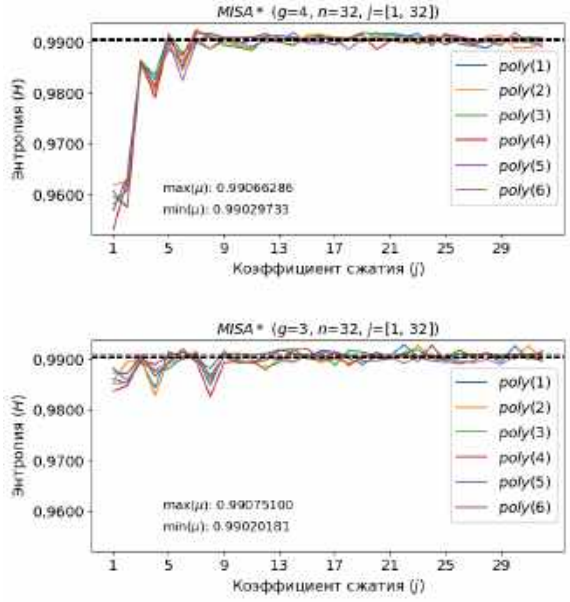


Рис. 2 – Графики зависимости энтропии по Шеннону от порождающего полинома и коэффициента сжатия

На рис. 2 представлены графики зависимости энтропии по Шеннону от порождающего полинома $\phi(x)$, разрядности входного слова и коэффициента сжатия. При этом использовались различные полиномы $\phi(x)$, заданные шестнадцатеричными значениями: {0x80000057, 0x80000062, 0x9D7826E1, 0xBD9D339, 0xFB3CFCDF, 0x80003474}, $g \in \{3, 4\}$ и $j = \overline{1, 32}$.

Из рис. 2 видно, что численное значение энтропии по Шеннону возрастает при увеличении коэффициента сжатия j. Для значений $j > 8$ при $g = 4$, а также $j > 10$ при $g = 3$, энтропии по Шеннону достигает придела, на котором наблюдается девиация матожиданий энтропии в диапазоне $[min(\mu), max(\mu)]$. Резкие провалы энтропии на графиках связаны с дублирование символов в выходной последовательности бит при малых значениях j, что видно при $HOK \neq 1$ для окна анализа, внутренней разрядности n и внутреннего сдвига $j \cdot g$. Вследствие чего можно заметить, что при достижении значения внутреннего сдвига $j \cdot g > n$, дублирование пропадет. Таким образом увеличение коэффициента сжатия j позволило получить прирост энтропии по Шеннону на 5 %, в сравнении с реализацией для $j = 1$.

III. Заключение

В ходе исследования характеристик метода постобработки на базе *MISA** модификация позволила избавиться от дублирования символов выходной последовательности, для значений $j \cdot g > n$, без увеличения количества ИЭ. Это позволило получить прирост выходной энтропии по Шеннону на 5 %, по сравнению с реализацией схемы при $j = 1$.

IV. Список литературы

1. NIST Special Publication (SP) 800–90B, Recommendation for the Entropy Sources Used for Random Bit Generation [Электронный ресурс] – режим доступа https://doi.org/10.6028/NIST.SP.800-90B. Дата доступа:24.10.2025.
2. Петровский Д. А Схема постобработки цифровой последовательности случайных чисел / Д. А. Петровский // Компьютерные системы и сети : сборник статей 60-й научной конференции аспирантов, магистрантов и студентов, Минск, 22 –26 апреля 2024 г. / Белорусский государственный университет информатики и радиоэлектроники. – Минск, 2024. – С. 729 –730
3. The CIFAR-10 dataset [Электронный ресурс] – режим доступа https://www.cs.toronto.edu/ kriz/cifar.html. Дата доступа:24.10.2025.

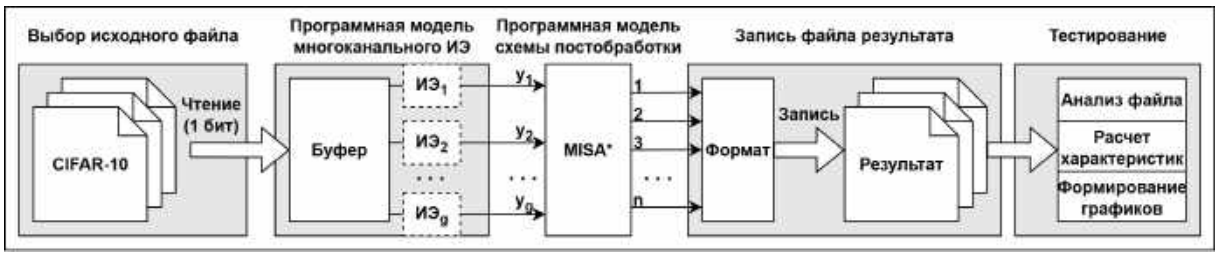


Рис. 3 – Схема эксперимента

FPGA РЕАЛИЗАЦИЯ ДВУХСЛОЙНОЙ НЕЙРОННОЙ СЕТИ ПРЯМОГО РАСПРОСТРАНЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

Субботенко О. Р., Вашкевич М. И.
Кафедра электронных вычислительных средств,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: subbotenkoolga2004@gmail.com, vashkevich@bsuir.by

В работе представлена аппаратная реализация двухслойной нейронной сети для распознавания рукописных цифр с использованием ПЛИС типа FPGA. Также проанализировано влияние точности представления параметров НС на её производительность и аппаратные затраты ПЛИС.

Введение

В настоящее время нейронные сети (НС) играют важную роль в задачах компьютерного зрения и обработки естественного языка. Зачастую для обучения и применения НС используют графические процессоры (GPU), так как они обеспечивают параллельную обработку данных. К недостаткам можно отнести высокое энергопотребление и невозможность выбора формата чисел [1].

Альтернативным решением является реализация НС на базе программируемых логических интегральных схемы (ПЛИС) типа FPGA [1-2]. К их преимуществам перед GPU можно отнести возможность выбора точности представления параметров, большая производительность и меньшее энергопотребление.

В работе представлена реализация двухслойной нейронной сети на ПЛИС для распознавания рукописных цифр из базы MNIST с использованием арифметики с фиксированной запятой. Проанализирована зависимость между разрядностью коэффициентов НС, аппаратными затратами и точностью распознавания.

I. Разработка двухслойной нейронной сети

Предлагаемая в работе НС для классификации изображений рукописных цифр имеет два полносвязных слоя. На рис. 1 приведена ее структура.

Для обучения и тестирования НС была использована база MNIST, содержащая изображения размером 28×28 пикселей рукописных цифр от 0 до 9 в оттенках серого. Она состоит из 70 тыс. изображений, из которых 60 тыс. относятся к обучающей выборке, а 10 тыс. – к тестовой. На вход НС подаются изображения, предварительно преобразованные в вектор.

Скрытый слой принимает входные данные и рассчитывает результат согласно выражению:

$$z_t^{[1]} = \sum_{s=0}^{783}(w_{t,s}^{[1]} \cdot x_s) + b_t^{[1]},$$

где $z_t^{[1]}$ – значения преактиваций скрытого слоя, $t \in [0;13]$; $w_{t,s}^{[1]}$ – веса 1-го слоя; x_s – входные значения (пиксели); $b_t^{[1]}$ – смещение t-го нейрона 1-го слоя.

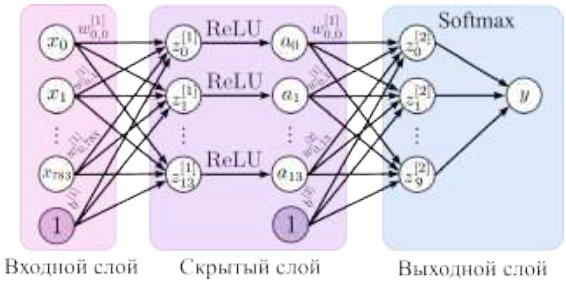


Рис. 1 – Структура двухслойной нейронной сети

Результаты вычислений поступают на вход активационной функции ReLU, которая описывается выражением:

$$f(z) = \mathrm{ReLU}(z) = \max(z,0).$$

Далее эти значения подаются на вход выходного слоя, где преобразуются в выходные по формуле:

$$z_t^{[2]} = \sum_{s=0}^{13}(w_{t,s}^{[2]} \cdot a_s) + b_t^{[2]},$$

где $z_t^{[2]}$ – значения преактиваций выходного слоя, $t \in [0;9]$; $w_{t,s}^{[2]}$ – веса 2-го слоя; a_s – выходы скрытого слоя; $b_t^{[2]}$– смещение t-го нейрона 2-го слоя.

Вычисленные 10 значений поступают на вход функции softmax:

$$\mathrm{softmax}(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^{n}\exp(z_j)},$$

где z_i – элемент входного вектора.

Обучение НС выполнялось с использованием языка Python и библиотеки PyTorch. В качестве функции потерь использовалась перекрестная энтропия, оптимизация выполнялась методом Adam со скоростью обучения $\eta = 0{,}0003$ и коэффициентом L2-регуляризации $\lambda = 0{,}0001$. Обучение проходило в 200 эпох, благодаря чему модели

Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
https://www.bsuir.by/ru/kaf-informatiki/yadro

удалось достигнуть высокой точности и избежать переобучения.

II. Реализация НС на FPGA

Для аппаратной реализации двухслойной нейронной сети была выбрана отладочная плата PYNQ Z2 на базе ПЛИС Zynq-7000. Zynq представляет собой систему на кристалле (СнК), объединяющую процессор ARM Cortex-A9 и программируемую логику FPGA. Для упрощения разработки и тестирования на данной платформе используется PYNQ (Python Productivity for Zynq). PYNQ – это специализированный дистрибутив Linux для плат на основе Xilinx Zynq и набор библиотек Python (pynq), которые предоставляют API для взаимодействия с программируемой логикой процессорной системой через IP-ядра, такие как DMA, GPIO и др.

IP-блок двухслойной НС описан на языке SystemVerilog. Полносвязные слои реализуются на базе MAC-ядер: 14 ядер находятся в скрытом слое и 10 в выходном. Каждое MAC-ядро производит операцию умножения входных значений на веса слоя, хранящиеся в блочной памяти FPGA. Функция ReLU реализуется при помощи мультиплексора, управляемого знаковым разрядом аргумента. Вместо softmax используется argmax, так как он требует меньших затрат. В блоке argmax осуществляется поиск индекса наибольшего элемента выходного вектора второго слоя. Тактовая частота IP-ядра составила 40 МГц.

Изображение из процессорной системы передается в блок IP-ядро НС по интерфейсу AXILite. По окончании вычислений результат распознавания передаётся обратно в процессорную систему по тому же интерфейсу (рис. 4).

III. Тестирование НС

Исследовалось влияние представления весов НС на ее точность и аппаратные затраты FPGA. Разрядность дробной части весов изменялась от 2 до 12. В каждом случае тестирование производилось на всех тестовых изображениях из базы MNSIT. Для анализа полученных результатов строились матрицы спутывания, одна из которых приведена на рис. 2.

С увеличением разрядности весов точность предсказаний растет. В то же время анализ аппаратных затрат FPGA показывает, что количество требуемых LUT (Look-Up Table) и FF (Flip-Flop) также растет(рис. 3).

По полученным результатам можно сделать вывод, что наиболее оптимальным является представление коэффициентов с 8 разрядами в дробной части. В этом случае точность достигает довольно высокого значения (94,24%), а аппаратные затраты принимают приемлемые значения.

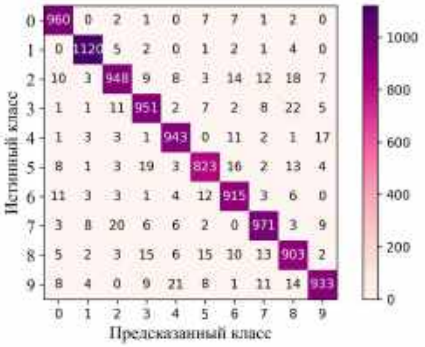


Рис. 2 – Матрица спутывания для коэффициентов НС с 10-разрядным представлением дробной части

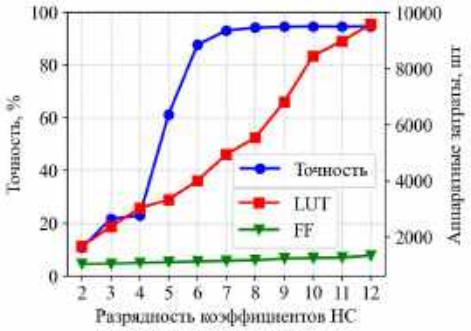


Рис. 3 – Точность и аппаратные затраты НС

IV. Заключение

В работе описан способ реализации двухслойной НС для распознавания рукописных цифр на базе платформы PYNQ Z2. Также была исследована зависимость между форматом представления весов НС и аппаратными затратами FPGA.

1. Sayed R., Draz H. H., Azmi H. HW-SW co-design of image classification accelerator on FPGA // The Journal of Supercomputing. – 2025. – Т. 81. – №. 8. – P. 1-26.
2. Кривальцевич Е. А., Вашкевич М. И. Исследование аппаратной реализации нейронной сети прямого распространения для распознавания рукописных цифр на базе FPGA // Доклады БГУИР. – 2025. – Т. 23. – №. 2. – P. 101-108.

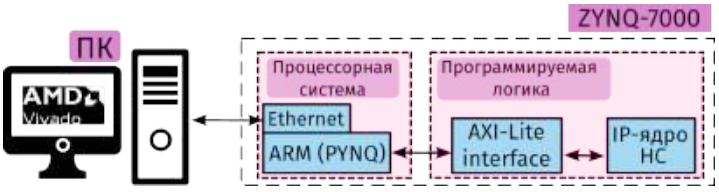


Рис. 4 – Реализация НС для распознавания рукописных цифр на базе платформы PYNQ Z2

СРАВНЕНИЕ АППАРАТНЫХ И ПРОГРАММНЫХ СРЕДСТВ РАСЧЁТА ОСНОВНЫХ ХАРАКТЕРИСТИК ФИЗИЧЕСКИ НЕКЛОНИРУЕМЫХ ФУНКЦИЙ

Трубач К. И., Иванюк А. А.
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: xenona11x@gmail.com, ivaniuk@bsuir.by

Проводится сравнение средств расчета основных характеристик (стабильности, единообразия, меж- и внутрикристальной уникальности) физически неклонируемых функций. Показано, что аппаратная реализация обеспечивает наивысшую скорость при оптимальных затратах ресурсов, Cortex-A9 предоставляет максимальную гибкость при минимальном быстродействии, а софт-процессорное ядро предлагает сбалансированное решение.

Введение

Физически неклонируемая функция (ФНФ) – специальная цифровая схема, которую легко спроектировать и реализовать, но практически невозможно воспроизвести [1]. Это свойство обуславливает широкое применение ФНФ в сферах защиты информации, физической криптографии, а также источниках случайности и системах идентификации.

Пригодность ФНФ к какому-либо из перечисленных направлений оценивается рядом характеристик, к которым относятся стабильность, надёжность, единообразие и меж- и внутрикристальная уникальность [2]. Регулярное измерение этих характеристик целесообразно и на стадии тестирования готовых ФНФ (в составе заказных ИС или реализованных на ПЛИС), поскольку они позволяют вовремя детектировать деградацию кристалла, а также обнаружить атаки как естественные (сюда относятся такие неблагоприятные условия, как колебания внешнего электромагнитного поля или некорретная работа источника питания кристалла), так и преднамеренные с целью компрометировать ФНФ.

При расчёте характеристик для научных целей важна гибкость, а для производственных – быстродействие и экономия ресурсов, поэтому необходимо осознанно подходить к выбору средств расчёта, которых можно выделить четыре: аппаратное процессорное ядро общего назначения Cortex-A9; софт-процессорное ядро общего назначения; софт-процессорное ядро со специализированным набором инструкций; аппаратная реализация на FPGA.

Все эксперименты в рамках данной работы проводились на плате быстрого прототипирования ZyboZ7 с FPGA Zynq7010 с использованием системы Jupyter Notebook и языка Python.

I. Обзор средств

Средство Cortex-A9 предполагает программный код, управляющий ФНФ (*PUF*) через регистровый файл (*RCU*). ФНФ, состоящая из конфигурируемых кольцевых осцилляторов (CRO) и счётчиков, формирует ответ (*Response*) на основе сравнения их показаний в заданном временном окне (*Window*). Описанная схема представлена на рис. 1.

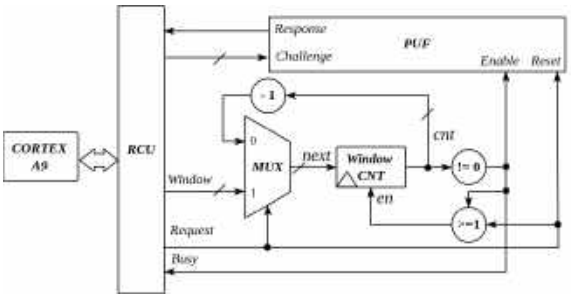


Рис. 1 – Схема ФНФ, контролируемая Cortex-A9

Подход минимизирует аппаратуру, переносящ расчёты на процессор общего назначения, и обеспечивает высокую гибкость и скорость разработки. Наиболее подходит для исследовательских работ, так как внедрение процессорного ядра в систему, изначально не предполагавшую его наличие, – это трудозатратный и дорогой процесс. Также ядро управляется языком программирования высокого уровня, что крайне негативно сказывается на общем затраченном времени.

Следующее средство – софт-процессорное микропрограммное ядро общего назначения с базовым набором инструкций. Оно способно работать с ФНФ и легко адаптируемо к любому их количеству. Предлагаемая архитектура, основанная на RISC-V и представленная на рис. 2, получила название *Verity RV* (далее – *VRV*).

Микропрограмма хранится в перепрограммируемом ЗУ (*PROM*). Загрузка инструкций в *PROM* и управление ядром (*MPCU*) осуществляются через интерфейс AXI4 Lite и регистровый файл (*RF*). ОЗУ разделено на две части: одна (*general purpose RAM*) для общего назначения, вторая (*RAM Mapped PUF CSRs*) для работы с

Работа выполнена в совместной учебной лаборатории БГУИР-YADRO
https://www.bsuir.by/ru/kaf-informatiki/yadro

ФНФ (*PUF*1, ..., *PUFn*). Управление ФНФ происходит на языке ассемблера, а Cortex-A9 лишь однократно загружает микрокод и запускает его. Главное преимущество *VRV* – микрокод, обеспечивающий реконфигурацию и гибкость.

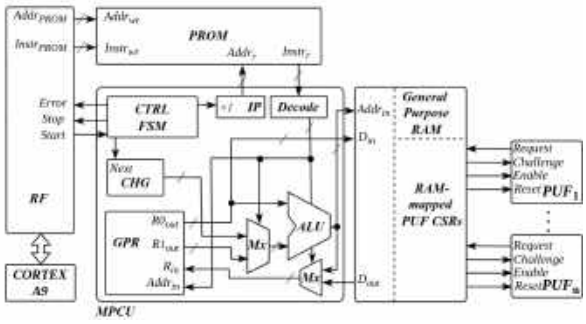


Рис. 2 – Схема софт-процессорного ядра *VRV*

VRV легко адаптируется под нестандартные расширения, что позволяет использовать софт-процессорное ядро со специализированным набором инструкций, созданных для максимального ускорения расчёта характеристик ФНФ и являющихся отдельным средством работы с ними – *VRV-S*. На рис. 3 показаны способ вычленения из алгоритма расчёта стабильности специализированной инструкции под названием *sta_ abs*, а также её аппаратная реализация. Ядро *VRV* требует затрат на разработку, но подходит как для исследований, так и для производства, сочетая гибкость и простоту реализации. Его универсальность выгодно выделяет *VRV* среди средств работы с ФНФ.

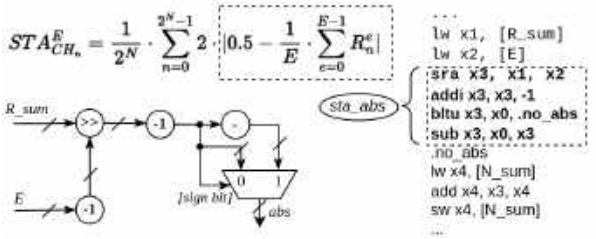


Рис. 3 – Пример получения специализированной инструкции

Последнее рассматриваемое средство — аппаратная реализация на FPGA. Она предполагает HDL-описание некоторой схемы расчёта (см. рис. 4) и её запуск на FPGA.

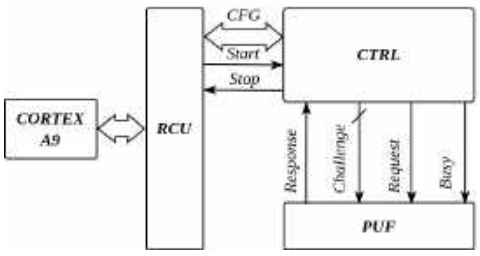


Рис. 4 – Схема аппаратной реализации

Контроль осуществляется через регистровый файл (*RCU*), включающего сигналы *Start* и *Stop* для сигнализации начала и конца расчёта, а также некоторой шины *CFG*, состав которой может варьироваться в зависимости от алгоритма реализуемой характеристики (например, размер окна измерения, количество итераций и пр.). Сигналы поступают на контрольное устройство *CTRL* с заданной логикой расчёта и прямым доступом к ФНФ. На рис. 4 ФНФ показана в единственном экземпляре для наглядности. Аппаратные схемы расчёта характеристик подходят для производства, не требуя больших затрат ресурсов и обеспечивая наибольшее быстродействие. Однако их переиспользование затруднено, и каждую схему (для каждой характеристики) необходимо разрабатывать и верифицировать отдельно.

II. Сравнение средств и выводы

Для количественного анализа выбрана характеристика стабильности, так как она включает вложенные циклы и требует значительного количества операций, что повышает точность расчёта. Результат показан в табл. 1, где t_{HW} и t_{SW} – время срабатывания аппаратуры и хоста соответственно, t_{SET} – время настройки аппаратуры. Затраты ресурсов для каждого средства не учитывают реализацию ФНФ, но учитывают логику доступа к ним (регистровый файл).

Таблица 1 – Эффективность рассмотренных средств при измерении стабильности ФНФ

Средство	t_{HW}, c	t_{SW}, c	t_{SET}, c	LUT6	FFs
Cortex-A9	429,7	$< 0{,}001$	0,3	324	263
VRV	1	0,2	2,3	1824	1410
VRV-S	1	0,07	2,1	1885	1419
Расчёт на FPGA	1	0,03	0,5	440	327

Таким образом, Cortex-A9 как средство требует минимальных ресурсов FPGA на реализацию ФНФ и доступа к ней, однако демонстрирует наихудшее время расчёта, что делает его пригодным для непроизводительных задач. Софт-процессор *VRV* является значительно более производительным решением, однако требует существенного увеличения ресурсов. Оптимизация *VRV* специализированными инструкциями улучшает скорость с незначительным ростом ресурсов. Наибольшее быстродействие при существенно меньших затратах ресурсов FPGA обеспечивает аппаратная реализация.

III. Список литературы

1. Secure System Design and Trustable Computing / ed.: Ch. H. Chang, M. Potkonjak. – Switzerland : Springer, 2016. – 549 p. DOI: https://doi.org/10.1007/978-3-319-14971-4.
2. Иванюк А. А. Исследование физически неклонируемой функции конфигурируемого кольцевого осциллятора // Информатика. – 2025. – Т. 22. – № 1. – С. 73-89.

ГРАФОВЫЕ МОДЕЛИ ЗАДАЧ, ВОЗНИКАЮЩИХ ПРИ ДЕКОМПИЛЯЦИИ КМОП СБИС

Черемисинова Л. Д.
Объединённый институт проблем информатики Национальной академии наук Беларуси
Минск, Республика Беларусь
E-mail: cld@newman.bas-net.by

Рассматриваются методы решения ключевых задач, возникающих при распознавании подсхем, реализующих КМОП-вентили, в плоской транзисторной схеме. Предлагаются графовые модели представления транзисторных схем при решении этих задач.

Введение

Современные КМОП СБИС содержат миллионы примитивных элементов уровня транзисторов, и их сложность быстро увеличивается. Инструментом, позволяющим существенно снизить трудоемкость тестирования схем такой сложности, служит средство их обратного проектирования (декомпиляции), в процессе которого представление схемы на транзисторном уровне заменяется функционально эквивалентным представлением на уровне логических элементов.

В настоящей работе рассматриваются некоторые ключевые задачи и методы, на которые разбивается процесс обратного проектирования КМОП схем на транзисторном уровне, заключающийся в экстракции из них структур на уровне логических элементов. Предлагаются графовые модели представления транзисторных схем при решении этих задач.

I. Декомпиляция транзисторных схем

Исходным объектом декомпиляции является плоское описание схемы из p-МОП и n-МОП транзисторов. Исходная плоская и результирующая иерархическая транзисторные схемы представляются в формате проектирования SPICE. Рассматривается наиболее распространенный стиль логики – логические комплементарные МОП-структуры и структуры на основе проходных транзисторов. Предполагается, что декомпилированная схема может содержать КМОП вентили, передаточные логические элементы и подсхемы на их основе, а также псевдоэлементы (нераспознанные как КМОП вентили).

В процессе декомпиляции выполняются следующие основные этапы [1]: 1) выделение подсхем, представляющих передаточные вентили и подсхемы (типа мультиплексоров) на их основе; 2) разбиение транзисторной схемы на подсхемы, состоящие из транзисторов, связанных по постоянному току; 3) распознавание «правильных» подсхем, представляющих КМОП вентили, а также реализуемых ими функций; 4) выделение групп функционально и топологически идентичных КМОП вентилей и псевдоэлементов – формирование библиотеки; 5) построение двухуровневой логической сети.

II. Задание транзисторных схем

Главной частью исходного плоского SPICE-описания является список транзисторов, в котором для каждого транзистора указываются имена цепей nd, ng, ns и nb, соединяющих его выводы (стока - drain, затвора – gate, истока – source, подложки – substrate) с остальными частями схемы: M<name> <nd> <ng> <ns> <nb> <model-name>, где name и model-name – название и тип (например, nmos или pmos) транзистора.

Связи транзисторной схемы удобно представлять помеченным неориентированным двудольным графом $G = (V_1, V_2, E)$, $V_1 \cap V_2 = \varnothing$. В нем вершины из первой доли V_1 соответствуют выводам экземпляров транзисторов и портов схемы, а вершинам из V_2 ставятся в соответствие цепи схемы. Например, структура приведенной на рис. 1 транзисторной схемы представляется графом G (рис. 2), для которого $V_1 = \{n_1d, n_1g, n_1s, \ldots, p_3d, p_3g, p_3s\}$, $V_2 = \{Vdd, Gnd, a, b, c, t_1, t_2, t_3, t_4\}$.

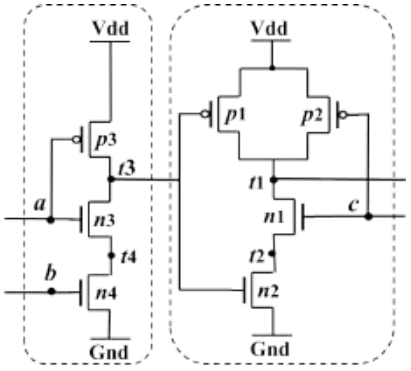


Рис. 1 – Пример транзисторной схемы

III. Распознавание групп транзисторов, связанных по току

Статическая КМОП-схема имеет четко определенную структуру, и соответствующая ей подсхема, представляет собой группу транзисторов, связанных по току. Такой группой является произвольная схема из МОП-транзисторов с тремя типами внешних соединений: 1) входы группы подаются только на затворы ее транзисторов; 2) выходы группы подаются только на затворы тран-

зисторов других групп; 3) имеются связи транзисторов с шинами питания (Vdd и Gnd).

В графовой интерпретации задача поиска групп транзисторов, связанных по току, сводится к нахождению компонент связности графа H (рис. 2), получаемого из графа G путем: 1) удаления вершин, соответствующих выводам затворов транзисторов и цепям питания; 2) соединения ребром вершин, соответствующих выводам стока и истока, для каждого транзистора. Группы транзисторов, соответствующие найденным двум компонентам связности графа H, выделены на рис. 1 пунктирными линиями.

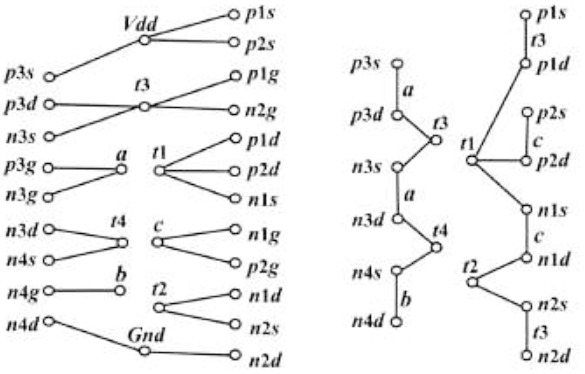


Рис. 2 – Графы G и H

IV. Распознавание подсхем, реализующих КМОП-вентили

Статический КМОП-вентиль содержит $n-$ и $p-$подсхемы, состоящие соответственно из n-МОП- и p-МОП-транзисторов. Эти подсхемы включены последовательно между цепями питания и разделены выходной цепью, а их проводимости комплементарны. Группа транзисторов, связанных по току, реализует КМОП-вентиль, если p- и n-блоки имеют одинаковое количество транзисторов и реализуют взаимно инверсные функции проводимости f_n и f_p. Группы, не удовлетворяющие этим двум условиям, относятся у псевдоэлементам. Например, левая из выделенных на рис. 1 подсхем является псевдоэлементом, так как не удовлетворяет первому условию.

В графовой постановке задача нахождения функций проводимости f_n и f_p сводится к поиску путей, связывающих выходную цепь подсхемы (на рис. 1 это t_1) с цепями питания (Gnd и Vdd). Задача решается на графе I, который получается графа G путем удаления вершин, соответствующих выводам затворов транзисторов и соединения вершин выводов стока и истока для каждого транзистора, ребром, которое помечается именем цепи, связанной с затвором данного транзистора (рис. 3). Просмотр цепей графа I позволяет найти функции, реализуемые $p-$ и $n-$блоками: $f_p = \overline{t_3} \vee \overline{c}$ и $f_n = c \wedge t_3$. Эти функции взаимно инверсны, что говорит о том, что правая подсхема на рис. 1 есть КМОП-вентиль 2И-НЕ, реализующий функцию $f = c \wedge t_3$.

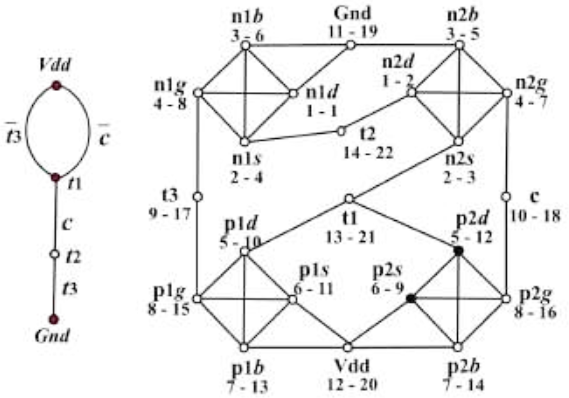


Рис. 3 – Графы I и T

V. Выделение классов топологически изоморфных элементов

На этом этапе множества подсхем, соответствующих распознанным КМОП-вентилям и псевдоэлементам, разбиваются на классы топологически изоморфных. Каждая из таких групп определяет элемент библиотеки подсхем второго уровня описания транзисторной схемы. Задача сводится к проверке изоморфизма помеченных графов для анализируемых подсхем путем приведения каждого из них к каноническому виду. Хеши канонических маркировок графов для изоморфных подсхем будут совпадать.

Для выполнения задачи канонизации граф, представляющий каждую из анализируемых подсхем, дополняется ребрами, связывающими попарно выводы каждого из его транзисторов. Каждая из подсхем задается помеченным графом $T = (W, F, h)$, где $h(w_i)$ – метка вершины из W. Метки 1–4 (5–8) присвоены вершинам графа, соответствующим выводам стока, истока, подложки и затвора n-МОП-транзисторов (p-МОП-транзисторов), затем идут метки входных и выходных полюсов схемы, Gnd, Vdd и, наконец, цепей схемы. Для найденной подсхемы И-НЕ на рис. 3 приведен граф T, в котором под именем каждой вершины указаны ее метки (начальная и полученная в результате канонизации).

VI. Заключение

Предложенные графовые модели и основанные на них методы решения ключевых задач распознавания подсхем КМОП-вентилей и построения иерархического описания на уровне логических элементов реализованы на языке C++ как часть программы декомпиляции плоской транзисторной схемы.

VII. Список литературы

1. Черемисинов, Д. И. Извлечение сети логических элементов из КМОП схемы транзисторного уровня / Д. И. Черемисинов, Л. Д. Черемисинова // Микроэлектроника. – 2019. – № 3 (48). – С. 224-234.

Секция «Математическое моделирование и компьютерная графика»

Section «Mathematical Modeling and Computer Graphics»

НЕЙРОСЕТЕВЫЕ ПОДХОДЫ К ВОССТАНОВЛЕНИЮ ТРЕХМЕРНОЙ ПОЗЫ ЧЕЛОВЕКА ПО ДВУМЕРНЫМ ДАННЫМ

Белаш В. А.
Кафедра информационных технологий в бизнесе,
факультет социально-экономических и компьютерных наук,
Национальный исследовательский университет Высшая школа экономики (Пермский филиал)
Пермь, Российская Фередрация
E-mail: vabelash@edu.hse.ru

В статье исследованы современные нейросетевые подходы для восстановления трехмерной позы человека на основе двумерных данных. Рассматриваются такие нейросетевые архитектуры, как SPIN, VideoPose3D и PoseFormer. Проведен их сравнительный анализ. Выявлено, что модели, использующие временной контекст (VideoPose3D, PoseFormer), демонстрируют преимущества при анализе динамических движений по сравнению со статическими методами (SPIN). Ключевые слова: нейросетевые модели, компьютерное зрение, распознавание движений, 2D-to-3D-lifting, спортивная аналитика.

Введение

Современная спортивная аналитика активно использует методы распознавания позы человека с помощью компьютерного зрения. Модели, ориентированные на работу с двумерными скелетными моделями, демонстрируют значительные успехи. Однако в ходе работы, направленной на распознавание и оценку движений фигуристов [1], было выявлено явное ограничение двумерного подхода. Оно заключается в сложности двумерного представления скелетных данных для анализа пространственных характеристик движения.

Во-первых, двухмерный скелет не позволяет надежно измерить ключевые показатели (например, поворот корпуса в вертикальной плоскости или пространственную траекторию движения). Во-вторых, сложная координация в таком виде спорта, как фигурное катание требует одновременного учета симметрии обеих сторон тела спортсмена (в двумерной плоскости для этого требуется производить съемку с разных сторон и затем объединять полученные кадры). Различные вращения, прыжки, расположение свободной ноги невозможно качественно проанализировать без трехмерного понимания биомеханики движений человека.

Актуальность преобразования двумерных поз в трехмерные также обусловлена заметным дисбалансе существующих наборах данных: на сегодняшний день гораздо больше размеченных 2D-датасетов по сравнению с 3D. Процесс разметки двумерных скелетов требует меньше трудозатрат, чем создание точных 3D-аннотаций с привлечением сложного оборудования и специализированного программного обеспечения. Это усиливает интерес к нейросетевым методам трансформации двумерных изображений в трехмерные.

Переход к трехмерному анализу предполагает использование более сложных вычислительных – «тяжеловесных» – моделей, что закономерно приводит к увеличению требований к вычислительным ресурсам. Однако для тех задач спортивной аналитики, где ключевое значение имеют точность оценки техники выполнения элементов и контроль биомеханически корректных положений, такое повышение вычислительной нагрузки оправданно.

I. Описание подходов преобразования двумерных скелетов

Основой для многих нейросетевых методов реконструкции трехмерной позы из двумерных данных является параметрическая модель SMPL – Skinned Multi-Person Linear Model [5]. Модель обучена на обширной наборе из трехмерных сканов человеческих тел и генерирует 3D меш человеческого тела с 6890 вершинами на основе двух наборов параметров: позовых (75 параметров, включающих вращения 24 суставов) и форменных (10 параметров, описывающих вариации телосложения).

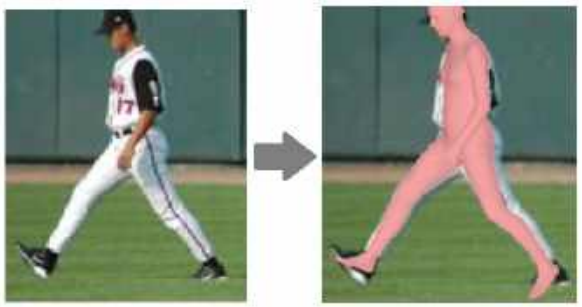


Рис. 1 – Трехмерное восстановление с помощью Synchronized Pose Image Network [2]

Так, нейросеть Synchronized Pose Image Network (SPIN) основана на интеграции нейросетевого регрессора с итеративной оптимизацией параметров модели SMPL. Особенность этого нейросетевого подхода заключается в реализации механизма самообучения, при котором 3D-модель человека устанавливается в двумерное пространство для соответствия исходным ключевым точ-

кам поданного на вход двумерного скелета (как изображение или в формате *json), чтобы последовательно уточнить параметры позы и формы тела в процессе обработки исходного одиночного изображения.

Пока модель SPIN фокусируется на статическом анализе, нейросеть VideoPose3D [3] от Facebook Research обрабатывает временные последовательности, принимая на вход последовательность 2D-поз. Архитектура строится на Temporal Convolutional Network, чье применение позволяет анализировать движение в рамках расширенного временного окна. Это дает не только повышенную точность реконструкции, но и плавность, устойчивость к шумам и естественность восстановленной 3D-анимации движения человека.

Современное развитие трансформерных архитектур нашло отражение в модели PoseFormer [4], реализующей пространственно-временной анализ последовательностей двумерных ключевых точек скелета. Архитектура использует механизм внимания на двух уровнях: пространственное внимание учитывает взаимосвязи между суставами в пределах отдельного кадра, а временное «внимание» отслеживает последовательность траектории движения отдельных суставов.

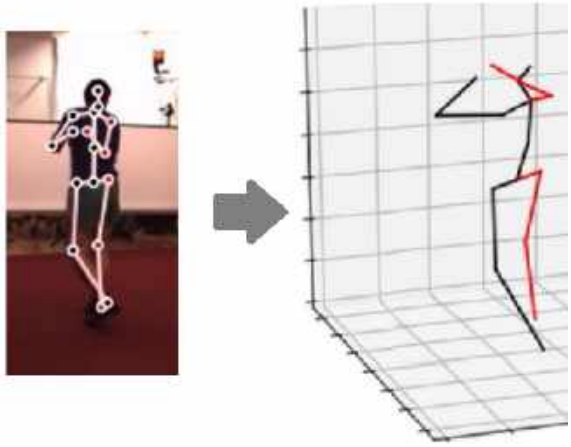


Рис. 2 – Пример преобразования с помощью PoseFormer [4]

II. Сравнение нейросетевых подходов

Описанные ранне подходы демонстрируют различные стратегии обработки данных: от работы с одиночными изображениями до анализа пространственно-временных видеопоследовательностей. Таблица 1 представляет собой результат сравнительного анализа.

Выбор критериев обусловлен спецификой задач спортивной аналитики (например, для анализа динамических движений в спорте важен учет временных зависимостей).

Таблица 1 – Сравнительный анализ

Критерий	SPIN	VideoPose3D	PoseFormer
Inputs	1 кадр	послед-ть 2D-поз	послед-ть 2D-поз
Outputs	3D-поза + меш	3D-скелет	3D-скелет
Временной контекст	–	свертки	внимание
Учет простр. контекста	огранич.	огранич.	полный
Ключевая идея	SMPL	TCN	SpTemp

Заключение

Перспективным направлением развития подходов к восстановлению трехмерной позы человека по двумерным данным представляется создание гибридных архитектур, сочетающих преимущества рассмотренных подходов: точность статических методов, временную согласованность видеопоследовательностей и способность учитывать пространственно-временные зависимости. Кроме того, важно адаптировать методы 3D-реконструкции для специфических требований различных спортивных дисциплин. Например, при анализе техники движений фигуристов разумно учитывать не только биомеханику тела спортсмена, но и положение коньков относительно ледовой поверхности, что требует расширения стандартных моделей скелетов.

1. Ученые НИУ ВШЭ в Перми учат искусственный интеллект разбираться в фигурном катании [Электронный ресурс] // Новости науки в НИУ ВШЭ. – 2025. – Режим доступа: `https://www.hse.ru/news/science/1034088150.html` (дата обращения: 25.10.2025).
2. Kolotouros, N. Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop / N. Kolotouros, G. Pavlakos, K. Daniilidis // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 4501-4510.
3. Pavllo, D. 3D human pose estimation in video with temporal convolutions and semi-supervised training / D. Pavllo, C. Feichtenhofer, D. Grangier, M. Auli // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 7753-7762.
4. Zheng, C. 3D Human Pose Estimation with Spatial and Temporal Transformers / C. Zheng, S. Zhu, M. Mendieta, T. Yang, C. Chen, Z. Ding // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – P. 11656-11665.
5. Loper, M. SMPL: A Skinned Multi-Person Linear Model / M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M. J. Black // ACM Transactions on Graphics (TOG). – 2015. – Vol. 34, № 6. – Article 248. – P. 1-16.

МАРКИРОВКА РАСТРОВЫХ ИЗОБРАЖЕНИЙ ДЛЯ ЗАЩИТЫ ОТ ФАЛЬСИФИКАЦИИ

Белобокова Ю. А., Колистратов М. В.
Кафедра общегуманитарных и етественнонаучных дисциплин,
Московский региональный социально-экономический институт,
Кафедра Инфокоммуникационных технологий,
Национальный исследовательский технологический университет МИСИС
Видное, Москва, Российская Федерация
E-mail: yulya.belobokova@mail.ru, kolistratov.mv@misis.ru

Предлагается методика цифровой маркировки изображений с применением технологии водных знаков (ЦВЗ) для выявления фальсификаций. Используются две метки: bitmap-логотипы для подтверждения авторства и электронные подписи для фиксации изменений. Метки внедряются в красный и синий каналы RGB-изображений, обеспечивая защиту от манипуляций. Подбор оптимальных параметров гарантирует надежность и незаметность меток. Эффективность методики доказана экспериментально и реализована в программном продукте на языке C#.

Введение

Благодаря развитию инструментов обработки изображений, публикуемые в электронных СМИ, а также социальных сетях цифровые фотографии становятся уязвимыми для возможных подделок. При этом в некоторых случаях изменение даже небольшого фрагмента может полностью изменить смысл исходного фото. Поэтому разработка методов противодействия фальсификации цифровых изображений является достаточно актуальной задачей.

I. Методы защиты цифровых изображений

Технология цифровых водяных знаков (ЦВЗ) востребована в качестве защиты цифрового контента еще с середины девяностых годов прошлого века. ЦВЗ представляет собой скрытую метку, внедряемую в изображение с помощью некоторого алгоритма в соответствии с определенным ключом. Встраиваемые метки оценивают по трем критериям, причем первые два противоречат друг другу. Это незаметность (отсутствие сильных искажений в изображении), устойчивость (робастность, стойкость к возможным искажающим воздействиям) и емкость (количество информации, содержащейся в ЦВЗ).

Цифровые фотографии, предназначенные для использования в сети интернет, чаще всего сохраняют в форматах JPEG и PNG, поскольку эти форматы поддерживаются практически всеми графическими редакторами. Изображения, сохраненные в формате JPEG, из-за режима сжатия с потерями проигрывают аналогичным PNG-изображениям в качестве, но при этом обладают меньшими размерами. При высокой степени сжатия у JPEG-изображения необратимо теряется часть графической информации, что может привести к потере маркировки (у PNG-изображений такая проблема не возникает). Но такое сжатие при небольших размерах исходной JPEG-фотографии может существенно «испортить картинку», поэтому его использование станет менее вероятным.

При фальсификации исходные цифровые изображения с большей вероятностью могут подвергаться кадрированию, добавлению или удалению каких-либо элементов. Менее вероятны смена формата или цветовой модели, масштабирование, добавление фильтров цвета, применение таких инструментов, как резкость, автоуровни, автоконтраст, зашумление.

Логично, что ЦВЗ, указывающие на авторство изображения, должны быть максимально устойчивыми к этим воздействиям. Если же поставлена задача определения фальсифицированных фрагментов, внедряемые метки должны указывать на факт фальсификации. Выполнение обоих задач требует внесения ЦВЗ с разной степенью робастности, поэтому была разработана специальная методика [1,2] на основе метода Коха и Жао [3], позволяющая различать оригинальные и сфальсифицированные фрагменты изображения благодаря внедрению двух типов невидимых меток. Первый тип – это собственно ЦВЗ (симметричные bitmap-логотипы), показывающие авторство изображения и подлинность его составных частей, второй – электронные сигнатуры, (хэш-коды, строки бит, вычисляемые на основе размеров изображения или его фрагментов), которые могут свидетельствовать о факте внесения изменений, например, кадрировании, ретуши или добавлении инородных фрагментов [1]. Значения электронных сигнатур зависят от свойств маркируемой фотографии, а ЦВЗ постоянен для всех её фрагментов. профиля на сайте конференции.

II. Методика маркировки распознавания

Разработанная методика включает в себя алгоритмы множественной защитной маркировки и проверки растровых изображений на аутентичность и целостность и применяется к изображениям, сохраненным в формате RGB, поскольку из-за

особенностей человеческого зрения встраивание маркировочных меток идёт только в красный и синий каналы.

Обработанное в форме битовой карты изображение представляется в виде матрицы и разбивается на блоки 64x64 пикселя (некратные блоки остаются снизу и/или справа изображения), после чего каждый из блоков разделяется на 3 матрицы (по числу цветовых каналов). Далее в полные b-матрицы (синий канал) встраивается ЭП и половина бит монохромного логотипа, а в полные r-матрицы (красный) – вторая половина бит логотипа. Неполные блоки размером менее 18x18 пикселей не маркируются, в остальные неполные блоки встраиваются только биты ЦВЗ. Способ встраивания подробно описан в [1]. Очень важен правильный выбор значения параметра, называемого коэффициентом силы встраивания: чем он выше, тем больше не только робастность маркировки, но и ее заметность. Серия экспериментов показала, что для маркировки низко- и среднечастотных областей фотографий рационально использовать коэффициент силы встраивания со значением 5..7, а для Высокочастотных областей – значения 7..20.

При проверке подлинности цифрового изображения из него извлекаются внедренные ЦВЗ и электронные сигнатуры. При этом изображение также представляется в виде матрицы, разбиваемой на блоки 64x64, и маркировка извлекается из красного и синего каналов. Если в верхнем левом блоке обнаружены биты ЦВЗ, предполагается, что найден базовый блок и проверяются остальные блоки. В противном случае генерируется сдвиг, изображение переразбивается, и в новых блоках также ищется ЦВЗ. По характеру расположения бит ЦВЗ в блоках синего канала ищутся значения электронных сигнатур. При проверке блоков возможны три исхода: блок аутентичен (либо применяемые к изображению воздействия не исказили его содержание); блок содержит ЦВЗ, но значение сигнатуры указывает на наличие искажающих воздействий; блок не содержит маркировки, что говорит о сильных разрушающих воздействиях или его фальсификации.

Для повышения защиты изображений целесообразно использовать базу данных, в которой хранятся не только сами промаркированные изображения, но и такая информация, как дата маркировки, геометрические размеры, образец встраиваемого монохромного логотипа и т.д. Разработанная методика была реализована в виде программного комплекса на языке программирования C#. Серия экспериментов подтвердила эффективность разработанной методики и ее практическую применимость

III. Результаты

Для проверки устойчивости метода была сформирована база файлов с растровыми изображениями с различающимися типами объектов, изображенных на них (природа, здания и сооружения, люди, предметы обихода). Программным способом в изображения по разработанной авторами методике в синий и красный каналы были встроены маркировочные знаки двух видов (рис. 1). Каждое изображение неоднократно маркировалось с использованием различных значений коэффициента силы встраивания, после чего в результате сравнения результатов с помощью визуального анализа группой экспертов и определения соотношения сигнал-шум выбиралось рациональное значение коэффициента p, позволяющее делать маркировку не только стойкой, но и незаметной.

Показан результат работы разработанной программы по выявлению внесённых изменений в промаркированные изображения (рис. 2). Выявлены измененные фрагменты изображения и предположительные действия с ним. Белая сетка – сигнатура неизменная, черная – слетели значения. Красные блоки указывают на то, что ЦВЗ изменен, зелёные – ЦВЗ распознан.

Рис. 1 – Промаркированное исходное изображение

Рис. 2 – Изменённое изображение, прошедшее процедуру распознавания встроенных меток в разработанном программном модуле

IV. Список литературы

1. Белобокова, Ю. А. Метод многократной маркировки цифровых фотографий для защиты от фальсификации / Ю. А. Белобокова, Е. В. Булатников // Известия высших учебных заведений. Проблемы полиграфии и издательского дела. – 2014. – №2 – С. 33–41
2. Белобокова, Ю. А. Защита информационного содержания цифровых фотографий методом многократной маркировки цифровыми водяными знаками / Ю. А. Белобокова, Э. С. Клышинский // Системный администратор. – 2014. – №4 – С. 70–73
3. Koch, E. TTowards Robust and Hidden Image Copyright Labeling / Koch E., Zhao J. // IEEE Workshop on Nonlinear Signal and Image Processing. Halkidiki, Greece – 1995 – P. 452–455.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЛИЯНИЯ ПАРАМЕТРОВ МАТЕРИАЛОВ МЕТАЛЛИЗАЦИИ НА ЭКСПЛУАТАЦИОННЫЕ ХАРАКТЕРИСТИКИ ЭЛЕМЕНТОВ ЦИФРОВЫХ ИНТЕГРАЛЬНЫХ СХЕМ

Винник М. А., Кривец К. А., Ловшенко И. Ю., Новиков П. Э., Корсак К. В., Степанов А. А.
Белорусский государственный университет информатики и радиоэлектороники
Национальный детский технопарк
Минск, Республика Беларусь
E-mail: lovshenko@bsuir.by

В работе исследуется влияние выбора материалов металлизации на надёжность ячеек статической памяти. В качестве ключевого параметра используется статический запас по шуму (SNM). Проведён сравнительный анализ SNM для семи приборных структур с различными материалами металлизации: медь, алюминий, вольфрам, никель, кобальт, золото и серебро. Установлено, что медь обеспечивает наивысшее значение SNM (0,289 В), что подтверждает её статус оптимального материала для современных интегральных схем. Результаты работы подчёркивают важность учёта параметров материала при проектировании цифровых микросхем с повышенной надежностью.

Введение

На ранних этапах развития микроэлектроники алюминий стал стандартным материалом для межсоединений в интегральных схемах благодаря относительной простоте технологий осаждения и совместимости с кремнием. Однако по мере перехода к нанометровым технологическим нормам, сопровождающемуся уменьшением сечения проводников и увеличением плотности тока, фундаментальными ограничениями алюминия являются высокое удельное сопротивление и низкая стойкость к электромиграции, что со временем стало критическим. Эти недостатки привели к росту задержек сигнала, увеличению тепловых потерь и, в целом, к снижению надёжности микросхем.

Естественным решением этой проблемы стала замена алюминия на медь, которая обладает гораздо меньшим удельным сопротивлением и значительно лучшей устойчивостью к электромиграции. Это началось с перехода на процессы по нормам 90 нм, что потребовало кардинальной перестройки технологий, например, внедрения барьерных слоёв [1-3]. Тем не менее, поиск оптимальных материалов для металлизации не остановился на меди. В условиях постоянного стремления к повышению энергоэффективности и надёжности цифровых схем, особенно в элементах памяти, актуальным становится вопрос о влиянии не только основного проводникового материала, но и альтернативных вариантов, таких как золото, серебро, кобальт или никель, на ключевые эксплуатационные характеристики.

I. Оценка надежности ячеек памяти

Надёжность ячеек памяти при чтении данных становится критически важной в современных нанометровых технологиях. Повышенная вариативность характеристик элементов памяти делает их более чувствительными к шумам и увеличивает разброс параметров, что в итоге снижает общую надёжность. В последнее время перспективным направлением для энергоэффективных устройств стало использование подпорогового режима работы памяти. Однако ключевой проблемой такого подхода является обеспечение стабильного чтения и записи. Снижение напряжения питания, часто применяемое для уменьшения энергопотребления, ещё больше усугубляет проблему надёжности. Таким образом, стабильность работы энергоэффективных ячеек памяти при чтении становится одним из важнейших требований [4]. Статический запас по шуму (англ. Static Noise Margin – SNM) – это мера устойчивости ячейки памяти к сохранению своих данных при наличии шума. SNM определяется как минимальное напряжение, присутствующее на битовых линиях без информационного сигнала, необходимое для изменения состояния ячейки. Существует два метода измерения SNM: графический подход, в котором SNM определяется как длина максимально возможного квадрата, вписанного в передаточную характеристику, и второй подход, предполагающий использование источников шума (в виде источников напряжения) на битовых линиях [5]. Для каждой конструкции построены передаточные характеристики графическим методом, и методом полного перебора рассчитан SNM. Качественный вид передаточной характеристики при операции чтения для каждой структуры является одинаковым и, в качестве примера, на рисунке 1 приведена передаточная характеристика одной из рассматриваемых структур. Выходное и входное напряжение обозначается как напряжение в точках int1 и int2, соответствующих разным битовым линиям. SNM измеряется в вольтах и характеризует способность ячейки сохранять данные при воздействии шума. Чем выше значение SNM, тем устойчивее ячейка.

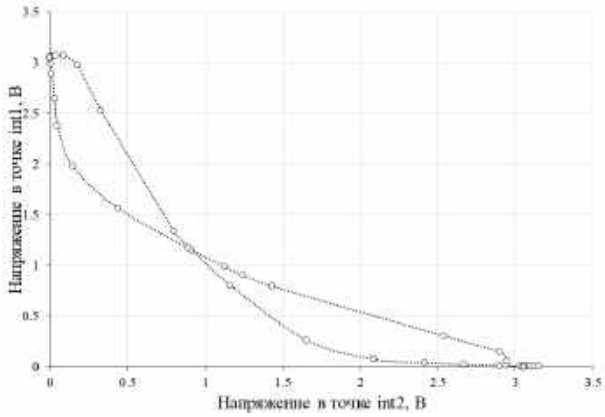


Рис. 1 – Типовая передаточная характеристика (структура № 2)

II. Определение оптимальных материалов металлизации

Исследовано влияние материалов металлизации на надёжность ячеек памяти при чтении данных посредством математического моделирования в среде Silvaco. Ячейка памяти представляет собой стандартную 6T реализацию [6], при моделировании исследовалось ее 3D-представление (рисунок 2). В исследуемой приборной структуре можно выделить 4 слоя металлизации. В таблице 1 представлены сочетания используемых материалов слоев металлизации, а также значения рассчитанных SNM для всех рассматриваемых приборных структур.

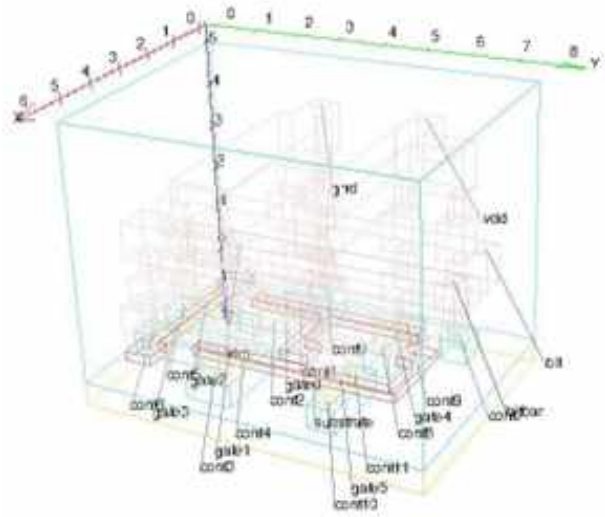


Рис. 2 – 3D-представление 6T ячейки памяти

Проведённый анализ данных показал, что выбор материала металлизации оказывает значительное влияние на надёжность ячейки памяти. Наибольшее значение SNM составляет 0,289 В и было получено при использовании меди во всех слоях металлизации. Это в первую очередь связано с её низким удельным сопротивлением, что минимизирует паразитные падения напряжения и тепловые шумы.

Таблица 1 – Варьирование материалов каждого слоя металлизации для приборных структур

Номер структуры	Материал металлизации	SNM, В
1	Медь	0,289
2	Алюминий	0,267
3	Вольфрам	0,231
4	Никель	0,270
5	Кобальт	0,266
6	Золото	0,275
7	Серебро	0,277

Среди остальных материалов наилучшие результаты показали серебро (0,277 В) и золото (0,275 В), что также объясняется их высокой проводимостью. Однако их применение в массовом производстве ограничено из-за высокой стоимости и технологических сложностей при реализации в традиционных кремниевых технологиях. Алюминий и кобальт продемонстрировали схожие значения SNM, 0,267 В и 0,266 В соответственно, что делает их приемлемыми альтернативами при определённых технологических ограничениях. Никель (0,270 В) показал немного лучший результат по сравнению с алюминием. Наименьшее значение SNM равно 0,231 В и было получено при использовании вольфрама, что обусловлено его относительно высоким сопротивлением и меньшей подвижностью носителей заряда.

Таким образом, для обеспечения максимальной надёжности ячеек памяти при чтении предпочтительным выбором остаётся медь, несмотря на определённые технологические ограничения, связанные с производством микросхем с металлизацией на ее основе.

1. Amrutur, B. S. Speed and Power Scaling of SRAM's / B. S. Amrutur, M. A. Horowitz // IEEE Journal of Solid-State Circuits. – 2000. – V. 35, № 2. – P. 175–185.
2. Andricacos, P. C. Copper on-chip interconnections: A breakthrough in electrodeposition to make better chips / P. C. Andrianos // The Electrochemical Society Interface. – 1999. – V. 8, № 1. – P. 32.
3. Copper interconnection integration and reliability / C-K. Hu, B. Luther, F. B. Kaufman [et al.] // Thin Solid Films. – 1995. – V. 262, № 1-2. – P. 84-92.
4. Gul, W. SRAM cell design challenges in modern deep submicron technologies: An overview / W. Gul, M. Shams, D. Al-Khalili // Micromachines. – 2022. – V. 13, № 8. – P. 1332.
5. Gadhe, A. Read stability and Write ability analysis of different SRAM cell structures / A. Gadhe, U. Shirode // International Journal of Engineering Research and Applications (IJERA). – 2013. – V. 3, № 1. – P. 1073–1078.
6. Osada, K. «Low power and reliable SRAM memory cell and array design» / K. Osada // Springer Series in Advanced Microelectronics. – 2011. – V. 31. – P. 5-10.

ТЕПЛОВОЕ МОДЕЛИРОВАНИЕ МИКРОСХЕМЫ УСИЛИТЕЛЯ МОЩНОСТИ X ДИАПАЗОНА

Журавлёв В. И., Наумович Н. М., Стежко И. К., Колбун В. С.
Центр 1.6 НИЧ
Белорусский государственный университет информатики и радиоэлектроники
г.Минск, Республика Беларусь
E-mail: vadzh@bsuir.by

Предложена тепловая модель бескорпусной монолитной микросхемы усилителя в составе приёмопередающего модуля. Результаты моделирования показывают, что при максимально допустимой постоянной входной мощности обеспечивается допустимый нагрев и равномерное распределение теплового поля в области размещения кристалла усилителя.

Введение

Бескорпусные твердотельные элементы, выполненные в виде микроволновых интегральных монолитных интегральных схем на диэлектрических подложках с высокой диэлектрической проницаемостью, всё чаще используются в качестве мощных усилителей X диапазона [1]. Такие компоненты находят применение в усилителях мощности приёмопередающих модулей и радиочастотных переключателях [2]. Несмотря на технологический прогресс, позволяющий увеличивать мощность микроволнового излучения, системы отвода тепла часто не справляются с необходимыми режимами охлаждения в условиях эксплуатации.

Наиболее сложной является задача поддержания стабильных температур при высокой плотности упаковки усилителей и переключателей мощности в условиях низкого атмосферного давления или его отсутствия для бортовых систем [3]. Управление тепловым режимом цепей радиочастотной мощности играет ключевую роль в обеспечении стабильной и надежной работы оборудования. Основная задача – эффективное удаление избыточного тепла из компонентов и участков цепи, чтобы предотвратить их перегрев, снизить риск ухудшения характеристик или выхода из строя. Даже несмотря на универсальный характер контроля тепловых режимов для всех элементов цепи, оптимальный подход должен начинаться именно с компоновки и проектирования печатных модулей. Правильное расположение компонентов, использование эффективных участков металлизации и организация хороших тепловых путей позволяют значительно повысить тепловую эффективность и обеспечить надёжную работу радиочастотных систем. Это требует разработки специальных решений для обеспечения эффективного теплоотвода и предотвращения перегрева активных элементов, что критически важно для поддержания их работоспособности и долговечности. Моделирование этих тепловых процессов играет важную роль в определении оптимальных параметров охлаждения и позволяет настроить как электрические, так и конструкционные свойства мкроволновых усилителей мощности.

I. Построение тепловой модели

В качестве примера рассматривается микросхема усилителя X диапазона CHA8100 (United Monolithic Semiconductors), выполненная в бескорпусной конструкции на основе GaAs и предназначенная для размещения на теплоотводящем основании. Рассмотрим анализируется конструкция с пассивным теплоотводом, где микросхема размещена на теплоотводящей пластине, что способствует повышению эффективности теплоотдачи и снижению температуры активных элементов, обеспечивая более стабильную работу устройства и увеличивая его надёжность.

В разработанной модели анализа теплового поля для описания распределения температуры применяются граничные условия второго рода. Для упрощения решения были введены следующие допущения. Распространение тепла в многослойной системе считается определяющимся распределением внутри того слоя, который занимает наибольшую площадь. При этом, несмотря на структуру микросхемы из нескольких слоёв, максимальные тепловые перепады в основном зависят от параметров материала полупроводниковой подложки, которая существенно крупнее сформированных слоёв. Хотя теплоотвод зачастую имеет размеры, сопоставимые с кристаллом, распределение тепловых потоков в нём аналогично распределению в кристалле, а возникающие тепловые градиенты в целом значительно меньше. При общем решении уравнения теплопроводности параметры полупроводника моделируются как постоянные величины, что значительно упрощает аналитическую и числовую обработку.

Топология усилителя мощности довольно неоднородна, поэтому применение численных методов является важным для оценки теплового нагрева как активных элементов, так и прилегающих областей структуры. В моделируемой сеточной модели каждая область представляется как самостоятельный источник тепла. Используя CFD-методы, можно получить распределение тепловых полей в неоднородной структуре, что

обеспечивает более точное представление о температурных градиентах и эффективных путях отвода тепла внутри устройства.

II. Анализ результатов

Результаты моделирования показывают, что при максимально допустимой постоянной входной мощности примерно 11 Вт обеспечивается допустимое и достаточно равномерное распределение теплового поля в области размещения кристалла усилителя (рис.1). В местах крепления и пайки теплопередача замедляется из-за увеличенного теплового сопротивления, однако это не оказывает значительного влияния на общую температуру. Анализ распределения температуры по кристаллу и теплоотводящей пластине усилителя выявил отсутствие существенных температурных градиентов, что важно для надежной работы СВЧ тракта и предотвращения термомеханических напряжений (рис.2). Максимальные температурные перепады вдоль кристалла составляют менее 2 °C (рис.3). Видно, что механизм теплопроводности эффективно и равномерно работает вдоль всего теплового пути. Это является допустимым в условиях эксплуатации при заданной мощности, позволяя своевременно прогнозировать и избегать локальных перегревов и структурных проблем на стадии проектирования.

Моделирование в импульсном режиме показало возможность снижения рассеиваемой мощности и минимизации рисков деградации из-за неравномерного нагрева. При длительности импульса 50 мкс и частоте повторения 3,5 кГц средняя рассеиваемая мощность в микросхеме составила около 0,45 Вт за один цикл. Максимальная температура нагретой зоны при этом не превышала 99 °C, что соответствует требованиям по эксплуатационной надёжности и сохранению электрических характеристик усилителя.

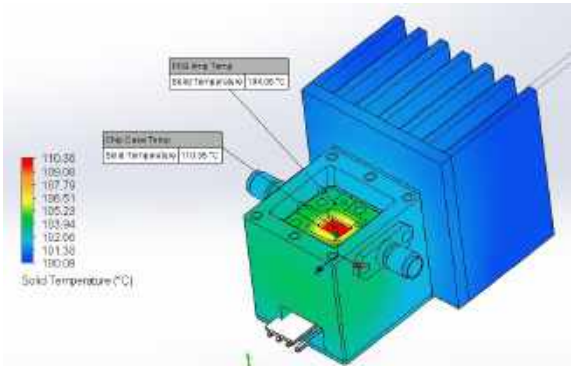


Рис. 1 – Результаты моделирование нагрева усилителя при максимально подводимой постоянной мощности

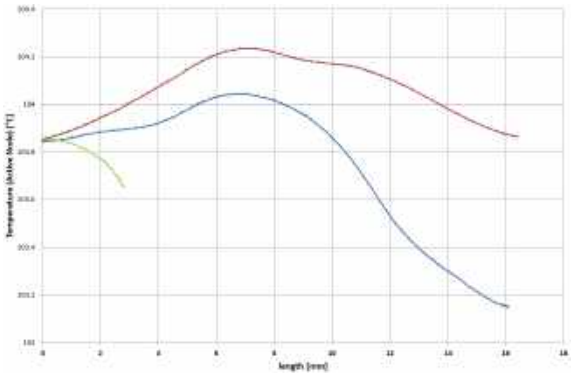


Рис. 2 – Градиенты температуры по осям пластины

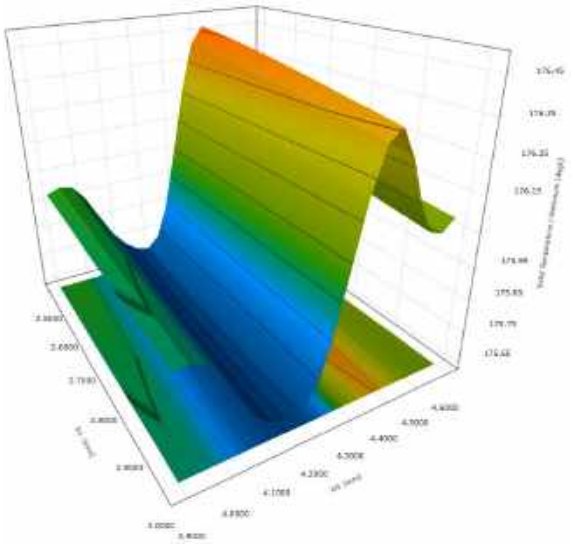


Рис. 3 – Расчётное распределение температуры кристалла

Заключение

Результаты позволяют использовать моделирование для анализа теплового распределения в усилителях мощности диапазона X и использовать меры для снижения потерь в СВЧ тракте без необходимости кардинальной смены конструкции, что способствует повышению эффективности системы и долговечности её компонентов.

Список литературы

1. E. Aras, *et al.* HEMTs With Side-Source-Via for Optimizing the Performance and Size of High-Power Amplifiers // IEEE Microwave and Wireless Technology Letters. – vol. 35, no. 7. – 2025. – P. 1061-1064.
2. S. T. Sriram, *et al.* A 42 dB Gain, 16 W X-Band MMIC Power Amplifier with 46% PAE for Satellite Applications / 2024 IEEE Microwaves, Antennas, and Propagation Conference, Hyderabad. – 2024. – P. 1-4.
3. I. Torralbo, *et al.* Thermal Analysis of the Solar Orbiter PHI Electronics Unit // IEEE Transactions on Aerospace and Electronic Systems. – vol. 56, no. 1. – 2020. – P. 186-195.

МОДЕЛИРОВАНИЕ СИЛОВОГО ПРЕОБРАЗОВАТЕЛЯ С УЧЁТОМ ТЕМПЕРАТУРНЫХ ЗАВИСИМОСТЕЙ

Журавлёв В. И., Свито И. Л., Петровский И. И., Шилин Л. Ю.
Кафедра теоретических основ электротехники
Белорусский государственный университет информатики и радиоэлектроники
г.Минск, Республика Беларусь
E-mail: vadzh@bsuir.by

Предложен подход адаптивного моделирования, при котором учитывается влияние температурной зависимости теплопроводности материалов на нагрев компонентов силового преобразователя

Введение

Тепловые модели критически важны для надёжности, эффективности и долговечности силовой электроники. Они позволяют оценивать тепловые режимы компонентов и предотвращать перегрев, отказы и осуществлять оптимизацию конструкции. Без отвода тепла возможно превышение допустимой температуры кристалла, что ведёт к деградации или выходу из строя. Тепловые модели позволяют заранее оценить тепловые потоки и грамотно разместить компоненты, радиаторы и, повышая эффективность охлаждения и общую надёжность. Однако адекватное тепловое моделирование силовых преобразователей сталкивается с рядом трудностей: от сложности геометрии и многослойности материалов до нелинейных эффектов и ограничений программных средств. Нелинейность в тепловой модели могут вносить многочисленные факторы. К ним относятся механизмы распространения тепла, процессы конвективного охлаждения, а также зависимость свойств материалов от температуры и другие аспекты. Влияние температуры на тепловые параметры часто не учитывается в тепловых моделях, что может вызывать существенные погрешности в расчетах от 4% до 12% [1-2]. Однако тепловое поведение силового преобразователя напрямую зависит от окружающей температуры, что необходимо учитывать при проектировании и эксплуатации.

I. Корректировка тепловой модели

В типовых силовых преобразователях активно используются такие классические материалы как кремний, алюминий, медь. Анализ тепловых свойств показывает, что теплопроводность кремния снижается с повышением температуры, в то время как удельная теплоёмкость увеличивается. При температуре около 250 °C теплопроводность кремния падает почти вдвое по сравнению с её значением при 25 °C. В то же время изменение плотности различных материалов, входящих в состав силовых полупроводниковых устройств, в диапазоне рабочих температур (от 25°C до примерно 200°C) является незначительным, поэтому влияние температуры на теплоёмкость можно считать пренебрежительным и не учитывать в моделях.

Для учёта зависимости свойств от температуры, такие параметры как теплопроводность и теплоёмкость материалов в тепловой модели представлены в виде функций от температуры, путём корректировки соответствующих характеристик. На рис.1 приведены зависимости теплопроводности материалов, используемых в модуле преобразователя, в диапазоне температур. Эти данные позволяют более точно моделировать тепловое поведение модуля при различных условиях работы.

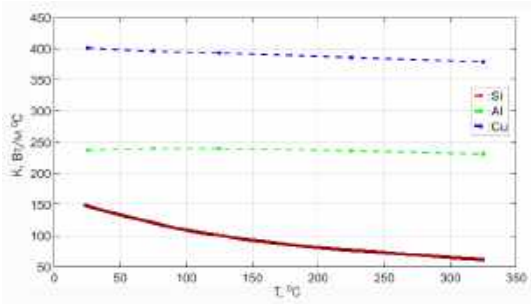


Рис. 1 – Влияние температуры на теплопроводность материалов

Для оценки погрешностей, вызванных игнорированием температурной зависимости тепловых свойств, выполняется решение полной нелинейной тепловой модели, в которой учитываются температурно-зависимые параметры материалов, такие как теплопроводность. Процесс переходного моделирования проводится итерационным методом при изменении температуры охлаждающей среды, которая обычно находится в диапазоне от 300 K до 370 K. На начальном этапе решается линейная модель, использующая свойства материалов при постоянной температуре, при различных значениях температуры окружающей среды. Аналогично, выполняется решение нелинейной модели, в которой свойства материалов зависят от температуры, при тех же условиях. После выполнения расчетов результаты сравниваются. Такой подход позволяет проанализировать влияние учета температурной зависимости свойств на точность моделирования тепловых процессов в силовых модулях.

II. Анализ и сравнение результатов

Для апробации предложенной корректировки модели использовались результаты экспериментального исследования понижающего преобразователя на широко используемый микросхеме LM2596 при наименьшем выходным напряжением (12В в 3,3В). Предварительно были получены результаты SPICE моделирования оценки рассеиваемой мощности (рис. 2 и 3). Сравнительные результаты приведены в табл.1 на основе разработанной ранее термоэлектрической модели, рассматриваемой адаптивной тепловой модели и измеренных экспериментально. В целом, профили температуры хорошо согласуются, хотя наблюдается небольшое отклонение около 2,2 °C. Поскольку теплопроводность кремния уменьшается с повышением температуры окружающей среды, тепловое сопротивление системы увеличивается. В результате при сравнении линейной модели, использующей свойства при температуре 300 K с нелинейной моделью при температуре 370 K, наблюдается погрешность примерно в 8%. Эта ошибка обусловлена температурной зависимостью теплопроводности кремния. В основном погрешность в тепловом импедансе возникает из-за изменения теплопроводности кремния и меньшей активной площади кремниевого слоя по сравнению с алюминиевым или медным слоем. Вклад теплового сопротивления слоя кремния в общее тепловое сопротивление системы невелик, и его влияние на общие тепловые характеристики можно считать незначительным. Таким образом, использование свойств при средней температуре позволяет снизить погрешность до 3% (при текущей температуре).

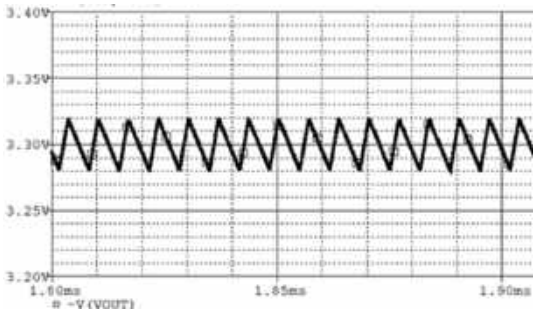


Рис. 2 – Форма напряжения на выходе

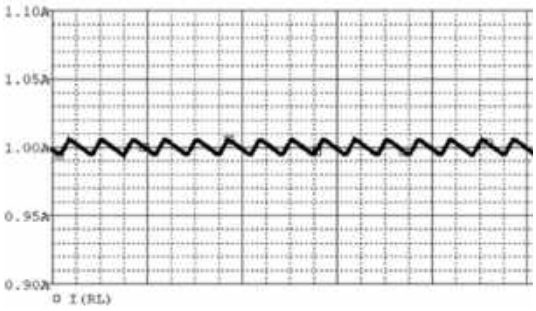


Рис. 3 – Форма тока нагрузки

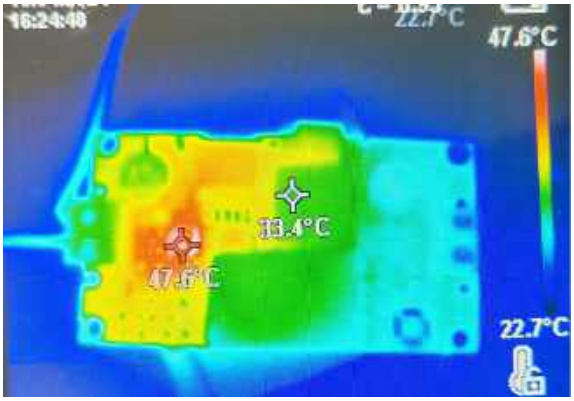


Рис. 4 – Вид измеренного распределения температуры

Таблица 1 – Значения температуры (°C)

Термоэлектрическая модель	Тепловая модель	Измерение
43,8	46.2	47,6

III. Заключение

В настоящее время не существует наилучшего способа полностью учесть влияние температуры на тепловые свойства материалов силовых приборов. Такие факторы, как наличие и расположение миниатюрных пассивных компонентов (конденсаторов, резисторов), наличие их потерь также могут влиять на распределение температуры. Эти потери, если их не учитывать в модели, могут стать источниками дополнительного тепла, что ведет к повышению температуры в области расположения микросхемы. Игнорирование этих потерь в тепловой модели приводит к несоответствиям между расчетными тепловыми профилями и фактическим распределением температуры, наблюдаемым на экспериментах или моделях с учетом всех источников тепла. Использование упрощённой зависимости свойств материалов от средней температуры на заданном шаге, позволяет повысить адекватность моделирования.

IV. Список литературы

1. S. Race, A. Philipp, etc. Circuit-Based Electrothermal Modeling of SiC Power Modules with Nonlinear Thermal Models // IEEE Transactions on Power Electronics. – vol. 37, no. 7. – 2022. – P. 7965-7976.
2. G. Wang, Y. Zhou Thermal Management Modeling for β-Ga_2O_3-Highly Thermal Conductive Substrates Heterostructures // IEEE Transactions on Components, Packaging and Manufacturing Technology – vol. 12, no. 4. – 2022. – P. 638-646.

РАЗРАБОТКА АЛГОРИТМА ВОССТАНОВЛЕНИЯ КОМПЬЮТЕРНЫХ ИЗОБРАЖЕНИЙ В УСЛОВИЯХ НЕДОСТАТОЧНОЙ ОСВЕЩЕННОСТИ AIR-NET

Кукин Д. П., Купчина Е. В., Панасик А. А.
Кафедра вычислительных методов и программирования,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {kukin, e.kupchina, a.panasik}@bsuir.by

Описываются основные аспекты алгоритма восстановления изображений плохо освещенных объектов, обеспечивающего улучшение видимости деталей при сохранении естественного вида сцены.

Введение

Процесс формирования компьютерных изображений в условиях недостаточной освещенности сопровождается существенным ухудшением основных характеристик визуальных данных. Ключевой проблемой является значительное снижение отношения сигнал–шум, возникающее вследствие ограниченного количества фотонов, достигающих светочувствительной матрицы. Это приводит к преобладанию шумовой составляющей над полезным сигналом, что особенно выражено в теневых областях и создает серьезные препятствия для последующей обработки и анализа изображений.

I. Основные положения

В рамках решения задачи восстановления плохо освещенных изображений нейронная сеть служит основным инструментом для построения сложного нелинейного преобразования, способного компенсировать искажения, вызванные недостаточной освещенностью.

Особенностью предлагаемого подхода является использование глубокой сверхточной архитектуры AIR–Net, которая позволяет анализировать пространственные взаимосвязи между различными участками изображения.

Предлагаемая архитектура объединяет надежную U-Net-подобную структуру с принципиально новыми модулями обработки. В качестве базового каркаса используется проверенная энкодер-декодер архитектура, которая дополнена специализированными блоками, разработанными для решения специфических проблем слабоосвещенных изображений.

Алгоритм AIR-Net представляет собой специализированную нейросетевую архитектуру, разработанную для эффективного восстановления компьютерных изображений, полученных в условиях недостаточной освещенности. Основу алгоритма составляет каскадная обработка, сочетающая анализ глобального распределения освещенности с локальной коррекцией деталей.

Ключевой особенностью алгоритма является использование адаптивных карт освещенности, которые позволяют дифференцированно обрабатывать различные участки изображения.

Архитектура AIR-Net реализована как энкодер-декодерная структура с тремя специализированными модулями:

1. Модуль AIM (Adaptive Illumination Map) реализован на основе иерархической обработки изображения с использованием пирамиды масштабов. Принцип работы основан на последовательном анализе распределения яркости от глобального к локальному уровню. На каждом масштабе применяются сверточные слои с увеличивающейся областью восприятия, что позволяет выявлять как общие закономерности освещения, так и локальные аномалии. Для генерации карт освещенности используется комбинация линейных и нелинейных преобразований, обеспечивающая адаптацию к различным типам компьютерных изображений.
2. Блоки IRB (Illumination Restoration Block) построены по принципу остаточного обучения с использованием глубинных сепарабельных сверток. Основной идеей является раздельная обработка пространственных и глубинных характеристик изображения, что позволяет эффективно восстанавливать детали при сохранении вычислительной эффективности. В каждом блоке реализованы механизмы внимания, которые автоматически определяют значимость различных областей изображения для последующей коррекции.
3. Модуль MSFF (Multi-Scale Feature Fusion) использует принцип взвешенного объединения признаков различного масштаба. Реализация основана на параллельной обработке признаков с разным разрешением и их последующей интеграции с обучаемыми весовыми коэффициентами. Это позволяет сохранять как мелкие детали (текст, границы интерфейсов), так и глобальную структуру компьютерного изображения.

Общим принципом реализации всех модулей является ориентация на специфику компьютерных изображений, включая сохранение резких

границ, точности цветопередачи и структурной целостности графических элементов.

Для алгоритма AIR-Net использована система критериев, учитывающих структурные особенности компьютерных изображений. Основной акцент делается на оценке сохранения функциональных характеристик интерфейсов: читаемости текстовых элементов, узнаваемости пиктограмм, сохранения пропорций графических компонентов. Вводится параметр «коэффициент информационной сохранности», измеряющий степень распознаваемости ключевых элементов интерфейса после восстановления. Дополнительно оценивается естественность цветопередачи стандартных палитр и сохранение резкости векторных элементов.

Качество работы алгоритма AIR-Net существенно зависит от соответствия входных данных установленным требованиям. Для компьютерных изображений критически важным является использование форматов без потерь, что позволяет сохранить точность графических элементов и текстовых компонентов. Минимальное разрешение обеспечивает необходимый уровень детализации для корректного распознавания интерфейсных элементов. Стандартное цветовое пространство гарантирует согласованность обработки с отраслевыми нормами компьютерной графики.

Заключение

Проведенные экспериментальные исследования подтвердили принципиальные преимущества разработанного алгоритма при обработке именно компьютерных изображений в сравнении с универсальными решениями. Архитектура AIR-Net обеспечивает сохранение семантической целостности интерфейсных элементов, графических компонентов и текстовой информации, что было верифицировано в ходе тестирования на специализированных наборах данных.

Перспективные направления дальнейшего развития алгоритма связаны с оптимизацией вычислительной эффективности для работы в режиме реального времени, расширением функциональных возможностей для обработки видеопотоков компьютерных изображений, а также адаптацией архитектуры для работы с новыми форматами цифрового контента.

Список литературы

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений. – М. : Техносфера, 2012. – 1104 с.
2. Фомин Я. А., Корякин А. В. Алгоритмы обработки и анализа цифровых изображений. – СПб. : СПбГУ, 2018. – 216 с.
3. Сойфер В. А. Методы компьютерной обработки изображений. – М. : Физматлит, 2019. – 784 с.
4. Журавлев Ю. И., Рязанов В. В., Сенько О. В. Распознавание. Математические методы. Программная система. Практические применения. – М. : Физматлит, 2019. – 296 с.
5. Иванов Д. Н., Петров А. В. Сравнительный анализ алгоритмов улучшения изображений при низкой освещенности // Компьютерная оптика. – 2020. – Т. 44. – № 3. – С. 412-420.
6. Сидоров А. А., Козлов С. В. Нейросетевые методы восстановления зашумленных изображений // Информационно-управляющие системы. – 2021. – № 2. – С. 45-53.
7. Васильев К. Д., Николаев Д. П. Алгоритмы сегментации изображений на основе глубокого обучения // Обработка изображений и дистанционное зондирование. – 2019. – № 4. – С. 28-35.
8. Тихонов А. Н., Арсенин В. Я. Методы решения некорректных задач. – М. : Наука, 2019. – 288 с.
9. Желтов С. Ю., Корякин А. В. Методы регуляризации в задачах обработки изображений // Вычислительные методы и программирование. – 2020. – Т. 21. – С. 156-165.
10. Белов М. Л., Грановский Ю. В. Цифровая обработка сигналов и изображений в системах видеонаблюдения. – М. : Горячая линия-Телеком, 2018. – 280 с.
11. Андреев А. М., Павлов А. В. Алгоритмы улучшения качества изображений в системах технического зрения // Датчики и системы. – 2021. – № 3. – С. 34-41.
12. Кузнецов Е. П., Орлов Ю. В. Методика сравнительного анализа алгоритмов обработки изображений // Информационные процессы. – 2019. – Т. 19. – № 2. – С. 145-156.
13. Смирнов А. В., Федоров Р. К. Отечественная база данных изображений с различными условиями освещенности // Цифровая обработка сигналов и ее применение. – 2020. – Т. 22. – № 1. – С. 67-72.
14. Девятков В. В., Голубков А. Е. Применение нейросетевых технологий в задачах компьютерного зрения. – М. : Радио и связь, 2018. – 184 с.
15. Рыжов А. П., Михайлов Б. Б. Математические модели в системах обработки изображений. – М. : Изд-во МГТУ им. Н.Э. Баумана, 2019. – 320 с.
16. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention. – 2015. – P. 234-241.
17. Wang Z., Bovik A.C., Sheikh H.R., Simoncelli E.P. Image Quality Assessment: From Error Visibility to Structural Similarity // IEEE Transactions on Image Processing. – 2004. – Vol. 13. – No. 4. – P. 600-612.
18. Guo X., Li Y., Ling H. LIME: Low-Light Image Enhancement via Illumination Map Estimation // IEEE Transactions on Image Processing. – 2017. – Vol. 26. – No. 2. – P. 982-993.
19. Dabov K., Foi A., Katkovnik V., Egiazarian K. Image Denoising by Sparse 3D Transform-Domain Collaborative Filtering // IEEE Transactions on Image Processing. – 2007. – Vol. 16. – No. 8. – P. 2080-2095.
20. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization // arXiv preprint arXiv:1412.6980. – 2014.
21. OpenCV Library Documentation [Электронный ресурс]. – Режим доступа: `https://docs.opencv.org/` (дата обращения: 10.01.2024)
22. TensorFlow Framework Documentation [Электронный ресурс]. – Режим доступа: `https://www.tensorflow.org/` (дата обращения: 15.01.2024)
23. PyTorch Framework Documentation [Электронный ресурс]. – Режим доступа: `https://pytorch.org/` (дата обращения: 20.01.2024)

ПРЕДСТАВЛЕНИЕ ДАННЫХ ТРАССЫ ПАРАЛЛЕЛЬНЫХ ПРОЦЕССОВ В ВИДЕ ТАБЛИЦ ДЛЯ РАБОТЫ С БАЗОЙ ДАННЫХ SQLITE

Логинова И. П.
Объединённый институт проблем информатики Национальной академии наук Беларуси
Минск, Республика Беларусь
E-mail: irilog@mail.ru

При проведении параллельных вычислений в системе логической оптимизации функционально-структурных описаний цифровых устройств FLC2 формируются файлы трасс, содержащие набор определенных событий.Посредством средств, предоставляемых библиотеками платформы Qt, осуществляется доступ к данным трассы параллельных процессов, из которых выбирают наборы событий для аппроксимации, фильтрования и визуализации. Для хранения информации о событиях, необходимой для анализа и последующей оценки эффективности параллельных вычислений, используется сериализация трасс в базе данных SQLite.

Введение

Анализ трассы параллельного приложения проводится: с учетом количества событий в трассе; временного интервала выполнения приложения; числа процессов параллельного приложения; текущих настроек прикладных программ оптимизации, используемых в различных проектных процедурах. После чтения данных о событиях из файла трассы, их сортировки и фильтрации, осуществляется визуализация трассы параллельных процессов путем формирования графического представления. В основу реализации параллельных вычислений в системе FLC2 положена организация параллельных процессов. Каждый процесс выполняется на ядре многопроцессорной вычислительной системы (ВС) [3]. В процессах производится запуск прикладных программ с помощью команды, аргументы которой определяют имя программы и исходный объект. Исходные входные объекты в параллельных процессах являются независимыми. Организовать совокупности независимых объектов для проведения параллельных вычислений с использованием программ оптимизации позволяет среда системы FLC2, в которой возможно проведение проектных операций для преобразования исходного объекта в одноуровневую сеть независимых элементов, представляющих входные объекты для операций, выполняемых в отдельных процессах. Таким образом, параллельные вычисления в системе FLC2 реализуются посредством запуска параллельных процессов, в каждом из которых выполняется программа оптимизации для одного элемента сети [4].

I. Концепция «Модель/Представление» для обработки данных трассы

Для организации сложных комплексных приложений платформа Qt применяет архитектуру View/Model (Представление/Модель), которая является модификацией паттерна MVC. Qt содержит набор классов, которые для управления связями между данными и способами их представления используют архитектуру «Модель-/Представление» (M/V). Такое разделение позволяет отображать одни и те же данные в нескольких разных представлениях и реализовывать новые типы представлений без изменения базовых структур данных. На рис. 1 представлена архитектура, который использует одну Модель (источник данных) для одновременного преставления данных из нее двумя способами: в виде таблице и в виде горизонтальных диаграмм с накоплением. Согласно данной схеме всю работу по обновлению данных между источником данных и Моделью берет на себя Представление.

Рис. 1 – Архитектура M/V при работе с трассой параллельных процессов

Процесс работы приложения в этой архитектуре выглядит следующим образом. Модель взаимодействует с источником данных (трасса параллельных процессов), обеспечивая интерфейс для других компонентов архитектуры. Представление через модель может взаимодействовать с данными, в частности, может получать данные и отображать их. Для отображения в представлении применяется программа визуализации [3]. Когда данные изменяются, программа визуализации напрямую взаимодействует с моделью, уведомляя ее об изменении данных.

II. Представление данных трассы параллельных процессов в виде таблицы

Платформа Qt предоставляет набор классов, которые реализуют архитектуру M/V для управления взаимосвязями между данными и различными представлениями. Архитектура обеспечивает гибкость настройки представления и позволяет комбинировать источники данных с широким спектром представлений. Базовым классом всех моделей является класс QAbstractItemModel. Этот класс определяет интерфейс, который используется для доступа к данным. При необходимости представления данных в виде таблицы при реализации модели, как правило, используется класс QAbstractTableModel. Представления (Views) реализуются объектами классов, унаследованных от абстрактного класса QAbstractItemView.Например, в программе визуализации [3] объект класса QTableView отображает данные из модели в виде таблицы.

На рисунках 2–4 приведены представления трассы параллельных процессо, полученной в результате проведения логической оптимизации в системе FLC2 для 26-ти объектов одноуровневой сети, выполненных на 8-ми ядрах процессора.

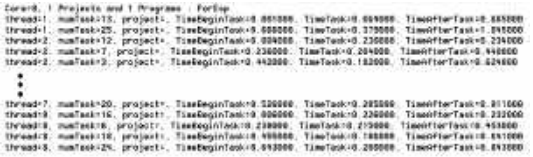
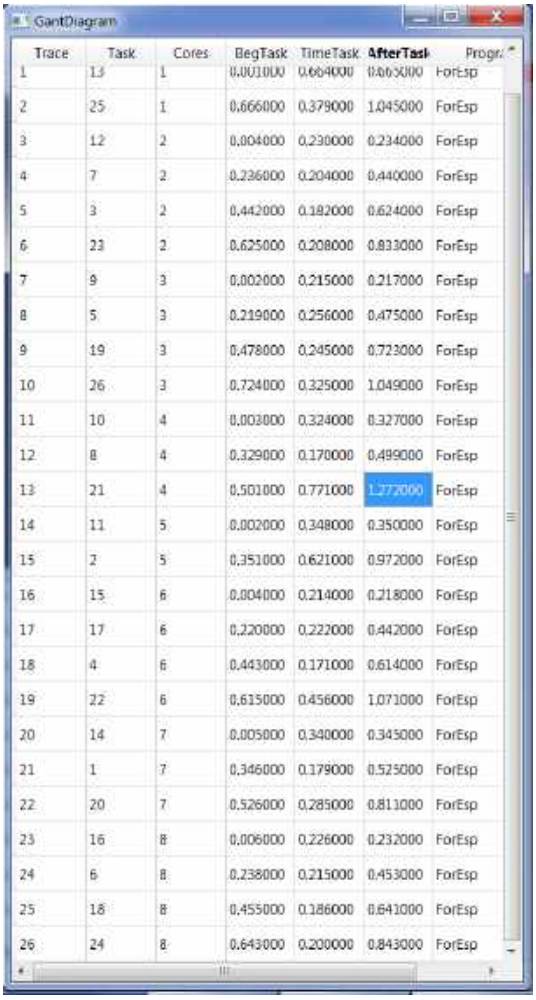


Рис. 2 – Фрагмент трассы параллельных процессов

На рис. 3 представление в виде горизонтальной диаграммы с накоплением демонстрирует распределение вычислительной нагрузки по ядрам процессора. На На рис. 4 представление в виде таблицы содержит всю информацию трассы с указанием всех событий, начиная от идентификатора процесса, номера задачи, номера ядра и всех временных меток начала, длительности и завершения каждой задачи.

Рис. 3 – Представление временных меток трассы в виде диаграммы Ганта



Trace	Task	Cores	BegTask	TimeTask	AfterTask	Progr
1	13	1	0.001000	0.664000	0.665000	ForEsp
2	25	1	0.666000	0.379000	1.045000	ForEsp
3	12	2	0.004000	0.230000	0.234000	ForEsp
4	7	2	0.236000	0.204000	0.440000	ForEsp
5	3	2	0.442000	0.182000	0.624000	ForEsp
6	23	2	0.625000	0.208000	0.833000	ForEsp
7	9	3	0.002000	0.215000	0.217000	ForEsp
8	5	3	0.219000	0.256000	0.475000	ForEsp
9	19	3	0.478000	0.245000	0.723000	ForEsp
10	26	3	0.724000	0.325000	1.049000	ForEsp
11	10	4	0.003000	0.324000	0.327000	ForEsp
12	8	4	0.329000	0.170000	0.499000	ForEsp
13	21	4	0.501000	0.771000	1.272000	ForEsp
14	11	5	0.002000	0.348000	0.350000	ForEsp
15	2	5	0.351000	0.621000	0.972000	ForEsp
16	15	6	0.004000	0.214000	0.218000	ForEsp
17	17	6	0.220000	0.222000	0.442000	ForEsp
18	4	6	0.443000	0.171000	0.614000	ForEsp
19	22	6	0.615000	0.456000	1.071000	ForEsp
20	14	7	0.005000	0.340000	0.345000	ForEsp
21	1	7	0.346000	0.179000	0.525000	ForEsp
22	20	7	0.526000	0.285000	0.811000	ForEsp
23	16	8	0.006000	0.226000	0.232000	ForEsp
24	6	8	0.238000	0.215000	0.453000	ForEsp
25	18	8	0.455000	0.186000	0.641000	ForEsp
26	24	8	0.643000	0.200000	0.843000	ForEsp

Рис. 4 – Представление трассы в виде таблицы

Данные можно хранить как в модели, так и в базе данных (БД). Qt предлагает несколько моделей, которые используются для доступа к базам данных: QSqlQueryModel, SqlTableModel и QSqlRelationalTableModel. Использование моделей БД и таблиц позволяют организовать сериализацию данных трассы в БД SQLite.

III. Список литературы

1. Бибило, П. Н. Система логической оптимизации функционально-структурных описаний цифровых устройств на основе продукционно-фреймовой модели представления знаний / П. Н. Бибило, В. И. Романов // Проблемы разработки перспективных микро- и наноэлектронных систем – (МЭС-2020) : IX Юбилейная Всероссийская научно-техническая конференция : Сборник трудов ; под общей ред. А. Л. Стемпковского. – Москва : ИППМ РАН, 2021 Москва : ИППМ РАН, 2020. – С. 9–16.
2. Бибило, П. Н. Экспериментальное сравнение эффективности программ минимизации систем булевых функций в классе дизъюнктивных нормальных форм / П. Н. Бибило, И. П. Логинова // Информатика. – 2022. – Т. 19, № 2. – С. 26-55.
3. Логинова, И. П. Построение и визуальный анализ параллельных решений в системе логической оптимизации функционально-структурных описаний дискретных схем // Компьютерное проектирование в электронике (EDA Conference 2024) : сб. трудов Междунар. науч.-практ. конф. (Республика Беларусь, г. Минск, 28 ноября 2024 г.) Минск : БГУИР, 2024. – С. 108–111.

ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

Панасик А. А., Кукин Д. П., Купчина Е. В.
Кафедра вычислительных методов и программирования,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {a.panasik, kukin, e.kupchina}@bsuir.by

Статья представляет собой обзор современных методов сегментации изображений, от классических алгоритмов до архитектур глубокого обучения. Проводится систематический анализ различных подходов.

Введение

Сегментация изображений – это процесс разделения цифрового изображения на множественные сегменты (множества пикселей) с целью упрощения или изменения представления изображения для более простого анализа [1]. По сути, это задача классификации каждого пикселя изображения. В отличие от простого детектирования объектов, сегментация предоставляет точные границы и маски объектов, что необходимо для детального понимания сцены.

Актуальность технологии подтверждается широким спектром применений:

1. В автономных транспортных системах для идентификации элементов дорожной сцены [2].
2. В медицинской диагностике для точного выделения анатомических структур и патологий [3].
3. В дистанционном зондировании для классификации земных покровов и мониторинга изменений.

Современные подходы к сегментации разделяются на классические алгоритмы, основанные на анализе низкоуровневых признаков, и методы глубокого обучения с использованием CNN для автоматического выделения признаков.

I. Классические методы сегментации

Методы, которые опираются на низкоуровневые признаки изображения, такие как яркость, цвет, текстура или градиент.

Пороговые методы:

Пороговая сегментация является одним из простейших методов. Ее идея заключается в разделении пикселей на классы на основе их интенсивности путем выбора одного или нескольких пороговых значений [4].

Метод Оцу: Алгоритм автоматически вычисляет оптимальный порог, минимизирующий внутриклассовую дисперсию или, что эквивалентно, максимизирующий межклассовую дисперсию. Метод эффективен для изображений с бимодальной гистограммой.

Методы, основанные на кластеризации:

Методы рассматривают пиксели как точки в пространстве признаков и группируют их в кластеры.

K-means: Алгоритм итеративно назначает пиксели k кластерам, центроиды которых постоянно пересчитываются. Главным недостатком является необходимость априорного задания числа кластеров k.

Mean Shift: Алгоритм не требует задания числа кластеров. Он итеративно смещает точки к режиму плотности распределения данных, что позволяет находить кластеры произвольной формы.

Методы, основанные на областях (Region-based):

Подходы этой группы основываются на поиске связных областей с однородными свойствами.

Region Growing: Алгоритм начинается с «затравки» (seed point) и последовательно присоединяет соседние пиксели, если они удовлетворяют заданному критерию однородности. Эффективность сильно зависит от выбора начальной точки и критерия.

Split and Merge: Метод начинается с всего изображения как одного региона, который рекурсивно разбивается на части, если регион неоднороден. Затем соседние однородные регионы объединяются.

Методы на основе активных контуров (Active Contours):

Модель «змей» (snakes) представляет собой энергетически-минимизирующую сплайну, которая эволюционирует под действием внутренних и внешних сил [5].

Внутренние силы отвечают за гладкость контура.

Внешние силы «притягивают» контур к границам объекта (например, вычисляются через градиент изображения).

Модель способна давать гладкие и замкнутые контуры, но требует аккуратной инициализации и может «застревать» в локальных минимумах. Основная цена за это — требование к большим размеченным датасетам и значительные вычислительные затраты на обучение и инференс.

II. Методы на основе глубокого обчения

С появлением больших размеченных датасетов и мощных GPU, глубокое обучение произвело революцию в сегментации изображений.

Семантическая сегментация (Semantic Segmentation):

Задача присвоения каждой пиксельной метки класса без разделения отдельных экземпляров объектов.

Fully Convolutional Networks (FCN) [6]: Замена полносвязных слоев на сверточные позволила обрабатывать изображения произвольного размера и генерировать карты сегментации. Для восстановления разрешения используются транспонированные свертки.

U-Net [7]: U-образная архитектура, ставшая стандартом в биомедицинской сегментации. Состоит из: энкодера (сжатие и извлечение признаков); декодера (восстановление разрешения); Skip-connections для передачи пространственной информации, утраченной при пулинге.

DeepLab Family - семейство архитектур, характеризующееся тремя ключевыми усовершенствованиями: Atrous (Dilated) Convolution, который увеличивает поле обзора без потери разрешения, Atrous Spatial Pyramid Pooling (ASPP), который обеспечивает многомасштабную обработку объектов, Encoder-Decoder with Depthwise Separable Convolution в котором используется декодер для уточнения границ объектов.

Сегментация экземпляров (Instance Segmentation):

Более сложная задача, которая требует не только классификации пикселей, но и разделения отдельных объектов одного класса.

Mask R-CNN: Является расширением Faster R-CNN [9], дополненное третьей ветвью для генерации бинарных масок объектов с помощью небольшой FCN. Ключевое улучшение - замена RoIPooling на RoIAlign - обеспечило существенное повышение точности пиксельного соответствия масок.

Классические методы эффективны в условиях ограниченных вычислительных ресурсов, для задач с сильной дифференциацией по яркости/-цвету или когда обучение модели невозможно.

Методы на основе глубокого обучения демонстрируют точность на сложных, зашумленных данных с неочевидными границами объектов. Они способны учитывать высокоуровневый контекст сцены.

Заключение

Проведен системный обзор методов сегментации изображений: от простых пиксельных и граничных подходов к сложным контекстнозависимым моделям глубокого обучения. Доминирование архитектур на основе CNN, является установившимся фактом для задачи семантической сегментации, в то время как Mask R-CNN и его производные лидируют в Instance Segmentation. Текущие исследования в области направлены на решение следующих проблем:

Эффективность: Разработка более легких и быстрых моделей для работы на мобильных и встраиваемых устройствах.

Новые архитектуры: Интеграция трансформеров (Vision Transformers, SegFormer) в задачи сегментации, что позволяет лучше моделировать глобальные зависимости в изображении.

Интерпретируемость: Создание моделей, не только дающих точный результат, но и позволяющих понять, на основе каких признаков было принято решение.

Таким образом, область сегментации изображений продолжает активно развиваться, двигаясь в сторону создания более умных, эффективных и универсальных моделей, способных работать в условиях реального мира.

Список литературы

1. Szeliski, R. Computer Vision: Algorithms and Applications / R. Szeliski. – Springer, 2010.
2. Chen, L. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation / L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A. L. Yuille // Computer Vision – ECCV 2018. – 2018. – P. 801-818.
3. Ronneberger, O. U-Net: Convolutional Networks for Biomedical Image Segmentation / O. Ronneberger, P. Fischer, T. Brox // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. – 2015. – P. 234-241.
4. Otsu, N. A Threshold Selection Method from Gray-Level Histograms / N. Otsu // IEEE Transactions on Systems, Man, and Cybernetics. – 1979. – Vol. 9(1). – P. 62-66.
5. Kass, M. Snakes: Active contour models / M. Kass, A. Witkin, D. Terzopoulos // International Journal of Computer Vision. – 1988. – Vol. 1(4). – P. 321-331.
6. Long, J. Fully convolutional networks for semantic segmentation / J. Long, E. Shelhamer, T. Darrell // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2015. – P. 3431-3440.
7. He, K. Mask R-CNN / K. He, G. Gkioxari, P. Dollár, R. Girshick // Proceedings of the IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2961-2969.
8. Vaswani, A. Attention Is All You Need / A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin // Advances in Neural Information Processing Systems (NeurIPS). – 2017. – Vol. 30.
9. Xie, E. SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, P. Luo // Advances in Neural Information Processing Systems (NeurIPS). – 2021. – Vol. 34. – P. 12077-12090.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ ПОЯСНИЧНОГО ОТДЕЛА ПОЗВОНОЧНИКА ЧЕЛОВЕКА НА ОСНОВЕ АНАЛИЗА КТ-ИЗОБРАЖЕНИЙ

Семенченя Т. С., Курочка К. С.
Кафедра информационные технологи,
Гомельский государственный технический университет имени П. О. Сухого
Гомель, Республика Беларусь
E-mail: levts@gstu.by

В данной работе предлагается исследование напряженно-деформированного состояния моделей поясничного отдела позвоночника человека с различными нагрузками. Построенные, на основе анализа изображений компьютерной томографии, модели могут быть использованы для анализа состояния сегмента позвоночника в норме, при нестандартных нагрузках и патологических изменениях.

Введение

Позвоночник представляет собой сложную трехмерную конструкцию, анатомические особенности которой обеспечивают движения как изолированно в любой из трех плоскостей, так и одновременно в нескольких. Наиболее активным является поясничный отдел позвоночника человека. При выборе тактики лечения, врачи стараются определить морфологические изменения в позвоночнике, которые ответственны за клинику заболевания [1, 2].

В последнее время с целью усовершенствования методик лечения и хирургических вмешательств применяется математическое моделирование. Модель дает значительно больше информации о биомеханике позвоночника, чем можно получить современными средствами измерений. При анализе поведения модели поясничного отдела позвоночника человека варьирование значений параметров структур позвоночного столба дает возможность определить влияние каждого из них на проявление синдрома и возможные последствия хирургических вмешательств.

Одним из наиболее информативных методов изучения напряженно-деформированного состояния является метод конечных элементов. К плюсам данного метода можно отнести: возможность моделирования тел с различными материалами, хорошая аппроксимация криволинейных границ, возможность уточнения результатов путем измельчения сетки конечных элементов, учет различных граничных условий.

Экспериментальная часть

Для исследования моделей было разработано программное средство, которое строит 3D-модель содержащую позвонки и межпозвонковые диски, на основе анализа КТ-изображений [3]. Полученная модель представлена на рисунке (см. рис. 1).

Рис. 1 – Математическая модель поясничного отдела позвоночника человека

Для сравнения НДС провели исследования на двух моделях поясничного отдела позвоночника человека – с межпозвонковыми дисками и с отсутствием одного межпозвонкового диска.

Для исследования использовали специализированный программный продукт ANSYS.

При исследовании НДС в сгенерированных программным средством конечноэлементных моделях предполагалось, что позвонки представляют собой линейноупругое тело, а межпозвонковые диски – упругопластическое тело.

Для проведения расчета НДС необходимо определить параметры материалов, из которых состоят позвонки и межпозвонковые диски. Материалы, используемые при построении модели, считаются однородными и изотропными. Модуль упругости и коэффициент Пуассона приведены в таблице 1.

Таблица 1 – Физические характеристики материалов, используемые при построение конечно-элементной модели поясничного отдела позвоночника человека

Материал	Модуль упругости Юнга, МПа	Коэффициент Пуассона
Кость	100	0.2
Межпозвонковый диск	5	0.49

Для проведения анализа нужно на основе трехмерной модели создать конечно-элементную сетку, так как программный комплекс использует конечные элементы для проведения расчетов. Сетка создается внутри программы, то есть не

требует никакого дополнительного программного обеспечения.

После создания сетки требуется указать ограничения, в данном случае необходимо только два: масса, которая имитирует давление на позвоночник человека в вертикальном положении сверху; закрепление, которое необходимо для того, чтобы математическая модель была корректной и при усилии на которую она не двигалась. В данном случае необходимо закрепить нижнюю поверхность позвоночника в районе крестца.

При учете этих ограничений математическая модель будет имитировать настоящий позвоночник в упрощенном виде.

Результаты и их обсуждения

Для проведения экспериментов использовалась серия КТ-изображений поясничного отдела позвоночника человека с различными нагрузками и моделями (с межпозвонковыми дисками и без них).

В процессе проведения напряженно-деформированного анализа получены значения напряжений, которые указаны в таблице 2.

Таблица 2 – Результаты напряженно-деформированного анализа поясничного отдела позвоночника человека

Нагрузка(кг)	Среднее напряжение модели без межпозвонкового диска (МПа) в Ansys	Среднее напряжение модели с межпозвонковыми дисками (МПа) в Ansys
40	3.95	0.88
65	6.42	1.44
85	8.4	1.88
120	11.85	2.66
150	14.81	3.33

По результатам расчета напряженно-деформированного анализа поясничного отдел позвоночника человека можно заметить, что зависимость напряжения прямо пропорциональна прикладываемой нагрузке. При этом в модели поясничного отдела позвоночника человека без межпозвонковых дисков среднее напряжение в несколько раз превышает значения модели с межпозвонковыми дисками, на основании чего можно сделать вывод, что трехмерная модель является корректной, а результаты расчеты приближены к реальному поведению настоящего позвоночника.

На рисунке показано исследование НДС поясничного отдела позвоночника человека с межпозвонковыми дисками (см. рис. 2). Анализ результатов показал, что наиболее напряженными участками являются области ножек дуги позвонков.

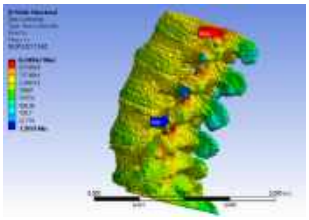


Рис. 2 – Распределение напряжения в модели поясничного отдела позвоночника с межпозвонковыми дисками

На рисунке показана трехмерная модель с распределением напряжения в случае отсутствия одного межпозвонкового диска (см. рис. 3).

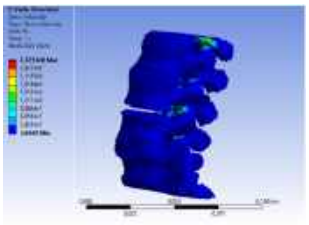


Рис. 3 – Распределение напряжения в модели поясничного отдела позвоночника с отсутствием одного межпозвонкового диска

На модели с напряжением можно увидеть области локализации повышенного напряжения. Одно из таких мест – место, где нет межпозвонкового диска. При необходимости замены удаленного межпозвонкового диска на протез, можно перед операцией провести анализ, в котором можно выяснить, как отразятся изменения на позвоночнике и на что нужно обратить внимание.

Заключение

Разработанная математическая модель и предложенная методика задания ее параметров позволяет повысить значимость расчетов. Конечно-элементную модель поясничного отдела позвоночника человека можно использовать для исследования напряженно-деформированного состояния при имитировании разнообразных патологий под действием вертикальной нагрузки, а также при моделировании различных хирургических вмешательств, тестировании и оптимизации дизайна искусственных позвонков и устройств для фиксации костей.

1. Чуйко, А. Н. Приближенный анализ анатомии, механических характеристик и напряженно-деформированного состояния позвоночника человека / А. Н. Чуйко // Травма. – 2014. – №6. – С. 100–109.
2. Цитко, Е. Л. Рентгенометрическая оценка кинематики пояснично-крестцового отдела позвоночника при остеохондрозе с помощью программного средства «ВОЛОТ» / Е. Л. Цитко // Проблемы здоровья и экологии. – 2017. – №4. – С. 35–41.
3. Курочка, К. С. Construction of an individual geometric 3D model of the lumbar spine of a person based on the analysis of medical images / К. С. Курочка, Т. С. Семенченя // Сборник научных трудов «Открытые семантические технологии проектирования интеллектуальных систем», 19 – 22 февраля 2020 г. – Минск: БГУИР, 2020. – с. 291–297.

МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ТРАЕКТОРИИ ПОЛЕТА ОДНОСТУПЕНЧАТОЙ ТВЕРДОТОПЛИВНОЙ БАЛЛИСТИЧЕСКОЙ РАКЕТЫ МАЛОЙ ДАЛЬНОСТИ

Свинарский М. В.
Служба системных исследований в области радиотехники
Государственное предприятие «Центр радиотехники Национальной академии наук Беларуси»
Минск, Республика Беларусь
E-mail: mechislav1993@gmail.com

Актуальность работы обусловлена широкой распространенностью баллистических ракет малой дальности (БРМД) с ракетными двигателями твердого топлива (РДТТ) и необходимостью точного прогнозирования их траекторий для разработки и модернизации средств противодействия. В работе представлена математическая модель траектории полета БРМД, основанная на системе дифференциальных уравнений движения центра масс. Данная модель учитывает тягу двигателя, аэродинамическое сопротивление, гравитацию, силу Кориолиса и центробежную силу. Для перехода от теоретической модели к вычислительной введен ряд допущений. Представлены графики изменения основных координат и параметров траектории в процессе полета.

Введение

Современные баллистические ракеты представляют собой одну из наиболее серьезных угроз национальной безопасности вследствие своих тактико-технических характеристик. Высокая скорость полета, значительная дальность действия, устойчивость к электромагнитным помехам и способность преодолевать системы противоракетной обороны делают их эффективным средством поражения стратегически важных объектов. Особую опасность представляет возможность оснащения таких ракет различными типами боевых частей, включая ядерные.

Эффективное противодействие баллистическим ракетам требует комплексного подхода, основанного на современных научных методах и технологиях. Задачами первостепенной важности являются своевременное обнаружение пуска, точное сопровождение на всех участках траектории и прогнозирование точки падения. Решение этих задач осуществляется с применением сложных алгоритмов, центральное место в которых занимают методы математического моделирования. Теоретической основой для построения таких алгоритмов является адекватная математическая модель траектории полета баллистической ракеты. При этом степень детализации и физической обоснованности этой модели напрямую влияет на точность и надежность решения перечисленных прикладных задач. В связи с этим разработка таких моделей, в частности, для широко распространенного класса одноступенчатых твердотопливных ракет малой дальности, представляет собой актуальную научно-практическую задачу.

I. Теоретические основы моделирования

В зависимости от действующих силовых факторов траекторию полета баллистической ракеты принято разделять на три основных участка [1]: активный, баллистический и участок возвращения в атмосферу. На активном участке движение происходит под действием тяги двигателя, силы тяжести и аэродинамического сопротивления. После прекращения работы двигателя ракета продолжает полет по баллистической траектории, где основными действующими силами являются гравитация и незначительное сопротивление в верхних слоях атмосферы. На этапе возвращения в атмосферу аэродинамическое сопротивление становится доминирующей силой, наряду с силой тяжести, вызывая интенсивное торможение объекта.

Поскольку движение рассматривается относительно вращающейся Земли, для адекватного описания траектории необходимо использовать неинерциальную систему отсчета. В такой системе, помимо основных физических сил, необходимо учитывать [2]: силу Кориолиса и центробежную силу. С учетом изложенного, векторное уравнение движения центра масс баллистической ракеты в геоцентрической системе координат может быть представлено в следующем виде:

$$\mathbf{A}_\Sigma = \mathbf{A}_\mathrm{T} + \mathbf{A}_\mathrm{A} + \mathbf{A}_\mathrm{G} + \mathbf{A}_\mathrm{C} + \mathbf{A}_\mathrm{F},$$

где $\mathbf{A}_\mathrm{T}$ – вектор ускорения обусловленный тягой двигателя, $\mathbf{A}_\mathrm{A}$ – вектор ускорения обусловленный аэродинамическим сопротивлением, $\mathbf{A}_\mathrm{G}$ – вектор ускорения обусловленный гравитационным полем Земли, $\mathbf{A}_\mathrm{C}$ – вектор ускорения обусловленный силой Кориолиса, $\mathbf{A}_\mathrm{F}$ – вектор ускорения обусловленный центробежной силой.

Важно отметить, что на практике учет всех возможных факторов, формирующих перечисленные ускорения является сложной задачей. Такие факторы, как нестационарность атмосферы, упругие колебания конструкции, изменение аэродинамических коэффициентов в широком диапазоне углов атаки и чисел Маха, точное знание программы работы двигателя и многие другие,

вносят значительную неопределенность. В связи с этим общепринятым подходом при баллистическом проектировании является введение в математическую модель обоснованных ограничений и допущений, позволяющих достичь приемлемого компромисса между точностью прогноза и вычислительной сложностью модели [1].

В связи с этим приняты следующие допущения при разработке математической модели:

- ракета аппроксимируется материальной точкой (учтены только уравнения движения центра масс);
- углы крена и скольжения ракеты в процессе полета равны нулю;
- модель гравитационного поля соответствует центральному ньютоновскому;
- атмосфера неподвижная с экспоненциальным законом изменения плотности;
- учет аэродинамических сил с поправкой на кривизну Земли;
- управление по углу тангажа осуществляется по программному закону от относительной массы топлива;
- аэродинамические коэффициенты аппроксимируются зависимостями от числа Маха и угла атаки;
- интегрирование системы дифференциальных уравнений движения осуществляется явным методом Эйлера.

II. Результаты математического моделирования

Моделирование проводилось для типовой БРМД со следующими параметрами: стартовая масса 3800 кг, масса боевой части 480 кг, время работы РДТТ 26 с, удельный импульс в 2460 м/с. программный закон изменения угла тангажа от 90° до 44°. На рисунках 1-4 представлены результаты математического моделирования, показывающие динамику изменения основных координат и параметров движения баллистической ракеты на всех этапах полета.

Анализ полученных результатов подтвердил, что разработанная математическая модель корректно воспроизводит физику процесса на всех этапах: активном участке, баллистическом подъеме и нисходящей ветви. Верификация модели показала хорошую сходимость с ожидаемыми характеристиками: расчетная дальность полета составила 515 км, максимальная скорость – 2532 м/с, а боковое смещение траектории под действием силы Кориолиса – 8.6 км. Таким образом, полученная модель может быть использована для предварительного баллистического проектирования, оценки влияния различных факторов на траекторию полета, решения задачи сопровождения, а также для прогнозирования точки старта и падения.

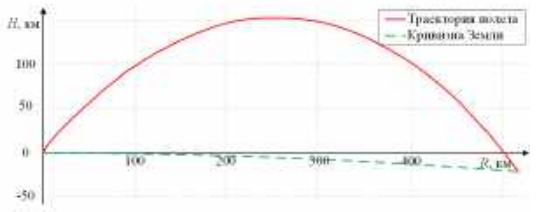


Рис. 1 – Траектория полета БРМД в вертикальной плоскости

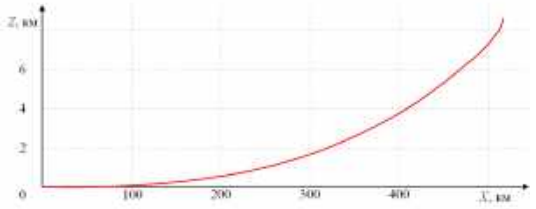


Рис. 2 – Траектория полета БРМД в горизонтальной плоскости

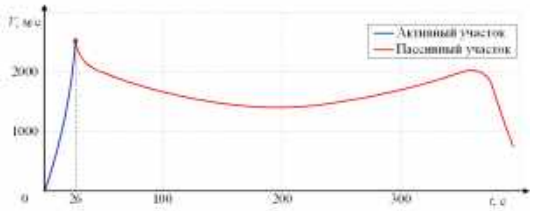


Рис. 3 – Изменение скорости БРМД в процессе полета

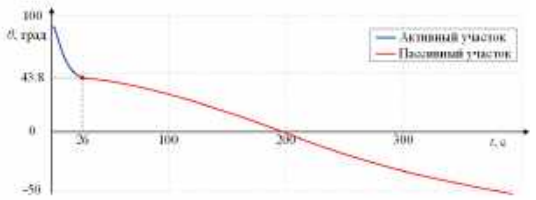


Рис. 4 – Изменение угла тангажа БРМД в процессе полета

Список литературы

1. Сихарулидзе, Ю. Г. Баллистика и наведение летательных аппаратов [Электронный ресурс] / Ю. Г. Сихарулидзе. – 2-е изд. (эл.). – М.: БИНОМ. Лаборатория знаний, 2013. – 407 с.
2. Лысенко, Л. Н. Внешняя баллистика: учебное пособие / Л. Н. Лысенко. – М.: Издательство МГТУ им. Н. Э. Баумана, 2018. – 328 с.

АНАЛИТИЧЕСКИЕ СВОЙСТВА РЕШЕНИЙ ДВУХ ДИНАМИЧЕСКИХ СИСТЕМ С ХАОТИЧЕСКИМ ПОВЕДЕНИЕМ

Цегельник В. В.
Кафедра высшей математики,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: tsegvv@bsuir.by

Исследован характер возможных подвижных точек решений двух динамических систем, обладающих (при определенных значениях входящих в них постоянных параметров) сложным хаотическим поведением.

Введение

В работах [1], [2] выполнено качественное исследование решений двух динамических систем. Численное моделирование показало, что каждая из рассмотренных систем при определенных значениях входящих в них постоянных параметров обладает сложным хаотическим поведением.

I. Постановка задачи

Целью настоящей работы является исследование характера подвижных точек (т.е. точек, положение которых зависит от начальных условий) решений систем дифференциальных уравнений

$$\dot{x} = -ax + yz, \tag{1}$$

$$\dot{y} = -bx + zx, \tag{2}$$

$$\dot{z} = z - xy \tag{3}$$

и

$$\dot{x} = p, \tag{4}$$

$$\dot{p} = -x - \alpha(u^2 + v^2)x - \beta p. \tag{5}$$

$$\dot{u} = -(\delta - x^2)v - \frac{\kappa}{2}u + l, \tag{6}$$

$$\dot{v} = (\delta - x^2)u - \frac{\kappa}{2}v. \tag{7}$$

В системах (1)-(3), (4)-(7) a, b и α, β, δ, κ, l – произвольные постоянные параметры; x, y, z, p, u, v – функции независимой переменной t.

Система (1)-(3) принадлежит [1] к классу трехмерных диссипативных систем с квадратичной нелинейностью,обладающих кристаллографической точечной симметрией. При определенных значениях $a < 0$, $b < 0$ в (1)-(3) наблюдаются хаотические аттракторы.

Характерной особенностью системы (4)-(7) с качественной точки зрения является наличие в ней гиперхаоса при определенных значениях параметров, в частности при $l = \frac{1}{2}$ [2].

Гиперхаос – это более сильная форма хаоса, при которой несколько степеней свободы одновременно способствуют экспоненциальному расхождению малых изменений. В [2] гиперхаос рассматривался в контексте резонаторной оптомеханики, в которой свет внутри оптического резонатора взаимодействует с подвешенной колеблющейся массой.

II. Алгоритм

Для решения поставленной задачи (в предположении, что неизвестные функции, входящие в системы (1)-(3), (4)-(7), являются функциями комплексной переменной t) использован тест Пенлеве [3], представляющий набор условий, необходимых для отсутствия у общего решения системы дифференциальных уравнений подвижных критических особых точек (свойство Пенлеве).

Выводы

1. Применение первого шага теста Пенлеве показывает, что все 3 компоненты решения системы (1)-(3) могут иметь простые подвижные полюсы, в окрестности которых справедливы формальные разложения

$$x = \frac{x_{-1}}{\tau} + x_0 + x_1\tau + \dots, \tag{8}$$

$$y = \frac{y_{-1}}{\tau} + y_0 + y_1\tau + \dots, \tag{9}$$

$$z = \frac{z_{-1}}{\tau} + z_0 + z_1\tau + \dots, \tag{10}$$

где $\tau = t - t_0$, t_0 – произвольное.

Подстановка разложений (8)-(10) в каждое из уравнений системы (1)-(3) и приравнивание коэффициентов при наинизшей степени t позволяет установить, что а) $x_{-1} = y_{-1}$, $z_{-1} = 1$ либо б) $x_{-1} = y_{-1}$, $z_{-1} = -1$.

Применение второго шага теста Пенлеве позволяет получить резонансное уравнение с корнями $r_1 = -1$, $r_2 = r_3 = 2$.

Применение третьего шага теста Пенлеве показывает, что для наличия в разложениях (8)-(10) трех произвольных постоянных необходимо, чтобы $a = b$.

Легко проверить, что при $a = b$ система (1)-(3) имеет первый интеграл

$$x^2 - y^2 = C^2 e^{-2at}, \tag{11}$$

где C – произвольная постоянная.

С помощью преобразований

$$x = \frac{C}{2} e^{-at} \left(u + \frac{1}{u} \right), y = \frac{C}{2} e^{-at} \left(u - \frac{1}{u} \right) \tag{12}$$

для определения функции u получим уравнение

$$u\ddot{u} = \dot{u}^2 + u\dot{u} - \frac{C^2}{4} e^{-2at}(u^4 - 1). \tag{13}$$

Поскольку $a \neq 0$, то согласно [4] уравнение (13) не является уравнением Пенлеве-типа.

Теорема 1. Система (1)-(3) в случае $a = b \neq 0$ не является системой Пенлеве-типа.

Теорема 2. Система (1)-(3) в случае $a \neq b$ не проходит тест Пенлеве.

2. Система (4)-(7) в случае $l = 0$ принимает вид

$$\dot{x} = p, \tag{14}$$

$$\dot{p} = -x - \alpha(u^2 + v^2)x - \beta p, \tag{15}$$

$$\dot{u} = -(\delta - x^2)v - \frac{\kappa}{2}u, \tag{16}$$

$$\dot{v} = (\delta - x^2)u - \frac{\kappa}{2}v \tag{17}$$

и имеет первый интеграл

$$u^2 + v^2 = He^{-\kappa t}, \tag{18}$$

где H – произвольная постоянная.

Второе уравнение системы (14)-(17) с учетом (18) принимает вид

$$\dot{p} = -\left(1 + \alpha H e^{-\kappa t}\right) x - \beta p.$$

Поскольку $p = \dot{x}$, то относительно x имеет уравнение

$$\ddot{x} + \beta\dot{x} + \left(1 + \alpha H e^{-\kappa t}\right) x = 0. \tag{19}$$

Уравнение (19), как линейное, вообще не имеет подвижных особых точек. Кроме того, заменой $e^{-\kappa t} = \rho$ оно сводится к уравнению (1.a) из [5, стр. 401], решения которого в общем случае выражаются через бесселевы функции первого и второго рода, а при $\alpha = 0$ оно интегрируется в элементарных функциях.

Из уравнения (16) следует, что $v = \frac{-\kappa\rho u' + \frac{\kappa}{2}u}{x^2 - \delta}$, где $u' = \frac{\mathrm{d}u}{\mathrm{d}\rho}$. Тогда из (17) находим уравнение для функции u:

$$\kappa^2\rho^2 u'^2 - \kappa^2\rho u u' + \left((x^2 - \delta)^2 + \frac{\kappa^2}{4}\right) u^2 - H\rho = 0. \tag{20}$$

Интегрирование уравнения (20) согласно [5, стр. 346] связано с интегрированием уравнения Абеля, общее решение которого содержит подвижные критические особые точки.

Теорема 3. Система (14)-(17) не является системой Пенлеве, хотя компоненты x, p ее общего решения вообще не имеют подвижных особых точек.

3. Рассмотрим систему (4)-(7) в случае $l \neq 0$. Следуя первому шагу алгоритма теста Пенлеве и подставляя формальные разложения

$$x = \frac{x_{-m}}{\tau^m} + \frac{x_{-m+1}}{\tau^{m-1}} + \ldots,$$

$$p = \frac{p_{-n}}{\tau^n} + \frac{p_{-n+1}}{\tau^{n-1}} + \ldots,$$

$$u = \frac{u_{-s}}{\tau^s} + \frac{u_{-s+1}}{\tau^{s-1}} + \ldots,$$

$$v = \frac{v_{-q}}{\tau^q} + \frac{v_{-q+1}}{\tau^{q-1}} + \ldots,$$

где m, n, s, q – неотрицательные целые числа, среди которых хотя бы одно отлично от нуля, получаем условие $m = \frac{1}{2}$, $n = \frac{3}{2}$, $s = q = 1$. Таким образом, справедлива

Теорема 4. Система (4)-(7) в случае $l \neq 0$ не проходит тест Пенлеве.

Замечание. Представляет интерес исследование аналитических свойств решений системы

$$\dot{x} = ac + yz,$$

$$\dot{y} = by + xz,$$

$$\dot{z} = z + xy,$$

в которой, согласно [1], хаотическое поведение невозможно.

III. Список литературы

1. Никифоров, А. И. Динамический хаос в трехмерной диссипативной системе с группой симметрии D_2 / А. И. Никифоров, Д. С. Рябов, Г. М. Чечин // Известия вузов. Прикладная нелинейная динамика. – 2004. – Т. 12, № 6. – С. 28–43.
2. Habol, L. Route to hyperchaos in quadratic optomechanics / L. Habol, I. Shamroni // ArXiv: 2410.11329. – P. 1–10.
3. Грицук, Е. В. К теории нелинейных систем дифференциальных уравнений со свойством Пенлеве / Е. В. Грицук, В. И. Громак // Весці НАН Беларусі. Серыя фіз.-мат. навук. – 2010. – № 3. – С. 25–30.
4. Айнс, Э. Л. Обыкновенные дифференциальные уравнения / Э. Л. Айнс // Харьков: ОНТИ, 1939. – 720 с.
5. Камке, Э. Справочник по обыкновенным дифференциальным уравнениям / Э. Камке // Москва: Наука, 1971. – 576 с.

МОДЕЛИРОВАНИЕ СИНХРОНИЗИРОВАННЫХ ГЕНЕРАТОРОВ ЛОРЕНЦА

Шилин Л. Ю., Хаджинова К. А.
Кафедра информационных технологий и управления,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {leonidchilin, kseniyakhadzhynava}@gmail.com

В статье рассмотрено численное исследование синхронизации двух генераторов Лоренца с и без обратной связи. В работе анализируется механизм и условия достижения устойчивой полной синхронизации, подтвержденные анализом фаз и траекторий системы.

Введение

Изучение синхронизации хаотических систем, таких как генераторы Лоренца, представляет особую актуальность в области нелинейной динамики и современных технологий защищенной коммуникации. Такие системы применяются для генерации хаотических сигналов, используемых в криптографических алгоритмах и системах шифрования, где важна устойчивость и надежность передачи информации [1]. Предыдущие исследования показывают, что наличие обратной связи способствует стабилизации синхронных состояний, что подтверждает потенциал использования таких моделей на практике [2].

Цель работы – провести численные эксперименты по моделированию двух взаимодействующих генераторов Лоренца, реализовать критерии оценки их синхронности и количественно определить роль токовой обратной связи с задержкой в процессе перехода систем к устойчивой синхронизации.

I. Математическая модель

Генератор Лоренца описывается системой уравнений:

$$\begin{cases} \dot{x}_1 = \sigma\ (y_1 - x_1) - R(x_1 - x_2), \\ \dot{y}_1 = x_1(\rho\ - z_1) - y_1, \\ \dot{z}_1 = x_1 y_1 - \beta\ z_1, \\ \dot{x}_2 = \sigma\ (y_2 - x_2) - R(x_2 - x_1), \\ \dot{y}_2 = x_2(\rho\ - z_2) - y_2, \\ \dot{z}_2 = x_2 y_2 - \beta z_2. \end{cases}$$

где параметры $\sigma = 10$, $\rho = 28$, $\beta = 8/3$ задают хаотический режим динамики [3].

При моделировании взаимодействия вводится параметр сопротивления R обратной связи с задержкой τ:

$$\begin{cases} \dot{x}_1 = \sigma(y_1 - x_1) - R(x_1 - x_2(t-\tau)), \\ \dot{y}_1 = x_1(\rho - z_1) - y_1, \\ \dot{z}_1 = x_1 y_1 - \beta z_1, \\ \dot{x}_2 = \sigma(y_2 - x_2) - R(x_2 - x_1(t-\tau)), \\ \dot{y}_2 = x_2(\rho - z_2) - y_2, \\ \dot{z}_2 = x_2 y_2 - \beta z_2, \end{cases}$$

Такая модель отражает двунаправленное взаимодействие с учетом временной задержки, влияющей на динамику синхронизации.

II. Методика определения синхронизаций

Для оценки синхронности каждого генератора вычисляются его фазовые траектории как угол:

$$\varphi_i(t) = \arctan 2(y_i(t), z_i(t)), \quad i = 1,2.$$

После чего определяется фазовая разность с учётом модуля 2π:

$$\Delta\varphi(t) = \min\left(|\varphi_1 - \varphi_2| \bmod 2\pi,\ 2\pi - |\varphi_1 - \varphi_2| \bmod 2\pi\right).$$

Синхронизация считается устойчивой, если в скользящем окне длиной 2 секунды среднее евклидово расстояние между траекториями и средняя фазовая разница не превышают порогов 1.0 и 0.05 соответственно [4].

Момент полной синхронизации – первый момент во времени, с которого эти условия стабильно выполняются, указывая на согласование генераторов по фазе и амплитуде.

III. Результаты моделирования

Для решения системы использовался классический метод Рунге-Кутта 4 порядка с шагом интегрирования 0.01 секунды [5]. Моделирование проводилось на интервале до 260 секунд. Этот метод обеспечивает высокую точность расчётов и стабильность при моделировании нелинейных хаотических систем с задержкой.

Для наглядного представления моделирования ниже представлены графики аттракторов и фазовых траекторий генераторов как в режиме обратной связи, так и без нее. На рисунках генератор 1 показан оранжевым цветом, а генератор 2 – синим.

На рис. 1 изображены аттракторы генераторов без обратной связи за 40 секунд. Явно проявляется хаотичное поведение систем с отсутствием схожести их траекторий.

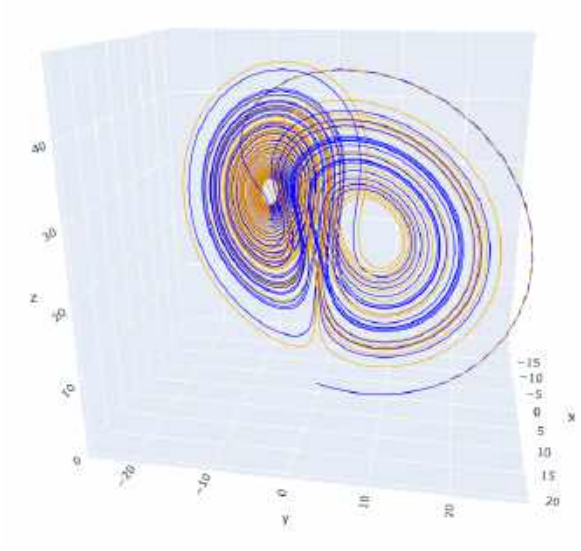


Рис. 1 – Аттракторы без синхронизации

Рис. 2 демонстрирует соответствующие фазовые траектории, где чётко видны значительные расхождения фаз, подтверждающие отсутствие синхронизации в данное время.

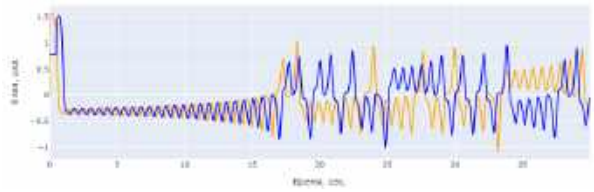


Рис. 2 – Фазовый портрет без синхронизации

Включение обратной связи приводит к принципиально иному поведению систем. На рис. 3 показаны аттракторы за интервал 260 секунд, на которых после 239.9 секунды начинается устойчивое совпадение траекторий, свидетельствующее о полной синхронизации генераторов.

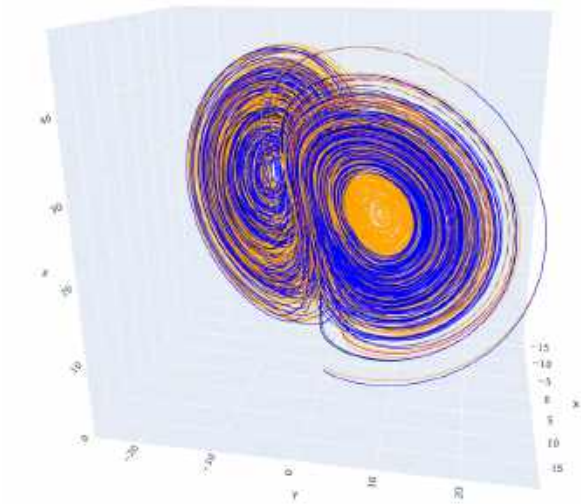


Рис. 3 – Аттракторы с синхронизацией

Рис. 4 иллюстрирует фазовые траектории, указывающие на устойчивое согласование фаз после указанного момента времени.

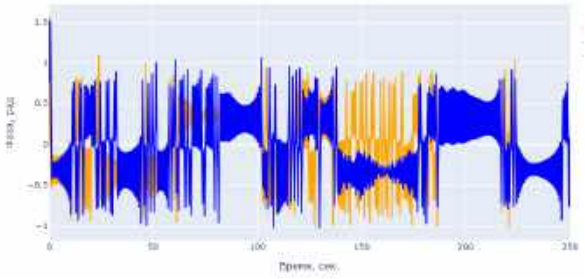


Рис. 4 – Фазовый портрет с синхронизацией

Представленные графики демонстрируют значительную роль обратной связи, обеспечивающей устойчивую синхронизацию генераторов Лоренца и подтверждают возможность практического управления хаотическими системами в целях защищённых коммуникаций [1,5].

Данная модель не учитывает внешние шумы и искажения. Параметр сопротивления R обеспечивает взаимное влияние двух генераторов, способствуя синхронизации.

Заключение

Проведённые исследования подтвердили, что токовая обратная связь с параметрами $R = 1.2$ и задержкой $\tau = 0.5$ обеспечивает устойчивую и полную синхронизацию генераторов Лоренца через примерно 239.(9) секунд. Отсутствие обратной связи приводит к независимой хаотической динамике. Результаты имеют важное значение для применения в системах защищённой связи и криптографии, расширяя возможности использования хаотических систем в управляемых сетях [1,4,5].

1. Шилин, Л. Ю., Навроцкий, А. А. Система шифрования данных на основе устройств фазовой синхронизации // Известия Гомельского государственного университета имени Ф. Скорины. – 2024. – № 3(144). – С. 84–89.
2. Хаджинова, К. А., Петруша, А. Д. Исследование хаотического поведения генератора Лоренца / К. А. Хаджинова, А. Д. Петруша // Сборник БГУИР 2025. – 2025.
3. E. N. Lorenz, "Deterministic Nonperiodic Flow *Journal of the Atmospheric Sciences*, 1963.
4. A. Pikovsky, M. Rosenblum, J. Kurths, *Synchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, 2001.
5. Кузнецов, С. П. Простые электрические генераторы хаоса / С. П. Кузнецов // Система Лоренца. – 2018. – № 1. – С. 46-47.

DECOMPOSITION OF BASIC GRAPHS IN FRACTIONAL LINEAR PROGRAMMING PROBLEMS WITH LINEAR CONSTRAINTS

Pilipchuk L. A.
Department of Web Technologies and Computer Simulation, Belarusian State University
Minsk, Belarus
E-mail: pilipchuk@bsu.by

Efficient computational technologies of decomposing basis graphs in linear-fractional multiflow programming problems with a nested constraint structure are considered. An example of basis graphs decomposition for solving large sparse linear systems for mathematical model of one linear-fractional programming problem are given.

Introduction

Decomposition of the sparse linear underdetermined system with incidence matrix of the graph are considered. Systems of this type appear in non-homogeneous flow programming problems.

We construct the algorithms of decomposition of the sparse underdetermined linear systems by separating the variables according to the sets of arcs of spanning trees, cyclic arcs and no basis arcs. Effective algorithms and decomposition technologies have been developed for solving sparse systems of linear algebraic equations with a multigraph incidence matrix in the linear-fractional multiflow programming problems with a nested constraint structure are considered. An implementation of the sparse systems solutions in the computer algebra system Wolfram Mathematica with using the technologies of root trees are considered. The example of decomposition for basis graphs of the sparse systems of linear algebraic equations is given.

I. Mathematical model of linear-fractional optimization problem with linear constraints

For the multinetwork $G = (I,U)$ we consider the following linear-fractional programming problems with linear constraints

$$f(x) = \frac{p(x)}{q(x)} = \frac{\sum\limits_{(i,j)\in U}\sum\limits_{k\in K(i,j)} p_{ij}^k x_{ij}^k + \beta}{\sum\limits_{(i,j)\in U}\sum\limits_{k\in K(i,j)} q_{ij}^k x_{ij}^k + \gamma} \longrightarrow \max, \tag{1}$$

$$\sum_{j\in I_i^+(U^k)} x_{ij}^k - \sum_{j\in I_i^-(U^k)} x_{ji}^k = a_i^k, i \in I^k, k \in K;$$

$$\sum_{k\in K_0(i,j)} x_{ij}^k \le d_{ij}^0, (i,j) \in U_0; \tag{2}$$

$$\sum_{(i,j)\in U}\sum_{k\in K(i,j)} \lambda_{ij}^{kp} x_{ij}^k = \alpha_p, p = \overline{1,l};$$

$$x_{ij}^k \ge 0, k \in K_0(i,j), (i,j) \in U_0; \tag{3}$$

$$0 \le x_{ij}^k \le d_{ij}^k, \in K_1(i,j), (i,j) \in U, x_{ij}^k \ge 0,$$

$$k \in K(i,j) \setminus K_1(i,j), (i,j) \in U \setminus U_0;$$

$$I_i^+(U^k) = \{j \in I^k : (i,j)^k \in U^k\};$$

$$I_i^-(U^k) = \{j \in I^k : (j,i)^k \in U^k\}. \tag{4}$$

Here K ($|K| < \infty$) is a set of different products (types of flow) transported through the multinetwork G. Without loss of generality, let's put $K = \{1,\dots,|K|\}$.

Let us denote the connected network corresponding to a certain type k of flow with $S^k = (I^k,U^k)$, where I^k is the set of nodes and U^k is the set of arcs which are available for the flow of type k, $k \in K$. Also, we define for each node $i \in I$ the set of types of flows $K(i) = \{k \in K : i \in I^k\}$ and for each multiarc $(i,j) \in U$ the set $K(i,j) = \{k \in K : (i,j)^k \in U^k\}$. We assume that the denominator $q(x)$ of the objective function (1) does not change sign on a set of multiflows X, $x \in X$.

The work is devoted to methods, algorithms and technologies decomposition of basis multigraphs for constructing the optimal solutions of linear-fractional programming problem (1)–(4). We use modern innovative technologies of sparse matrix analysis [2], algorithmic graph theory and theoretical computer science.

II. Example of basis graphs decomposition

For a multinetwork $G = (I,U)$, $I = \{1,2,3,4,5\}$, $U = \{(1,2),(2,3),\ (4,1),(4,2),\ (4,3),(5,1),(5,2),\ (5,4)\}$ consider a sparse underdetermined system of linear algebraic equations (5) – (6). Multinetwork G presented as a combination of networks S^k (see Fig. 1): $S^k = (I^k,U^k)$, $k \in K = \{1,2,3\}$,

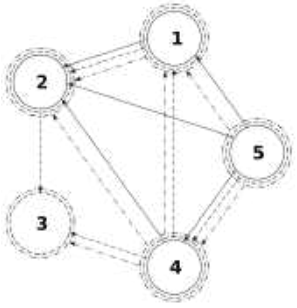


Figure 1 – Multinetwork $G = (I,U)$, Combination of networks S^k, $k \in K = \{1,2,3\}$

$U^1 = \{(1,2)^1,(4,2)^1,(5,1)^1,(5,2)^1,(5,4)^1\}$,
$U^2 = \{(1,2)^2,\ (2,3)^2,\ (4,1)^2,\ (4,3)^2,\ (5,4)^2\}$,

$$U^3 = \{(1,2)^3,\ (4,1)^3,\ (4,2)^3,\ (5,1)^3,\ (5,4)^3\}.$$

$$x^1_{1,2} - x^1_{5,1} = 5, \quad -x^1_{1,2} - x^1_{4,2} - x^1_{5,2} = -13$$
$$x^1_{4,2} - x^1_{5,4} = -1, \quad x^1_{5,1} + x^1_{5,2} + x^1_{5,4} = 9$$

$$x^2_{1,2} - x^2_{4,1} = -1, \quad x^2_{2,3} - x^2_{1,2} = 7$$
$$-x^2_{2,3} - x^2_{4,3} = -13, \quad x^2_{4,1} + x^2_{4,3} - x^2_{5,4} = -1$$
$$x^2_{5,4} = 8$$

$$x^3_{1,2} - x^3_{4,1} - x^3_{5,1} = -4, \quad -x^3_{1,2} - x^3_{4,2} = -10$$
$$-x^3_{4,3} = -6, \quad -x^3_{4,1} + x^3_{4,2} + x^3_{4,3} - x^3_{5,4} = 12$$
$$x^3_{5,1} + x^3_{5,4} = 8 \tag{5}$$

$$x^1_{1,2} + 4x^2_{1,2} + 7x^3_{1,2} + 4x^2_{2,3} + 5x^2_{4,1} +$$
$$+4x^3_{4,1} + 3x^1_{4,2} + 6x^3_{4,2} + 6x^2_{4,3} + 2x^3_{4,3} + 9x^1_{5,1} +$$
$$+3x^3_{5,1} + 10x^1_{5,2} + 4x^1_{5,4} + 2x^2_{5,4} + 9x^3_{5,4} = 328$$

$$10x^1_{1,2} + 6x^2_{1,2} + 5x^3_{1,2} + 2x^2_{4,1} + 4x^3_{4,1} + 9x^1_{4,2} +$$
$$+10x^2_{4,3} + 4x^3_{4,3} + 4x^1_{5,1} + 2x^3_{5,1} + 4x^1_{5,2} + 7x^1_{5,4} +$$
$$+7x^2_{5,4} + 10x^3_{5,4} = 412$$

$$5x^1_{1,2} + 8x^2_{1,2} + x^3_{1,2} + 7x^2_{2,3} + 9x^2_{4,1} + 5x^3_{4,1} +$$
$$+2x^3_{4,2} + 6x^2_{4,3} + 5x^3_{4,3} + 4x^1_{5,1} + +x^3_{5,1} + 7x^1_{5,2} +$$
$$2x^1_{5,4} + 8x^2_{5,4} + 5x^3_{5,4} = 359 \tag{6}$$

Basis graphs(support) $U^k_T \bigcup U^k_C$, $k \in K = \{1,2,3\}$ for the multinetwork $G = (I,U)$ for the system (5) – (6) [1] is represented on figures 2 – 4, where $U^1_T = \{(1,2)^1,\ (4,2)^1,\ (5,4)^1\}$, $U^2_T = \{(1,2)^2,(2,3)^2,(4,3)^2,(5,4)^2\}$, $U^3_T = \{(1,2)^3,(4,2)^3,(4,3)^3,(5,4)^3$ – sets of arcs of spanning trees U^1_T, U^2_T, U^3_T of the graphs $S^1 = (I^1,U^1)$, $S^2 = (I^2,U^2)$, $S^3 = (I^3,U^3)$ respectively (marked with bold lines), $U_C = U^1_C \bigcup U^2_C \bigcup U^3_C$ – set of cyclic arcs, $U^1_C = \{(5,1)^1,(5,2)^1\}$, $U^2_C = \{(4,1)^2\}$, $U^3_C = \emptyset$.

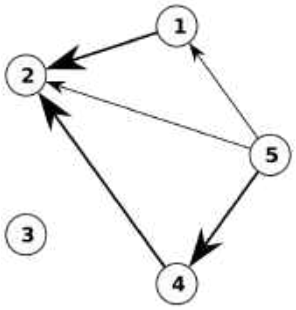


Figure 2 – Basis graph $U^1_T \bigcup U^1_C$ of the system (5)–(6) for $k = 1$. The arcs of basis graph ($k = 1$) for the system (5) are marked in bold

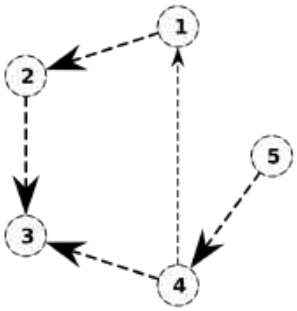


Figure 3 – Basis graph $U^2_T \bigcup U^2_C$ of the system (5)–(6) for $k = 2$. The arcs of basis graph ($k = 2$) for the system (5) are marked in bold

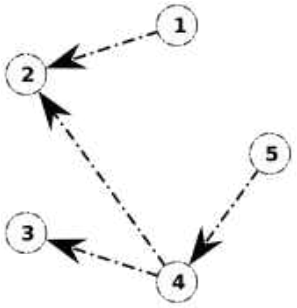


Figure 4 – Basis graph $U^3_T \bigcup U^3_C$ of the system (5)–(6) for $k = 3$. The arcs of basis graph ($k = 3$) for the system (5) are marked in bold

We are build a general solution to a sparse underdetermined system (5) – (6) relative to the basis multigraph (support) (see Fig. 2 – 4) $U^k_T \bigcup U^k_C$, $k \in K = \{1,2,3\}$ of the network $S = (I,U)$ for the system (5) – (6).

Based on the decomposition of basis multiflows we find the flows for the arcs of the set U^k_C, $k \in K = \{1,2,3\}$:

$$x^1_{5,2} \to \frac{1}{34}\left(-449 + 123y^3_{4,1} - 85y^3_{5,1}\right),$$
$$x^1_{5,1} \to \frac{1}{34}\left(3311 - 889y^3_{4,1} + 697y^3_{5,1}\right),$$
$$x^2_{4,1} \to -\frac{504}{17} + \frac{152y^3_{4,1}}{17} - 7y^3_{5,1}.$$

Then we compute the flows for the arcs of the set U^k_T, $k \in K = \{1,2,3\}$.

References

1. Pilipchuk L. A. Linear-fractional extremal inhomogeneous problems network flow programming. – Minsk : BSU, 2013. – 235 p. (in Russian).
2. Pilipchuk L. A. Sparse Linear Systems and Their Applications. – Minsk : BSU, 2013. – 235 p.

Секция «Автоматизированные системы обработки информации»

Section «Automated information processing systems»

СЦЕНАРИИ РЕАГИРОВАНИЯ НА КИБЕРАТАКИ В ОБРАЗОВАТЕЛЬНОМ КИБЕРПОЛИГОНЕ CYBERLAB

Белоусова Е. С., Вербило Н. А., Филиппов А. Н.
Кафедра защиты информации,
Белорусский государственный университет информатики и радиоэлектороники
Национальный детский технопарк
Минск, Республика Беларусь
E-mail: belousova@bsuir.by, nickolay3132@gmail.com, filipov_andrew@mail.ru

В данной статье представлено описание разрабатываемого авторами образовательного киберполигона CyberLab для профориентации учащихся средний учебных заведений и развитие навыков реагирования на кибератаки у студентов в области информационной безопасности. Киберполигон CyberLab является полностью автономным и может быть самостоятельно установлен и настроен пользователем. Для удобства пользователей авторами разрабатываются сценарии, а также реализована онлайн платформа для проверки правильности их прохождения и получения актуальной информации о появлении обновлений.

Введение

Киберполигон – это инфраструктура для отработки практических навыков специалистов, экспертов разного профиля, руководителей в области информационной безопасности и информационных технологий, а также для тестирования программного и аппаратного обеспечения путем моделирования компьютерных атак и отработки реакций на них [1].

На сегодняшний день существует множество киберполигонов различных по масштабу (локальные и национальные), архитектуре (виртуальные и гибридные) и реализации (индивидуальные и командные). Кроме того для прохождения киберполигона используются разные виды сценариев (CTF, Red Team, Blue Team, Purple Team).

Авторами статьи ведется разработка и популяризация образовательного киберполигона CyberLab, предназначенного для изучения учащимися средних и высших учебных заведений распространенных кибератак и способов их блокировки.

I. Назначение CyberLab

CyberLab – образовательный киберполигон с сегментированной сетевой архитектурой, включающей как одноуровневую, так и двухуровневую демилитаризованные зоны (DMZ), что позволяет моделировать различные сценарии кибератак и защиты в условиях, приближенных к реальным. Для образовательного киберполигона CyberLab авторами разрабатываются и внедряют сценарии по моделям Red Team и Blue Team.

Для популяризации и получение актуальной информации о версиях образовательного киберполигона CyberLab был разработан веб-ресурс [2], на котором подробно описаны архитектура, функциональность административного инструментария CyberLab Management Tool, системные требования для развертывания образовательного киберполигона и др. На рис. 1 представлен фрагмент страницы сценариев веб-ресурса CyberLab, предназначенной для отображения и запуска интерактивных обучающих сценариев, связанных с различными типами кибератак в области информационной безопасности.

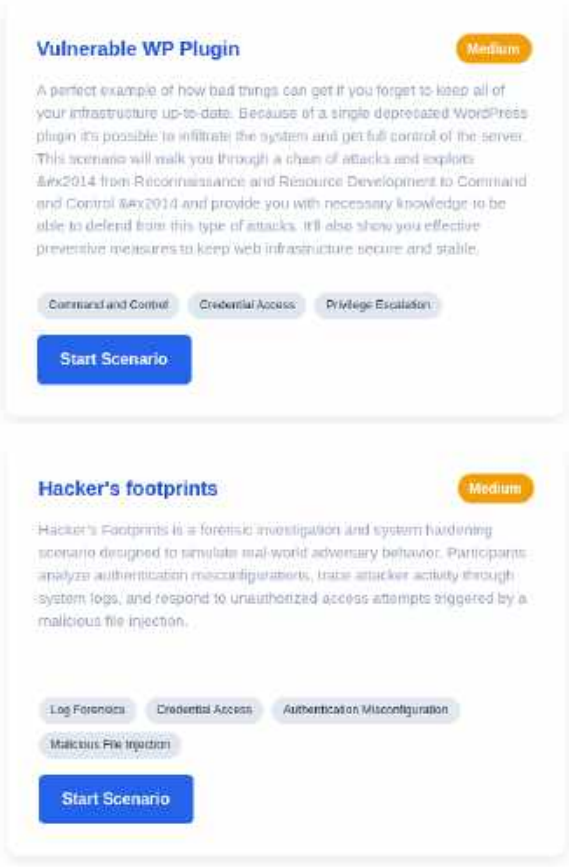


Рис. 1 – Внешний вид веб-ресурса для ознакомления с содержанием сценариев образовательного киберполигона CyberLab

Интерфейс для ознакомления с этапами прохождения сценариев и проверки правильности их выполнения реализован в виде набора фильтров и карточек, каждая из которых представляет отдельный сценарий с названием,

меткой сложности, кратким описанием, списком категорий кибератак. Как видно из рис. 1 на текущий момент реализовано два сценария: Vulnerable WP Plugin, имитирующий действия RedTeam; Hacker's footprints, имитирующий действия BlueTeam.

Для автоматизации локальной установки и управлением образовательным киберполигоном CyberLab была разработана программа CyberLab Management Tool [3], которая предоставляет быстрое взаимодействие со всеми виртуальными машинами в составе образовательного киберполигона.

II. Сценарий Vulnerable WP Plugin

В качестве примера в данной статье приведено описание сценария Vulnerable WP Plugin, основной целью которого является получение доступа к веб-серверу, расположенному в демилитаризованной зоне образовательного киберполигона CyberLab, посредством эксплуатации уязвимости CVE-2015-10144.

Сценарий Vulnerable WP Plugin демонстрирует вектор кибератаки нарушителя, в ходе которого осуществляет сканирование сервера утилитой wpscan с целью получение информации о CMS WordPress и его плагинах, BruteForce атака для получения имени пользователя и пароля к CMS WordPress, получение несанкционированного доступа к серверу посредством инструмента Metasploit для внедрения вредоносного кода, повышение привилегий и настройка удалённого доступа к серверу посредством backdoor. Таким образом, при прохождении сценария учащийся проходит все этапы, которые могут быть реализованы нарушителем. Формирование у учащегося знаний и навыков проведения кибератак позволит развить его понимание способов отслеживания действий нарушителя и их блокировку. Сценарий Vulnerable WP Plugin разделен на следующие этапы:

1. Знакомство с операционной системой Kali Linux;
2. Разведка;
3. Получение данных администратора;
4. Несанкционированный доступ к веб-серверу;
5. Повышение привилегий.

III. Пример прохождения сценария

В процессе прохождения сценария учащийся должен вводить флаги и таким образом проверять правильность выполнения заданий в сценарии. На рисунке 2 представлен пример прохождения одного из этапов сценария Vulnerable WP Plugin и результат нахождения флагов. На рисунке 3 продемонстрирован процесс проверки флага в результате выполнения одного из задания сценария.

Рис. 2 – Пример процесса прохождения задания сценария Vulnerable WP Plugin

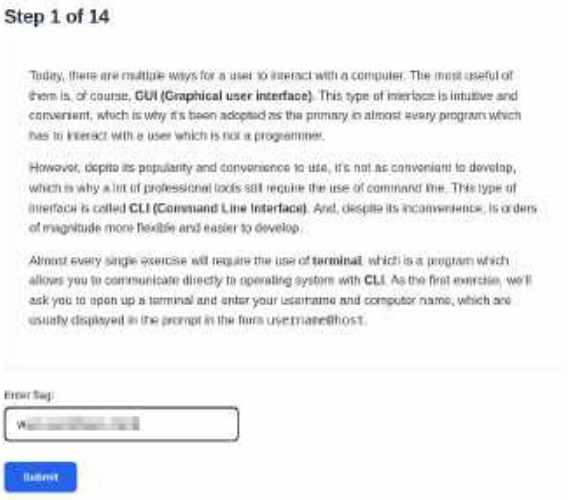


Рис. 3 – Проверка правильности выполнения сценария Vulnerable WP Plugin

На момент написания статьи авторами продолжается работа по разработке и внедрению новых сценариев в образовательный киберполигон CyberLab, который рекомендуется для внедрения в учреждениях среднего и высшего образования для профориентации учащихся в сфере информационной безопасности.

1. Об утверждении методик расчета показателей федеральных проектов национальной программы [Электронный ресурс] / Министерство цифрового развития, связи и массовых коммуникаций Российской Федерации. – Москва, 2025. – Режим доступа: `https://digital.gov.ru/documents/prikaz-minczifry-vhc-143`. – Дата доступа: 26.09.2025.
2. CyberLab Releases [Электронный ресурс] / CyberLab. – Минск, 2025. – Режим доступа: `https://techno-cyber-lab.store`. – Дата доступа: 06.09.2025.
3. Программа для автоматизации образовательного киберполигона CyberLab Management Tool [Электронный ресурс] / Нац. центр интеллект. собств. – Минск, 2025. –Режим доступа: `https://search.ncip.by/depon/index.php?page=3&target=2315`. – Дата доступа: 26.09.2025.

АВТОМАТИЗАЦИЯ ПРИНЯТИЯ РЕШЕНИЙ В СИСТЕМАХ ВИБРАЦИОННОГО КОНТРОЛЯ

Бранцевич П. Ю.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: branc@bsuir.by

Обеспечение безаварийной работы предприятий критической инфраструктуры требует создания системы контроля технического состояния промышленного оборудования. Для сложных механизмов и агрегатов предприятий энергетики одним из важнейших параметров является вибрация. Предлагается модель системы автоматического принятия решения об аварийном отключении турбоагрегата по вибрации.

Введение

Основной задачей современных систем вибрационного контроля и мониторинга является предотвращение аварийного повреждения наблюдаемого технического объекта при внезапном возникновении неисправностей или механических повреждений в его узлах, или при существенном отклонении каких-либо технологических параметров от номинальных. Однако факт возникновения ситуации, требующей останова механизма или агрегата, во многих случаях имеет неоднозначное отображение в параметры вибрации. Стандартизованные критерии защиты отражают наиболее общие взаимосвязи, полученные на основе длительного опыта эксплуатации и исследования механизмов с вращательным движением, и далеко не всегда в полной мере могут удовлетворить эксплуатирующий и управляющий персонал [1].

Системы вибрационного контроля и защиты, построенные на базе компьютерной техники, позволяют реализовать разнообразные и сложные алгоритмы защиты, ориентированные на конкретные типы дефектов и аварийных ситуаций. Это, в свою очередь, позволяет избежать необоснованных («ложная тревога») срабатываний защитного отключения и не допустить «пропуска дефекта» [2, 3].

I. Модель базового элемента системы принятия решений

Состояние наблюдаемого технического объекта описывается параметрами и характеристиками.

Параметр – свойство (показатель) объекта или системы, которое можно измерить. Результатом измерения параметра системы является число или величина, а саму систему можно рассматривать как множество параметров, которые требуется измерить для моделирования или оценки ее поведения. Иногда параметрами называют также величины, очень медленно изменяющиеся по сравнению с другими величинами (переменными). Примерами параметров вибрации являются: среднее квадратическое значение виброускорения (виброскорости), размах колебаний, амплитуда колебаний на определенной частоте, вычисляемые в процессе обработки вибрационного сигнала, формируемого первичными преобразователями (датчиками), установленными на подшипниковой опоре или корпусе механизма и преобразующими механические колебания в электрический сигнал.

Характеристика – это совокупность отличительных свойств чего-либо. Характеристика в технике – есть графическое или табличное выражение зависимости одного параметра от другого, а также функция, выражающая или описывающая эту зависимость. Например, характеристикой объекта является амплитудный спектр вибрационного сигнала, возбуждаемого на корпусе подшипниковой опоры или отрезок временной реализации вибрационного сигнала.

Чтобы выполнять оценку состояния наблюдаемого объекта, требуется некая система принятия решений или поддержки принятия решений. Предлагается следующая модель базового элемента системы принятия решений по оценке состояния наблюдаемого объекта [3,4].

Входными данными базового элемента являются: x_i – значение параметра i, $i = 1 \dots N$; $w_i(y_{j,1}, \dots, y_{j,k})$ – характеристика j при дискретных значениях аргумента y_j, $j = 1 \dots M$; $w_j(y_j(t))$ – характеристика j при непрерывном значении аргумента y_j, $j = 1 \dots M$.

По отношению к входным исходным параметрам и характеристикам применяются функции первичной обработки $f_l(x_i)$, где $l = 1 \dots B$, и $g_m(w_i)$, где $m = 1 \dots C$. Причем разные функции f_l могут применяться по отношению к одному и тому же параметру x_i, а функции g_m к одной и той же характеристике w_i.

Также используются комплексные многопараметрические и многохарактеристические функции $q_n(x_i, \dots, x_j, \dots, x_k, w_l, \dots, w_m, \dots, w_p)$, где $n = 1 \dots D$; $i,j,k = 1 \dots N$; $l,m,p = 1 \dots M$.

По отношению к набору функций $f_l(x_i)$, $g_m(w_i)$, $q_n(x_i, \dots, x_j, \dots, x_k, w_l, \dots, w_m, \dots, w_p)$ применяются обобщающие функции $y_k = Q[f_l(x_i),\ l = 1 \dots B;\ g_m(w_i),\ m = 1 \dots C;\ q_n(x_i, \dots, x_j, \dots, x_k, w_l, \dots, w_m, \dots, w_p),\ n = 1 \dots D;\ k = 1 \dots L$. И уже по отношению к y_k

применяются разнообразные решающие функции $S_t(y_k)$, $t = 1 \dots P$.

Результат функции $S_t(y_k)$ определяет одно из возможных состояний анализируемого объекта, тип самого объекта, принимаемое решение.

II. Применение базового элемента системы принятия решений в автоматике защитного отключения турбоагрегата по параметрам вибрации

Реализован и прошел апробацию на ряде турбоагрегатов алгоритм защитного отключения по вибрации, в котором учитывается несколько факторов.

1. Фактор низкочастотной составляющей вибрации. Под низкочастотной вибрацией (НЧВ) понимается среднее квадратическое значение виброскорости в зоне частоты, равной половине оборотной. Сигнал защитного отключения вырабатывается в том случае, если для любой подшипниковой опоры турбоагрегата возникла следующая ситуация: СКЗ виброскорости НЧВ, измеренное для вертикального направления и для поперечно-горизонтального направления любой подшипниковой опоры, на протяжении 4-6 секунд превышает v мм/с и, при этом, хотя бы для одного из этих направлений, оно на протяжении этого же времени превышает $3v$ мм/с. Уровень v определяется типом и рабочими частотами механизма.

2. Фактор оборотной составляющей вибрации. Под оборотной составляющей вибрации понимается СКЗ виброскорости спектральной составляющей с частотой, равной частоте вращения вала (ротора) агрегата:

2.1. Величина СКЗ оборотной составляющей. Для каждой подшипниковой опоры и каждого из направлений измерения вибрации устанавливается значение СКЗ виброскорости оборотной составляющей, соответствующее аварийному уровню, который выбирается с учетом конструктивных, функциональных и эксплуатационных особенностей контролируемого механизма. Сигнал защитного отключения вырабатывается в том случае, если в четырех или более точках контроля СКЗ виброскорости оборотной составляющей превысило заданный, для соответствующей точки, аварийный уровень;

2.2. Вектор приращения оборотной составляющей. Для каждой подшипниковой опоры и каждого из направлений измерения вибрации устанавливается значение вектора приращения оборотной составляющей, соответствующее аварийному уровню. Сигнал защитного отключения вырабатывается в том случае, если в четырех или более точках измерений вектор приращения оборотной составляющей превысил заданный, для соответствующей точки измерений, аварийный уровень.

3. Фактор высокочастотной составляющей вибрации. Под высокочастотной составляющей вибрации (ВЧВ) понимается СКЗ виброскорости в частотной полосе, нижняя граница которой равна двойной оборотной частоте, а верхняя - верхней границе частотного диапазона, в котором производится вибрационный контроль наблюдаемого механизма. Сигнал защитного отключения вырабатывается в том случае, если для любых двух направлений измерения вибрации для любой подшипниковой опоры высокочастотная вибрация превысила значение аварийного уровня, установленного для данного объекта в течение 3–6 секунд. Сигнал на защитное отключение контролируемого механизма вырабатывается в том случае, если он выработан по одному из указанных критериев, или по нескольким критериям одновременно. Для реализации данной системы принятия решения о защитном отключении в качестве входных данных применяется характеристика: $TI_j(x_0, \dots, x_{N-1})$, $j = 1 \dots M$ – дискретная временная реализация вибрационного сигнала; M – число точек контроля для наблюдаемого турбоагрегата. Для каждой подшипниковой опоры контроль вибрации осуществляется в трех направлениях: вертикальном, поперечно-горизонтальном, осевом.

Функция $g_{1,j,k}[TI_j(),dt_k]$, $j = 1 \dots M$; $k = 1,2,\dots$ предназначена для вычисления СКЗ оборотной составляющей вибрации $A_{CCV,j}$. Функция $g_{2,j,k}[TI_j(),dt_k]$, $j = 1 \dots M$; $k = 1,2,\dots$ предназначена для вычисления фазы оборотной составляющей вибрации $Faz_{CCV,j}$. Функция $g_{3,j,k}[TI_j(),dt_k]$, $j = 1 \dots M$; $k = 1,2,\dots$ предназначена для вычисления СКЗ низкочастотной вибрации. Функция $g_{4,j,k}[TI_j(),dt_k]$, $j = 1 \dots M$; $k = 1,2,\dots$ предназначена для вычисления СКЗ высокочастотной вибрации. Функция $f_{1,j,k}(A_{CCV,j},Faz_{CCV,j},dt_{k-1},dt_k)$, $j = 1 \dots M$; $k = 1,2,\dots$ предназначена для вычисления вектора приращения оборотной составляющей вибрации.

Результатом обобщающих функций первого и второго уровня является решение о срабатывании защитного отключения.

III. Список литературы

1. Правила технической эксплуатации электрических станций и сетей: СТП 33240.20.501-23. – 3-е изд. – Минск : Белэнерго, 2023. – VI, 477 с.
2. Бранцевич, П. Ю. Компьютерные системы и комплексы обработки вибрационных сигналов / П. Ю. Бранцевич. – Минск: Бестпринт, 2023. – 282 с.
3. Бранцевич, П. Ю. Оценка технического состояния механизмов с вращательным движением на основе анализа вибрационных характеристик пусков и выбегов / П. Ю. Бранцевич. – Минск : Четыре четверти, 2021. – 236 с.
4. Brancevich, P. Y. Implementation of Decision-Making Systems Based on a Typical Decisive Element / P. Y. Brancevich // Doklady BGUIR. – 2023. – Vol. 21, – № 5. – P. 96-103.

АНАЛИЗ ТЕХНОЛОГИЙ И МЕТОДОВ ДЛЯ АВТОМАТИЗИРОВАННОГО ВЫЯВЛЕНИЯ ЗАБОЛЕВАНИЙ РАСТЕНИЙ

Быстрова В. А., Боброва Н. Л.
Кафедра информатики, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: veronikabystrova94@gmail.com, bobrova@bsuir.by

В работе проведен анализ современных методов и инструментов для разработки интеллектуальных систем обнаружения заболеваний растений с использованием технологий компьютерного зрения. Рассмотрены особенности выбора алгоритмов, типов камер и вариантов размещения моделей, что позволяет определить оптимальные технологические решения для таких систем.

Введение

ИИ-технологии активно внедряются в сельское хозяйство, формируя основу для перехода отрасли к принципам умного и точного земледелия. Их использование позволяет повысить эффективность мониторинга состояния растений, оптимизировать применение ресурсов и снизить влияние человеческого фактора при принятии агротехнических решений. Сегодня системы на основе искусственного интеллекта применяются на всех уровнях агропроизводства – от конвейерных линий сортировки продукции до автономных роботов и беспилотных дронов, осуществляющих мониторинг состояния посевов.

В данной статье рассматриваются современные подходы к применению технологий искусственного интеллекта в задачах выявления заболеваний растений. Особое внимание уделяется компьютерному зрению – одному из наиболее динамично развивающихся направлений, обеспечивающему автоматическую интерпретацию визуальных данных с изображений растений.

I. Методы компьютерного зрения

Эффективность интеллектуальных систем, основанных на методах компьютерного зрения, во многом определяется корректным выбором методов обработки изображений, стратегий обучения моделей и используемого оборудования. От качества предобработки данных, архитектуры применяемых алгоритмов и параметров съемки напрямую зависят точность, устойчивость и практическая применимость системы в реальных условиях. Разнообразие типов культур, условий освещения и стадий развития болезни требует адаптации моделей и методов к специфике конкретной задачи и среды эксплуатации. Далее будут рассмотрены основные методы компьютерного зрения.

Классификация изображений применяется в системах, где необходимо определить общее состояние растения или тип заболевания на основе одного снимка. С развитием технологий большое распространение получили глубокие сверточные нейронные сети (CNN), такие как VGG16, ResNet50 и InceptionV3. Эти архитектуры способны автоматически извлекать сложные пространственные и текстурные признаки, что делает их особенно эффективными для диагностики болезней листьев, особенно при использовании методов трансферного обучения – дообучения предобученных моделей на специализированных агроданных. Подход классификации находит применение в мобильных приложениях и стационарных диагностических станциях, где изображение анализируется как единое целое, а также во встроенных и роботизированных системах, предназначенных для автоматического выявления и устранения зараженных растений.

Обнаружение объектов используется в системах, где важно не только определить наличие заболевания, но и локализовать его на изображении. В отличие от простой классификации, данный подход обеспечивает не только идентификацию болезни, но и пространственную привязку поражений, что значительно повышает информативность анализа. Методы семейства R-CNN обеспечивают высокую точность локализации, но требуют значительных вычислительных ресурсов. Для задач, где приоритетом является скорость, применяются модели YOLO и SSD, способные анализировать поток изображений в реальном времени. Например, в одном из исследований была разработана система распознавания заболеваний пшеницы в режиме реального времени на основе модели YOLOv8, интегрированной в платформу БПЛА, что позволило не только локализовать пораженные участки, но и передавать координаты для формирования карт-заданий по обработке полей [1].

Сегментация изображений используется на более детализированном уровне анализа, когда требуется определить точные границы пораженных областей и вычислить процент повреждения листовой поверхности. Для этого применяются архитектуры U-Net, SegNet и DeepLabv3+, обеспечивающие пиксельную классификацию изображения. Более сложные модели, такие как Mask R-CNN, позволяют одновременно выделять несколько различных типов поражений и проводить их раздельный анализ. Этот подход особенно поле-

зен в тепличных комплексах и исследовательских лабораториях, где требуется точная количественная оценка степени болезни, например, для отслеживания динамики лечения или устойчивости сорта к патогену.

II. Средства съемки и размещение моделей

Наряду с выбором алгоритма, тип используемой камеры играет ключевую роль в построении интеллектуальной системы, поскольку напрямую влияет на структуру входных данных и, следовательно, на архитектуру модели и выбор подходящего набора данных для ее обучения. RGB-камеры являются наиболее распространенным типом сенсоров в агротехнологических системах. Их популярность обусловлена сочетанием низкой стоимости, технологической простоты и совместимости с существующими методами анализа изображений. Они обеспечивают высокое пространственное разрешение и четкое изображение, что позволяет различать мелкие визуальные признаки заболеваний. Однако, несмотря на это, их спектральная ограниченность (диапазон 380-750 нм) влияет на способность обнаружения заболеваний на ранних стадиях, когда внешние проявления еще отсутствуют.

Мультиспектральные камеры регистрируют отражение растений в нескольких спектральных диапазонах, включая инфракрасный и красный край, и позволяют вычислять вегетационные индексы (например, NDVI), которые отражают уровень хлорофилла и здоровье растений [2]. Такие системы применяются на дронах, обеспечивая высокое покрытие поля, но с меньшей детализацией отдельных листьев. Мультиспектральные данные чаще используются для обнаружения стрессов или аномалий на уровне участка, а не для точной диагностики конкретного заболевания.

Гиперспектральные камеры предоставляют еще более богатую спектральную информацию, снимая сотни узких полос, что позволяет анализировать биохимический состав тканей растения. Например, исследование [3] показало, что при раннем заражении капусты грибками выявляются специфические изменения спектрального профиля (в диапазонах 400–520 нм и 700–722 нм) до появления визуально заметных симптомов. Несмотря на высокую информативность, гиперспектральная визуализация имеет ряд ограничений, сдерживающих ее широкое применение в полевых условиях. Эти системы характеризуются высокой стоимостью оборудования, значительными требованиями к вычислительным ресурсам и чувствительностью к внешним условиям съемки, таким как освещенность, влажность и движение растений.

Тип размещения модели также существенно влияет на выбор архитектуры и методов обработки данных. Во встроенных системах, таких как роботы или дроны, модель должна быть компактной, энергоэффективной и способной работать в реальном времени на ограниченных вычислительных ресурсах, что часто предполагает использование облегченных CNN-моделей или YOLO-подобных детекторов. В качестве примера наземного применения в полевых условиях в работе [4] описан роботизированный комплекс для сортофитопрочистки картофеля, который с помощью встроенной модели компьютерного зрения идентифицирует и удаляет зараженные растения в автоматическом режиме. В стационарных диагностических станциях или облачных платформах требования к ресурсам менее критичны, что позволяет применять более тяжелые и точные модели, обеспечивая высокую точность и детальную визуализацию пораженных областей. При этом, встроенные решения чаще ориентированы на немедленные действия, тогда как стационарные системы – на аналитику и прогнозирование.

Заключение

В данной работе рассмотрены современные подходы к применению методов компьютерного зрения для выявления заболеваний растений, а также влияние выбора сенсоров и архитектуры модели на эффективность интеллектуальных агросистем. Для разрабатываемой системы полевого мониторинга оптимальным решением является использование модели YOLO для обнаружения и локализации пораженных участков, совместно с веб-приложением, обрабатывающим данные с RGB-камер. Такой подход обеспечивает оперативную диагностику, визуализацию и передачу результатов агрономам для принятия своевременных решений.

1. Мифтахов, И. Р. Разработка диагностической платформы на базе БПЛА для определения заболеваний растений на основе глубокого обучения: дис. канд. техн. наук: 35.2.004.05 / И. Р. Мифтахов – Уфа, 2025. – 263 с.
2. Radócz, L. Multispectral UAV-Based Disease Identification Using Vegetation Indices for Maize Hybrids / Radócz, L. [etc.] // Agriculture. – 2024. – Vol. 14, iss. 11. – Mode of access: https://doi.org/10.3390/agriculture14112002. – Date of access: 23.10.2025.
3. Szechyńska-Hebda, M. A. Application of Hyperspectral Imaging for Early Detection of Pathogen-Induced Stress in Cabbage as Case Study / Szechyńska-Hebda, M. A. [etc.] // Agronomy. – 2025. – Vol. 15, iss. 7. – Mode of access: https://doi.org/10.3390/agronomy15071516 – Date of access: 23.10.2025.
4. Тетерин В. С. Обоснование цифровой системы идентификации инфекционных заболеваний и удаления зараженных растений картофеля / Аграрная наука – 2025. –Т. 1, № 6. – С. 148–155.

МЕТОДОЛОГИИ ОБЕСПЕЧЕНИЯ КАТАСТРОФОУСТОЙЧИВОСТИ В ВИРТУАЛИЗИРОВАННЫХ ИНФРАСТРУКТУРАХ: СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПОДХОДОВ

Буцкевич Е. М., Кравченко Е. Д., Белоусова Е. С.
Кафедра защиты информации,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {lbutskevich, ekat25163}@gmail.com, belousova@bsuir.by

Проведён сравнительный анализ современных методологий обеспечения катастрофоустойчивости в виртуализированных IT-инфраструктурах. Особое внимание уделено роли резервного центра обработки данных (ЦОД) и механизмам синхронизации. В работе предложена концептуальная модель автоматизированной системы аварийного восстановления, основанная на политиках, управляемых бизнес-метриками (RTO/RPO). Данная модель обеспечивает адаптивный выбор сценария переключения и минимизирует влияние человеческого фактора в процессе реагирования на инциденты.

Введение

В эпоху цифровизации и активного внедрения технологий виртуализации обеспечение непрерывности бизнес-процессов становится одной из ключевых задач управления IT-инфраструктурами. Нарушение доступности сервисов вследствие техногенных или природных катастроф может привести к значительным финансовым и репутационным потерям. Для минимизации таких рисков организации внедряют стратегии аварийного восстановления (Disaster Recovery Plan, DRP), базирующиеся на создании географически удалённого резервного ЦОД. Эффективность любой стратегии оценивается по двум ключевым бизнес-метрикам: RTO (Recovery Time Objective) – целевое время восстановления, и RPO (Recovery Point Objective) – целевая точка восстановления. Целью настоящей работы является сравнительный анализ основных подходов к обеспечению катастрофоустойчивости, их оценка по критериям RTO/RPO, и обоснование необходимости разработки автоматизированной системы, способной адаптивно выбирать механизм переключения при сбое.

I. Анализ базовых методологий катастрофоустойчивости

Фундаментом катастрофоустойчивой инфраструктуры является резервный ЦОД. Существуют четыре базовых подхода, отличающихся по принципу действия, стоимости и достигаемым показателям RTO и RPO.

Восстановление из резервных копий (Backup and Restore). Данная методология основана на периодическом (как правило, ежесуточном) создании образов виртуальных машин. Принцип действия заключается в развертывании инфраструктуры «с нуля» из последней копии в резервном ЦОД. Из-за периодического характера и необходимости ручного вмешательства данный подход характеризуется высокими RTO и RPO, измеряемыми в часах или днях.

Репликация виртуальных машин. Технология обеспечивает непрерывное копирование данных ВМ на резервную площадку в асинхронном режиме. Принцип действия состоит в том, что на резервной площадке поддерживаются актуальные копии ВМ, готовые к запуску. Это позволяет достичь низкого RPO (минуты), однако сам процесс переключения (failover) часто требует ручного или автоматизированного запуска.

Географически распределённый кластер (Metro Cluster). Этот подход объединяет два ЦОД в единую отказоустойчивую систему с синхронной репликацией данных. Принцип действия заключается в том, что система хранения данных обрабатывает обе площадки как единое целое. Это позволяет автоматически переключать нагрузку при сбое, обеспечивая RPO, близкий к нулю, и RTO в пределах минут.

Бесшовная отказоустойчивость (Fault Tolerance, FT). Технология создает полную, синхронно работающую копию ВМ. Принцип действия основан на технологии lockstep, когда каждая операция выполняется одновременно на основной и теневой ВМ. Это обеспечивает мгновенное переключение без простоя и потерь данных (RTO=0, RPO=0), являясь самым дорогостоящим решением.

Сводная сравнительная характеристика рассмотренных подходов приведена в таблице 1.

II. Архитектура системы автоматизации аварийного восстановления на основе политик

Анализ существующих подходов показывает, что их ключевым недостатком является зависимость от человеческого фактора на этапе принятия решения. Для устранения этого недостатка предлагается концептуальная модель автоматизированной системы, реализующей адаптивный,

управляемый политиками выбор сценария восстановления.

Ядром предложенной концептуальной модели является модуль принятия решений (Policy Engine). Работа системы, как показано на рис. 1, инициируется событием от модуля мониторинга, которое передается в модуль анализа для классификации инцидента. Получив классифицированный инцидент (например, «отказ хоста» (физического сервера, на котором работают виртуальные машины) или «полный отказ ЦОД»), модуль принятия решений обращается к внешней матрице политик (Базе знаний). В данной матрице заранее определены правила, сопоставляющие тип инцидента и уровень критичности сервиса (определяемый его RTO/RPO) с конкретным сценарием восстановления (Recovery Workflow). На основе этого сопоставления система делает адаптивный выбор: для некритичного сервиса при отказе одного хоста может быть выбран сценарий перезапуска средствами HA, тогда как для критичного сервиса при полном отказе ЦОД будет инициирована активация Metro Cluster. Выбранный сценарий передается на исполнение, где запускаются соответствующие технические процедуры и отправляются уведомления администраторам.

Рис. 1 – Архитектура системы адаптивного восстановления

Такая архитектура позволяет системе не просто реагировать на сбой, а адаптивно выбирать наиболее релевантный и экономически оправданный способ восстановления, минимизируя как время простоя, так и избыточные затраты.

III. Выводы

Катастрофоустойчивость является критически важным аспектом функционирования современных виртуализированных IT-инфраструктур. Проведенный сравнительный анализ показал, что, хотя выбор технологии аварийного восстановления и должен определяться метриками RTO/RPO, эффективность существующих решений ограничена зависимостью от человеческого фактора.

Для решения этой проблемы предложена концептуальная модель автоматизированной системы, управляемая политиками. Она устраняет задержки и ошибки, связанные с ручным вмешательством, путем автоматического выбора оптимального сценария восстановления на основе критичности сервиса. Такой подход трансформирует процесс реагирования на инциденты, позволяя перейти от реактивного к проактивному управлению. Это повышает адаптивность системы, сокращает время реакции при переключении между ЦОДами и минимизирует риски для бизнеса.

IV. Список литературы

1. Катастрофоустойчивость корпоративного датацентра как услуга [Электронный ресурс] / Хабр – Режим доступа: `https://habr.com/ru/companies/safedata/articles/273947/`. – Дата доступа: 24.12.2015.
2. Как сделать ИТ-инфраструктуру VMware катастрофоустойчивой? [Электронный ресурс] / OBLAKO – Режим доступа: `https://oblako.kz/iaas-blog/kak-sdelat-it-infrastrukturu-vmware-katastrofoustojchivoj`. – Дата доступа: 19.03.2018.
3. Microsoft StorSimple – автоматическое аварийное восстановление [Электронный ресурс] / Хабр – Режим доступа: `https://habr.com/ru/companies/comparex/articles/312064/`. – Дата доступа: 07.10.2016.

Таблица 1 – Сравнительная характеристика подходов

Критерий	Резервное копирование	Репликация	Metro Cluster	Fault Tolerance
RTO	от 4 до 48 часов	от 15 мин до 1 часа	от 1 до 5 минут	Ноль (мгновенно)
RPO	от 12 до 24 часов (зависит от частоты копирования)	от 15 сек до 5 минут (зависит от канала связи)	Ноль (синхронная репликация)	Ноль (без потерь данных)
Сложность	Низкая	Средняя	Высокая	Высокая

СОВЕРШЕНСТВОВАНИЕ АЛГОРИТМИЧЕСКОЙ И АППАРАТНОЙ СОСТАВЛЯЮЩИХ КОМПЬЮТЕРНОГО АНАЛИЗА ФРАКЦИОННОГО СОСТАВА МЕТОДОМ МИНИ-ДИСТИЛЛЯЦИИ

Дубровский А. В.
Кафедра вычислительных систем и сетей,
Полоцкий государственный университет имени Евфросинии Полоцкой
Новополоцк, Республика Беларусь
E-mail: a-dubrovskij@yandex.ru

В работе представлен обзор трех поколений приборов мини-дистилляции. Рассмотрены особенности алгоритмов обработки результатов измерений с применением математического моделирования постепенной перегонки. Показана возможность использования приборов мини-дистилляции для анализа фракционного состава сложных смесевых жидкостей и жидкостей с узким диапазоном температур кипения. Предложены пути развития метода мини-дистилляции.

Введение

Развитие методов анализа фракционного состава идет в сторону автоматизации и минимизации затрат времени, при этом основа анализа остается той же – постепенная перегонка. Современные методы основаны на автоматизированных приборах, которые следуют международным стандартам (например, ASTM D86, ГОСТ Р ЕН ИСО 3405), повышая точность, воспроизводимость и скорость анализа по сравнению с ручными методами. Параллельно совершенствованию аппаратной части абсолютно естественным является развитие алгоритмов обработки результатов измерений, в т. ч. с использованием методов математического моделирования.

I. Приборы мини-дистилляции

Следующей ступенью развития методов анализа фракционного состава является мини-дистилляция, позволяющая в 5...10 раз сократить время анализа. При этом сочетание мини-дистилляции с алгоритмами математического моделирования и компьютерной обработки результатов измерений позволяет решать различные аналитические и технологические задачи. В основе этого подхода лежит математическая модель постепенной перегонки в дифференциальной форме: $dv/df = kv(1-v)/f$, в интегральной форме: $v = af^k/(1+af^k)$, где v – доля отгона; f=(T-Тнк)/(Т-Ткк) – безразмерная температура перегонки; T, $Тнк$, $Ткк$ – температура соответственно текущая, начала и конца кипения; k – коэффициент интенсивности перегонки при T=$Tср$; a – коэффициент симметричности перегонки.

Работы по изучению закономерностей постепенной перегонки и детерминированных математических описаний этого процесса привели к созданию приборов трех поколений.

Поколение 1 – «Полоцк-1»: масса 7 кг, время анализа до 10 мин, объём пробы 10 мл, изготовлен в единственном экземпляре.

Поколение 2 – «PMD»: масса 14 кг, время анализа до 10 мин, объём пробы 10 мл, серийно выпускался компанией ISL (Франция) начиная с 2002 года.

Поколение 3 – «ТЭП-Полоцк»: Масса 8 кг, время анализа до 10 мин, объём пробы 10мл, выпущена опытно-промышленная партия в 3 экземплярах, проведена метрологическая аттестация в ВНИИМ им. Д. И. Менделеева (г.Санкт-Петербург).

Рис. 1 – Прибор «ТЭП-Полоцк»

II. Анализ выявленных проблем метода мини-дистилляции

Первое по значимости место в анализе фракционного состава по методу мини-дистилляции занимает определение начала $Тнк$ и конца $Ткк$ кипения. В приборе «ТЭП-Полоцк» для измерения температуры используется только один датчик – термопара типа ХК (хромель-копель) диаметром не более 2мм. При этом алгоритм фиксации моментов $Тнк$ и $Ткк$ основан на поиске экстремумов второй производной температуры Т", что, как видно из рис.2, является в определённых случаях недостаточным для надёжной идентификации этих точек. Второе по значимости место в определении фракционного состава занимает измерение избыточного давления (Р) в колбе в процессе перегонки, дальнейшая обработка которого по определённым алгоритмам позволяет рассчитать расход паровой фазы продукта, произвести пересчёт паровой фазы в жидкую и далее найти

температуры, соответствующие определённой доле отгона. При этом отсутствие математической обработки кривой избыточного давления (P) снижает возможности прибора «ТЭП-Полоцк» в части точного определения *Тнк* и *Ткк*, что так же проиллюстрировано на рис.2

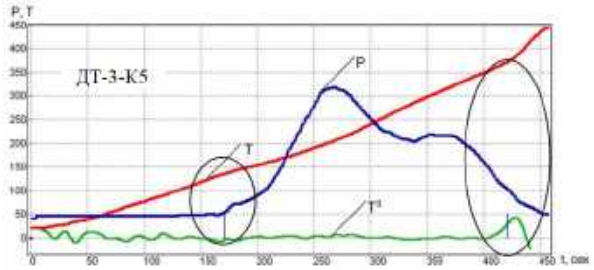


Рис. 2 – Диаграммы постепенной перегонки методом мини-дистилляции (– избыточное давление в колбе, Па; – температура жидкой фазы, °С; “ – вторая производная T)

Кроме того, в ходе проведения экспериментов выявлено, что постепенная перегонка смесевых жидкостей, например бензинов с добавлением октаноповышающих соединений (спиртов, эфиров и др.), модели постепенной перегонки не подчиняется.

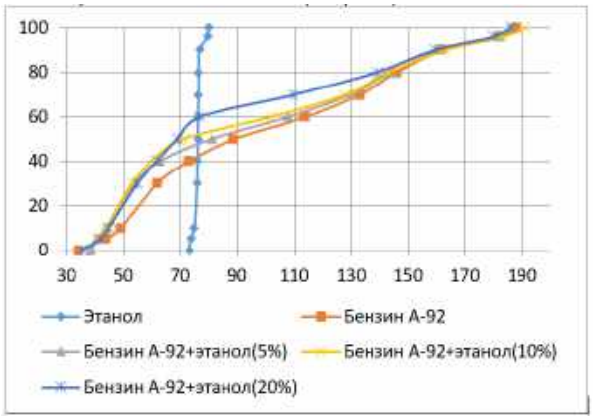


Рис. 3 – Кривые фракционного состава для смесей бензина и этанола

III. Направления развития метода мини-дистилляции

Основываясь на анализе выявленных проблем метода мини-дистилляции можно предложить следующие направления развития этого метода как в аппаратной, так и в алгоритмической части.

1) Измерение температуры перегонки в приборе «ТЭП-Полоцк» осуществляется только в одной точке – возле дна перегонной колбы и этого явно недостаточно для надёжной фиксации моментов *Тнк* и *Ткк*. Поэтому представляется обоснованным использование ещё двух дополнительных точек измерения температуры: вторая точка в кипящем слое на расстоянии около 15мм от дна колбы, третья точка на выходе паровой фазы из колбы. Такое решение потребует некоторого усложнения измерительной части прибора, но при этом обеспечит как надёжное определение температур *Тнк* и *Ткк*, так и упростит структуру алгоритма пересчёта температуры жидкой фазы в температуру паровой фазы.

2) Введение дополнительных алгоритмов обработки кривой избыточного давления в колбе. В первую очередь такое решение позволит гибко и адаптивно управлять процессом перегонки любых жидкостей без необходимости предварительной настройки нагревательной системы прибора и регулировки температуры нагрева в процессе проведения анализа.

3) Причина частичной неработоспособности модели постепенной перегонки заключается не в существовании различных законов постепенной перегонки, а в том, что постоянное взаимодействие испаряющейся (v) и оставшейся ($1\breve{}v$) к определенному моменту времени жидкости не сохраняется до конца перегонки для смесевых и разделенных по физическим характеристикам жидкостей. Поэтому предлагается модифицировать модель постепенной перегонки путём введения в её структуру точки ТМ (точки немонотонности) с координатами *[v*; T*]* и далее подходить к моделированию перегонки с использованием разных коэффициентов модели a и k для участков *Тнк* – и – *Ткк*. Такой подход так же позволит применять моделирование в отношении жидкостей, у которых *Тнк=Ткк=const*.

IV. Заключение

Прибор анализа фракционного состава методом мини-дистилляции «ТЭП-Полоцк», по существу, представляет из себя компьютерную систему. Поэтому дальнейшее развитие как в направлении совершенствования аппаратной части, так и работы по модификации математического описания постепенной перегонки и алгоритмов обработки результатов измерений позволят перейти к созданию приборов четвёртого поколения, а так же получить точные и надёжные результаты при анализе фракционного состава даже в самых сложных и трудоёмких случаях.

V. Список литературы

1. Прашкович, Я. В., Троц, В. А. Определение фракционного состава углеводородов с кислотосодержащими высокооктановыми компонентами на приборе PMD-100 // Электронный сборник трудов молодых специалистов Полоцкого государственного университета. Выпуск 40. Промышленность. Химичекие технологии. – 2021. – с. 131–134.
2. Абаев, Г. Н. Математическое описание фракционной разгонки нефтепродуктов и современные методы её определения // Вестник Полоцкого государственного университета. Серия В. Промышленность. Прикладные науки. – 2013. – № 3. – с. 145–149.
3. Абаев, Г. Н., Дубровский А. В., Кушнир, Е. В. Развитие методов анализа фракционного состава нефтепродуктов // Промышленный сервис. – 2013. – № 3. – с. 15–21.

ИСПОЛЬЗОВАНИЕ BIG DATA ДЛЯ ОПТИМИЗАЦИИ ИТ-РЕШЕНИЙ: КОМПЛЕКСНЫЙ АНАЛИЗ ТЕХНОЛОГИЙ И МЕТОДОВ

Жуков А. М., Навроцкий А. А.
Кафедра Информационных Технологий Автоматизированных Систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: alxzhukov20001@gmail.com , navrotsky@bsuir.by

В статье представлен всесторонний анализ современных методов обработки больших данных с акцентом на оптимизацию производительности ИТ-решений. Рассмотрены теоретические основы распределенных вычислений, проведено сравнительное исследование инструментов (Apache Spark, ClickHouse, Pandas, Dask) и систем хранения (колоночные vs. строковые СУБД). Разработана оригинальная методика тестирования производительности для различных объемов данных (1 Гб – 1 Тб). На практических примерах доказано, что комбинация колоночных СУБД и распределенных вычислений сокращает время обработки в 8-15 раз по сравнению с традиционными подходами. Предложены конкретные рекомендации по выбору технологий для различных сценариев использования.

Введение

Современные ИТ-системы сталкиваются с беспрецедентным ростом объемов данных. Согласно исследованиям IDC (2022), мировой объем данных достигнет 175 Зб к 2025 году[1]. При этом:

- 60% организаций испытывают сложности с обработкой данных в реальном времени;
- 70% вычислительных ресурсов используются неэффективно;
- Традиционные реляционные СУБД демонстрируют падение производительности на 90% при объемах >1 Тб [3].

Основная проблема: отсутствие системного подхода к выбору технологий BIG DATA приводит к:

- Перерасходу бюджета на 40-70%;
- Увеличению времени обработки в 5-10 раз;
- Сложностям масштабирования [5].

I. Архитектура распределенных систем

Современные системы обработки больших данных основаны на трех фундаментальных принципах [1]:

1. Горизонтальная масштабируемость;
2. Оптимальное размещение вычислений.
   2.1. Автоматическое восстановление при сбоях;
   2.2. Линейная масштабируемость;
3. Локализация данных.
   3.1. Минимизация сетевого трафика;
   3.2. Оптимальное размещение вычислений;

Теорема CAP (Brewer,2002): Любая распределенная система может гарантировать только два из трех свойств[1]:

1. Consistency (Согласованность)
2. Availability (Доступность)
3. Partition tolerance (Устойчивость к разделению)

II. Эволюция систем хранения

Сравнительный анализ показывает радикальные различия между подходами[2,8]:

Таблица 1 – Сравнительный анализ стороковых и колоночных СУБД

Критерий	Строковые СУБД (PostgreSQL)	Колоночные СУБД (Clickhouse)
Чтение	Полный скан таблицы (O(n))	Выборка только нужных колонок (O(1))
Сжатие	2-5x	10-100x
Запись	Быстрая (10K записей/сек)	Медленная (1K записей/сек)
Оптимальный случай	OLTP - транзакции	OLAP - аналитика)

Математическая модель производительности колоночных СУБД:

$$T_{query} \propto \frac{1}{\text{compression_ratio} \times \text{selectivity}}$$

где selectivity = (размер результата)/(размер таблицы)[2].

III. Методика Исследования. Тестовая среда

Конфигурация соответствует промышленным стандартам (Apache Spark Tuning Guide, 2023)[7]:

Таблица 2 – Характеристики тестовой среды

Компонент	Характеристика
Кластер	5 узлов (AWS EC2 r5.4xlarge)
CPU	16 vCPU на узел
RAM	128 Гб на узел
Хранилище	EBS gp3 (10K IOPS)
Сеть	10 Gbps

Генерация тестовых данных была произведена путем реализованного алгоритма на Python[6].Данные имеют различную кардинальность и сохранены в отдельных форматах.

IV. Результаты и обсуждение. Сравнительный анализ инструментов

Результаты тестирования (среднее по 10 запускам, 95% доверительный интервал):

Таблица 3 – Характеристики тестовой среды

Объем	Pandas	PySpark	Dask	Clickhouse
1Гб	12.3 ± 0.5	8.1 ± 0.2	10.2 ± 0.3	1.9 ± 0.1
10Гб	–	15.7 ± 0.4	18.9 ± 0.6	4.8 ± 0.2
100Гб	–	42.3 ± 1.1	55.7 ± 1.8	12.4 ± 0.5

Статистическая значимость подтверждена критерием Стьюдента ($p < 0.01$) Визуализация результатов:

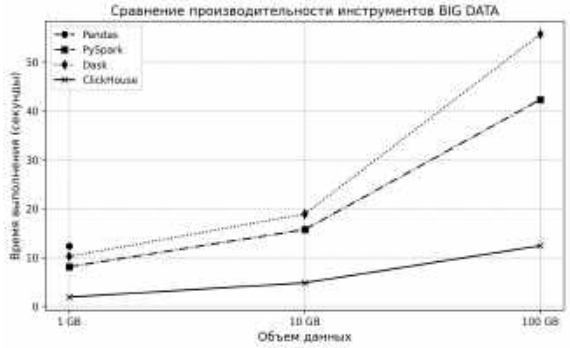


Рис. 1 – Сравнение производительности инструментов BIG DATA

V. Практические кейсы оптимизации. Оптимизация DWH в ритейле.

Исходные данные:

1. Объём обрабатываемых данных: 50 ТБ ежедневно;
2. Время генерации отчётов: 6 часов.

Реализованные оптимизации:

1. Партиционирование данных по дням;
2. Использование материализованных представлений[6];
3. Переход на колоночный формат хранения **Parquet**[2,8].

Результат: время генерации отчётов сокращено до **45 минут** (ускорение в 8 раз).

Также была произведена оптимизация ETL-пайплайнов[5,7].

Ключевые улучшения:

1. Параллельное выполнение задач;
2. Оптимальное управление памятью[4];
3. Инкрементальная загрузка данных[5].

Заключение

Проведённое исследование продемонстрировало ключевые преимущества использования современных технологий **BIG DATA** для оптимизации ИТ-решений. Результаты показали, что:

- **Колоночные СУБД (ClickHouse)** обеспечивают до **15-кратного ускорения** обработки аналитических запросов по сравнению с традиционными строковыми СУБД благодаря[2,8]:
  1. эффективному сжатию данных (10–100×);
  2. оптимизированным алгоритмам сканирования;
  3. поддержке векторных операций.
- **Распределённые вычисления (Apache Spark)** позволяют[4,7]:
  1. обрабатывать терабайты данных за минуты;
  2. линейно масштабировать производительность;
  3. обеспечивать отказоустойчивость.
- **Комплексная оптимизация ETL-процессов** достигается за счёт[5,6,7]:
  1. партиционирования данных;
  2. материализованных представлений;
  3. использования оптимальных форматов хранения (Parquet).

Практические кейсы подтвердили, что предложенные подходы позволяют сократить время обработки данных **с 6 часов до 45 минут** при одновременном уменьшении требуемых вычислительных ресурсов.

Результаты работы имеют практическую ценность для организаций, сталкивающихся с задачами обработки больших объёмов данных и требующих значительного повышения эффективности своих ИТ-инфраструктур.

1. Dean, J., Ghemawat, S. MapReduce: Упрощённая обработка данных на больших кластерах // OSDI'08. – 2008.
2. Abadi, D. и др. Проектирование и реализация современных колонно-ориентированных баз данных // Foundations and Trends in Databases. – 2013.
3. Stonebraker, M. SQL базы данных против NoSQL баз данных // Communications of the ACM. – 2010.
4. Zaharia, M. и др. Apache Spark: Унифицированный движок для обработки больших данных // Communications of the ACM. – 2016.
5. Kimball, R., Ross, M. Инструментариум хранилищ данных. – 3-е изд. – Wiley, 2013. – 600 с.
6. VanderPlas, J. Руководство по анализу данных с Python. – O'Reilly, 2018. – 548 с.
7. Apache Spark Tuning Guide [Электронный ресурс]. – 2023. – Режим доступа: `https://spark.apache.org/docs/latest/tuning.html`.
8. ClickHouse Documentation [Электронный ресурс]. – 2023. – Режим доступа: `https://clickhouse.com/docs`.

АВТОМАТИЗАЦИЯ СОЗДАНИЯ КУРСОВ НА ОСНОВЕ УЧЕБНЫХ ПЛАНОВ В СИСТЕМЕ MOODLE

Зайцев А. С., Богуш Р. П.
Полоцкий государственный университет имени Евфросинии Полоцкой
Новополоцк, Республика Беларусь
E-mail: a.zaytsev@psu.by, r.bogush@psu.by

Для автоматизированного развёртывания учебных курсов на платформе Moodle с использованием формализованных учебных планов разработан программный модуль, плагин Autocourses. Данный компонент включает парсер табличных планов, нормализацию и валидацию данных, оркестратор создания иерархии категорий (факультет, специальность, группа), курсов и дисциплин, систему идемпотентных операций и централизованное логирование. Описаны политики формирования fullname/shortname, семестровая логика, стратегия миграции категорий и требования к безопасности. Оценивается влияние на сокращение человекочасов при массовом развёртывании курсов.

Введение

Современная система высшего образования претерпевает глубокую трансформацию, обусловленную глобализацией, развитием ИТ-инфраструктуры и ограничениями, связанными с пандемией. Рост дистанционных и смешанных форматов обучения сделал платформы, такие как Moodle, центральным узлом, связывающим образовательные, административные и вспомогательные сервисы университета [2]. Например, интеграция с системами учета студентов или расписания стала неотъемлемой частью учебного процесса, особенно в условиях перехода на онлайн-обучение.

Преподаватели сталкиваются с необходимостью ручной настройки курсов: структурирование разделов, загрузка контента, настройка тестов, форумов и прав доступа для ассистентов и студентов. При большом количестве дисциплин или частых обновлениях учебных планов такие задачи могут занимать до нескольких дней в семестр. Автоматизация с помощью плагина Autocourses снимает это бремя, обеспечивая единообразие структуры, быстрый отклик на изменения и минимизацию ошибок. Плагин генерирует полностью настроенные курсы – от метаданных до назначения ролей. Основная цель исследования — разработка архитектуры и алгоритма, которые сократят время запуска дисциплин и улучшат взаимодействие с LMS, адаптируясь к растущим потребностям образовательных учреждений.

I. Краткий анализ Moodle

Выбор Moodle для автоматизации курсов обусловлен открытостью кода, модульной архитектурой [3], соответствием стандартам (SCORM, IMS LTI, xAPI) [4] и готовностью инфраструктуры. Открытость позволяет внедрять сложные плагины без изменения ядра, что важно для интеграции внешних сервисов или переработки интерфейса. Например, настройка кастомных модулей тестирования стала возможной благодаря этой гибкости.

Поддержка стандартов обеспечивает переносимость курсов между платформами и совместимость с внешними материалами, что ценно при закупке контента [4]. Многоязычный интерфейс и локализации адаптируют систему под студентов с разными культурными и языковыми особенностями, что актуально для международных программ. Активное сообщество и регулярные обновления безопасности [5] гарантируют поддержку, а отсутствие лицензий делает Moodle экономичным – расходы ограничиваются серверным обслуживанием и бэкапом.

В Полоцком государственном университете уже действует кластер Moodle с LDAP-аутентификацией и автоматическим резервным копированием, что ускорило разработку Autocourses. Отсутствие аналогов – системы либо коммерческие (например, Open University) [1,6], либо неполные (скрипты на GitHub) – подчеркивает уникальность подхода, где Autocourses охватывает весь цикл автоматизации, включая назначение ролей и контроль качества.

II. Архитектура плагина

Архитектура Autocourses направлена на автоматизацию создания курсов в Moodle с единообразием структуры, надёжностью (атомарные транзакции, логирование) и масштабируемостью (очереди задач, REST-запросы) [3]. Интеграция с SIS и LDAP через гибкий ETL-механизм поддерживает разные форматы данных (PDF, Excel), а авторизованные токены и валидация обеспечивают защиту от ошибок и атак. Далее по списку перечислены основные элементы в архитектуре плагина(в графическом виде представлено на рис. 1):

1. Local plugin: UI, настройки, управление правами.
2. Парсер: PDF/Excel - унифицированная JSON-схема.
3. ETL: нормализация, мэппинг (specialty - faculty), валидация.
4. Очередь задач: Moodle Task API.

5. Оркестратор: идемпотентные операции create/update.
6. Интеграция: internal API или REST с retry/backoff.
7. Логи: таблица логов, UI для фильтрации.

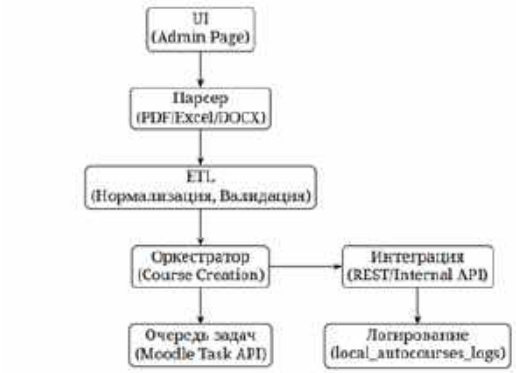


Рис. 1 – Архитектура плагина Autocourses

UI реализован как админ-страница в Moodle для удобного управления. Парсер преобразует данные в JSON, ETL нормализует и валидирует их, оркестратор управляет созданием с проверкой существующих категорий. Очередь задач обеспечивает фоновую обработку, модуль интеграции – надежную связь с Moodle через REST с механизмом повтора, а логирование упрощает аудит и диагностику ошибок.

III. Реализация прототипа

Прототип Autocourses разработан как local/autocourses для Moodle 4.0.4 (PHP 7.4+, mbstring, xml, zip, Composer опционально). Структура включает файлы (version.php, settings.php), классы (helper для категорий, planprovider для парсинга, coursegenerator для курсов, task для фоновых задач) и таблицы (specs, tasks, logs).

Парсеры используют Tabula-style для PDF с нормализацией колонок регулярными выражениями. ETL добавляет yearprefix, проверяет схемы и отправляет ошибки в очередь для ручной проверки, что повышает точность обработки данных. Более подробно процесс ETL показан ниже(см. рис. 2).

Идемпотентность реализуется через поиск по idnumber/name с созданием при отсутствии. Фоновая обработка использует pending-задачи с exponential backoff для устойчивости. Интеграция поддерживает internal и REST-режимы с обработкой HTTP 429. UI включает загрузку файлов, mapping и dry run для тестирования, CLI – утилиты для миграции, а логи – двухуровневые (event log, detailed logs) для детального анализа.

Заключение

Плагин Autocourses автоматизирует создание курсов в Moodle на основе учебных планов через парсинг, ETL и интеграцию. Вместо ручного ввода достаточно загрузить документ с данными (например, таблицу с дисциплинами и часами), и система создаёт структуру курсов, минимизируя ошибки и сокращая время с нескольких часов до минут. Такой подход стандартизирует курсы, упрощает их адаптацию и открывает перспективы для интеграции с новыми сервисами, такими как системы аналитики успеваемости.

1. Смирнова С. Б. Современные LMS: анализ функционала и применение в вузах // Информационные технологии и образование. –2021. – № 4. – С. 12-20.
2. Wright, S. Digital Transformation in Higher Education: Emerging Models and Practices // Educational Technology Society. – 2022. – Vol. 25, No. 1.
3. Moodle. Moodle Documentation: Plugin Development [Electronic resource] : Moodle. – 2023. – Mode of access: `https://docs.moodle.org/dev/`
4. Learning Tools Interoperability (LTI) [Electronic resource] : IMS Global, – 2024. – Mode of access: `https://www.imsglobal.org/activity/learning-tools-interoperability`.
5. Dougiamas, M. The Future of Moodle: Community-Driven Development [Electronic resource] : MoodleMoot Global 2022. – 2022. – Mode of access: `URL:https://dougiamas.com/edmedia2003/`.

Рис. 2 – Процесс ETL

ЗАЩИТА ТРАФИКА РАСПРЕДЕЛЕННЫХ СИСТЕМ НА ОСНОВЕ ПРОТОКОЛА MUTUAL TLS

Крагель В. А., Ярмош А. Д., Скиба И. Г.
Отдел информационных технологий, Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {v.kragel, a.iarmosh, i.skiba}@bsuir.by

В работе рассмотрен протокол Mutual TLS (mTLS) как средство обеспечения аутентификации и конфиденциальности трафика в распределенных системах. Проведен анализ принципов работы mTLS, его преимущества перед односторонним TLS, приведены ключевые аспекты практического внедрения в контексте парадигмы Zero Trust.

Введение

Современные информационные системы эволюционировали в сторону распределенных архитектур, таких как микросервисы и облачные сервисы. В таких системах взаимодействие между множеством независимых компонентов составляет основу их функционирования. Однако эта же особенность создает проблемы безопасности. Традиционный периметровый подход, основанный на защите границ сети, становится неэффективным, так как злоумышленник, получив доступ к одному компоненту, может беспрепятственно перемещаться по внутренней сети системы.

В связи с этим на первый план выходит парадигма Zero Trust («Никому не доверяй, всегда проверяй»), которая предполагает отсутствие неявного доверия к любым сетевым субъектам. Для ее реализации необходимы механизмы, обеспечивающие строгую аутентификацию и шифрование для каждого сетевого взаимодействия. Стандартный протокол TLS решает задачу конфиденциальности и аутентификации сервера для клиента, но не проверяет подлинность клиента перед сервером. Именно этот пробел восполняет Mutual TLS (mTLS).

В данной работе рассматривается протокол mTLS как механизм защиты трафика распределенных систем, его архитектурные преимущества и практические аспекты внедрения.

I. Принцип работы и отличия от стандартного TLS

В стандартном TLS-рукопожатии только сервер аутентифицируется перед клиентом, предъявляя свой цифровой сертификат [1]. Клиент может оставаться анонимным, например, при посещении веб-сайта, или аутентифицироваться на уровне приложения, используя логин и пароль. Это создает уязвимость: установленное TLS-соединение может быть использовано любым клиентом, в том числе и злоумышленником, если он получил доступ к сети.

mTLS добавляет в процесс рукопожатия дополнительный шаг – аутентификацию клиента [2–3]. Для этого клиент также должен обладать своим цифровым сертификатом, подписанным доверенным удостоверяющим центром (Certificate Authority, CA).

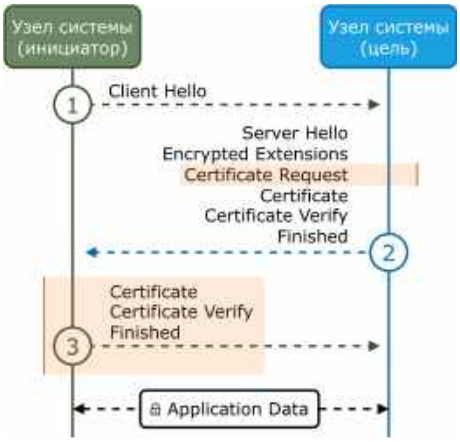


Рис. 1 – Схема процесса рукопожатия с использованием Mutual TLS

Процесс рукопожатия в TLS 1.3 с использованием mTLS осуществляется следующим образом:

1. Client Hello. Стандартное начало TLS-рукопожатия.
2. Server Hello. Наряду с базовыми сообщениями сервер отправляет Certificate Request. Это ключевое отличие от одностороннего TLS, которое указывает на требование аутентификации клиента и приводит к добавлению еще одного этапа рукопожатия.
3. В ответ на запрос сервера, клиент отправляет свой сертификат (Certificate) и сообщение Certificate Verify, которое служит криптографическим доказательством владения приватным ключом. Это обеспечивает аутентификацию клиента перед сервером.

Таким образом, mTLS обеспечивает взаимную аутентификацию (mutual authentication), гарантируя, что каждая сторона соединения знает, с кем она взаимодействует. После успешного завершения рукопожатия между аутентифицированными сторонами устанавливается защищенное соединение для обмена данными.

II. Преимущества использования mTLS в

распределенных системах

Внедрение mTLS предоставляет следующие преимущества для безопасности распределенных систем [4–5]:

- Сквозная аутентификация: каждый узел уверен в подлинности своего собеседника. Это предотвращает атаки подмены (spoofing), когда злоумышленник выдает себя за легитимный сервис, например, за сервис аутентификации или базу данных.
- Конфиденциальность трафика: как и в стандартном TLS, весь обмен данными шифруется, что защищает его от перехвата (sniffing) при передаче между узлами.
- Независимость от сетевого уровня: безопасность реализуется на транспортном уровне модели OSI. Примером преимущества такого подхода служит то, что политики безопасности остаются в силе при изменении IP-адресов сервисов, что является обычной практикой в контейнерных средах, таких как Kubernetes.
- Основа для Zero Trust Architecture: mTLS является практической реализацией принципа «проверяй каждый запрос», ликвидируя концепцию доверенной внутренней сети.

III. Ключевые аспекты внедрения

Основная сложность использования mTLS заключается не в самом протоколе, а в управлении жизненным циклом цифровых сертификатов:

- Масштабируемость: в системе из сотен узлов необходимо генерировать, распределять и обновлять сотни сертификатов. Ручное управление невозможно.
- Автоматизация: в условиях динамически изменяющейся инфраструктуры автоматическая выдача (provisioning) и ротация (rotation) сертификатов становится необходимостью.
- Безопасность приватных ключей: приватные ключи сертификатов должны надежно храниться на каждом сервисе, поскольку обладание таким ключом эквивалентно получению полного доверия со стороны зависимых сервисов, что приводит к нарушению конфиденциальности системы.

Для решения этих задач в распределенных системах применяются специализированные инструменты, которые абстрагируют сложность управления mTLS.

Инструменты, такие как Istio и Linkerd, используют архитектурный паттерн sidecar-прокси – это дополнительный контейнер, который развертывается рядом с каждым экземпляром сервиса и перехватывает весь его сетевой трафик. В сочетании с другими сетевыми функциями образуется так называемая сервисная сетка (Service Mesh). Она использует собственный корневой CA, на основе которого выдает уникальные сертификаты для каждого сервиса через sidecar-прокси. Получая входящий зашифрованный трафик, sidecar-прокси выполняет всю работу по mTLS: проверяет сертификат клиента, расшифровывает данные и передает их основному сервису в открытом виде, используя loopback-интерфейс. Аналогично, исходящий трафик от сервиса перехватывается прокси, который устанавливает mTLS-соединение с удаленным узлом. Для прокси автоматически выполняется ротация сертификатов, жизненный цикл которых поддерживается кратким. Помимо этого, интеграция сервисной сетки осуществляется без вмешательства в код приложения.

Инструменты управления секретами, наподобие HashiCorp Vault, действуют как полнофункциональный доверенный корневой CA для инфраструктуры. Приложения могут программно обращаться к API Vault для динамического запроса короткоживущих сертификатов непосредственно перед установлением соединения. Vault подписывает их, проверяя права приложения на получение сертификата, например, через роль в Kubernetes.

Заключение

Протокол Mutual TLS представляет собой стандартизированный механизм, обеспечивающий взаимную аутентификацию сторон и конфиденциальность передаваемых данных, что делает его инструментом защиты трафика в распределённых системах. Применение mTLS позволяет предотвратить несанкционированный доступ, подмену или перехват данных, формируя основу для реализации архитектур, соответствующих парадигме Zero Trust.

Несмотря на дополнительные затраты на администрирование сертификатов, mTLS обеспечивает высокий уровень доверия между сервисами. Процессы внедрения и управления решаются средствами оркестрации контейнеров и инструментами автоматизации сетевого взаимодействия, что делает использование mTLS масштабируемым и управляемым на практике.

1. Ristić, I. Bulletproof SSL and TLS / I. Ristić – London : Feisty Duck Limited, 2014. – 516 p.
2. RFC 8446 – The Transport Layer Security (TLS) Protocol Version 1.3 [Electronic resource] – Mode of access: `https://datatracker.ietf.org/doc/html/rfc8446`. – Date of access: 26.09.2025.
3. RFC 5246 – The Transport Layer Security (TLS) Protocol Version 1.2 [Electronic resource] – Mode of access: `https://datatracker.ietf.org/doc/html/rfc5246`. – Date of access: 29.09.2025.
4. Whitman, M. E. Principles of Information Security, Fifth Edition / M. E. Whitman, H. J. Mattord – Boston : Cengage Learning, 2014. – 722 p.
5. Горлов А. В., Ноженко К. Э. Анализ и методика проведения перехвата сетевого трафика // Мировая наука. – 2025. – №1. – С. 46–56.

РЕКУРРЕНТНАЯ НЕЙРОННАЯ СЕТЬ НА ОСНОВЕ LSTM-БЛОКОВ С МЕХАНИЗМОМ ВНИМАНИЯ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИЙ В РЕЧИ

Краснопрошин Д. В., Вашкевич М. И.
Кафедра ЭВС, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: daniil.krasnoproshin@gmail.com, vashkevich@bsuir.by

В работе предлагается механизм мультивекторного мягкого внимания для рекуррентных нейронных сетей на основе LSTM для задачи распознавания эмоций в речи. Эксперименты проводились на наборе данных RAVDESS. Для автоматизированного подбора оптимальных гиперпараметров сети использовался метод байесовской оптимизации. Результаты экспериментов показывают, что увеличение количества векторов внимания с одного до 64 приводит к улучшению среднего значения метрики UAR на 0.9%, что является статистически значимым результатом и подтверждает целесообразность использования предложенного механизма внимания в архитектурах на основе LSTM для задачи классификации эмоций в речи.

Введение

Распознавание эмоций в речи является одной из ключевых задач в области обработки естественного языка. В настоящее время продолжается активный поиск архитектур нейронных сетей, которые имеют умеренную сложность и высокую точность работы [1–3].

Целью данного исследования является разработка мультивекторного механизма мягкого внимания и количественная оценка данного механизма на эффективность LSTM-моделей при распознавании эмоций в речи. Предполагается, что увеличение числа векторов внимания позволит модели учитывать более разнообразные аспекты эмоционального контекста.

I. Система распознавания эмоций на основе РНН

Для параметризации речевого сигнала в работе используются мел-частотные кепстральные коэффициенты (МЧКК). Сигнал разделяется на короткие фреймы (кадры) после чего из каждого фрейма извлекается n-мерный вектор МЧКК [2]. В результате речевой сигнал преобразуется в последовательность векторов:

$$X = [\mathbf{x}_0, \mathbf{x}_1 \dots \mathbf{x}_{T-1}], \mathbf{x}_t \in \mathbb{R}^n, \tag{1}$$

где T – длина последовательности.

За основу в работе берется архитектура РНС с мягким механизмом внимания, предложенная в [3]. Входная последовательность подается в РНС типа LSTM [4], которая формирует последовательность скрытых состояний $\mathbf{h}_0, \mathbf{h}_1, \dots, \mathbf{h}_T$. Далее рассчитывается взвешенная сумма скрытых состояния, которая называется вектором контекста:

$$\mathbf{h}_{wp} = \sum_{t=0}^{T-1} \alpha_t \mathbf{h}_t, \tag{2}$$

где α_t – весовые коэффициенты отражающие значимость вектора состояния $\mathbf{h}_t$ в формировании вектора контекста.

На основании вектора контекста $\mathbf{h}_{wp}$ выполняется классификация эмоции. Т. е. вектор $\mathbf{h}_{wp}$ подается на полносвязный слой с активационной функцией softmax. Общая схема описанного подхода представлена на рис. 1.

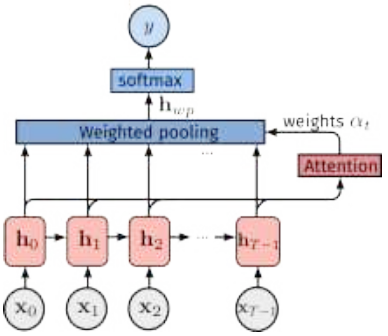


Рис. 1 – Классификация эмоций при помощи РНС с механизмом внимания

В оригинальной работе [3] для формирования весов использовался механизм мягкого внимания (англ. *soft attention*):

$$a_t = \text{softmax}(e_t) = \frac{\exp(e_t)}{\sum_{t=0}^{T-1} \exp(e_t)}, \tag{3}$$

где $e_t = \mathbf{u}^T \mathbf{h}_t$ – оценка внимания (англ. *attention score*), где $\mathbf{u}$ – вектор внимания (обучаемый параметр).

II. Мультивекторный механизм мягкого внимания

В настоящей работе предлагается мультивекторный механизм мягкого внимания, согласно которому оценка внимания рассчитывается по формуле

$$e_t = \text{maxpool}_{1D}\left(\begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \dots \\ \mathbf{u}_{N_{att}^T} \end{bmatrix} \mathbf{h}_t\right), \tag{4}$$

где N_{att} – число векторов внимания, а $\mathbf{u}_i$ – набор векторов внимания, каждый из которых отвечает

за независимое взвешивание векторов скрытых состояний.

Следует отметить, что обычный механизм мягкого внимания является частным случаем мультивекторного механизма внимания, когда $N_{att} = 1$.

III. Описание эксперимента

Для экспериментальной проверки мультивекторного механизма внимания использовалась однослойная РНС с LSTM-блоками. Входные данные представляют собой нормализованные МЧКК размерностью 1×34, размерность скрытого состояния РНС выбиралась равной 64. Количество векторов внимания изменялось в диапазоне от одного до 64: $N_{att} \in \{1, 2, 4, 8, 16, 32, 64\}$.

Для автоматизированного поиска оптимальных гиперпараметров (скорость обучения, размера батча, вероятность дропаута, скорость «затухания» весов и количество циклов в планировщике косинусного отжига) использовался метод байесовской оптимизации, реализованный в библиотеке Optuna языка Python.

Оптимизация выполнялась для базовой модели с одним вектором внимания. Полученные оптимальные параметры использовались далее для всех экспериментов с различным числом векторов внимания.

Эксперименты проводились на наборе данных RAVDESS [1], включающем аудиозаписи с восьмью эмоциональными категориями. Для каждой конфигурации модели обучение повторялось 10 раз с различными начальными значениями весов. Для оценки качества использовался показатель UAR (*Unweighted Average Recall*). Результаты экспериментов приведены в табл. 1.

Табл. 1. Результаты экспериментов

Model	UAR	UAR (max)
LSTM-v-1	0,5510 ± 0,0071	0,5625
LSTM-v-2	0,5566 ± 0,0104	0,5742
LSTM-v-4	0,5548 ± 0,0065	0,5618
LSTM-v-8	0,5593 ± 0,0079	0,5710
LSTM-v-16	0,5586 ± 0,0112	**0,5775**
LSTM-v-32	0,5581 ± 0,0109	0,5729
LSTM-v-64	**0,5598 ± 0,0083**	0,5742

На рис. 2 представлены сглаженные распределения значений UAR для моделей с одним и 64-мя векторами внимания. Модель с большим числом векторов внимания демонстрирует более высокие значения UAR.

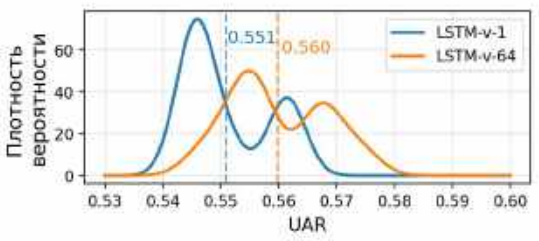


Рис. 2 – Сглаженные распределения точностей моделей: LSTM-v-1 и LSTM-v-64

Статистический анализ различий между моделями проводился с использованием t-теста для независимых выборок. Полученные результаты показали наличие статистически значимой разницы ($p < 0{,}05$) между моделью с одним и 64 векторами внимания. Среднее улучшение UAR составило 0,9%.

В целом, рост точности можно объяснить тем, что каждый вектор внимания учится фокусироваться на различных аспектах эмоциональных признаков – интонации, тембре, спектральной энергии и т. д. Таким образом, использование нескольких таких векторов создаёт более богатое представление входного сигнала, повышая способность модели выделять эмоционально релевантные сегменты речи.

Несмотря на то что абсолютное улучшение метрики невелико (порядка 1%), оно является стабильным и статистически подтверждённым. Это указывает на потенциал дальнейших исследований в данном направлении.

IV. Заключение

В работе проведено исследование влияния количества векторов внимания на качество распознавания эмоций в речи с использованием LSTM-блоков. Эксперименты на датасете RAVDESS показали, что увеличение числа векторов внимания до 64 приводит к статистически значимому росту метрики UAR. Данный результат подтверждает, что мультивекторный механизм мягкого внимания способствует более эффективному обучению эмоциональных представлений в речевом сигнале.

V. Список литературы

1. C. Luna-Jiménez, et al., “Multimodal emotion recognition on RAVDESS dataset using transfer learning,” Sensors, vol. 21, 2021, pp. 1-29.
2. D. V. Krasnoproshin and M. I. Vashkevich “Speech emotion recognition method based on support vector machine and suprasegmental acoustic features,” Doklady BGUIR, vol. 22, no. 3, 2024, pp. 93-100. (In Russain)
3. S. Mirsamadi, E. Barsoum and C. Zhang “Automatic speech emotion recognition using recurrent neural networks with local attention, Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 2227-2231.
4. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, 1997, pp. 1735–1780.

АЛГОРИТМ АНАЛИЗА СРЕДНЕЙ ДЛИНЫ СЛОВА В СЛОВАРЕ В.И. ДАЛЯ

Кузнецов Е. В., Приходько Н. Н., Чехомов Е. Г., Косарева Е. М.
Факультет компьютерного проектирования, Кафедра ПИКС, БГУИР
Минск, Республика Беларусь
E-mail: {yauheni.kuznetsou, nikitaprix50, genchek1, kksrvvv }@gmail.com

В работе представлен алгоритм статистического анализа лексического состава русского языка на основе Толкового словаря живого великорусского языка В. И. Даля. С использованием языка программирования Python и библиотек pandas, matplotlib реализована обработка текстовых данных словаря, включающая извлечение словарных единиц, вычисление средней длины слова и построение распределения длин слов. Результаты анализа показывают, что средняя длина слова в словаре Даля составляет 8,49 символа, что превышает средние показатели для современного русского языка. Исследование демонстрирует применение методов Data Science для лингвистического анализа.

Введение

Толковый словарь живого великорусского языка В. И. Даля – один из крупнейших словарей русского языка, содержащий около 200 тысяч слов и 30 тысяч пословиц, поговорок и загадок [1, 2]. Словарь был создан в середине XIX века и отражает состояние русского языка того периода, включая народную речь, областные диалекты и профессиональную лексику [3].

Статистический анализ лексического состава языка позволяет выявить закономерности в структуре словарного запаса и провести сравнительные исследования [4]. Современные методы обработки данных с использованием языка программирования Python и специализированных библиотек открывают новые возможности для количественного анализа лингвистических материалов.

Цель данного исследования – разработка и реализация алгоритма анализа средней длины слова в словаре Даля с применением инструментов Data Science.

I. Описание исходных данных

Исходным материалом для исследования послужил электронный текстовый файл, содержащий словарные статьи из словаря Даля. Данные представлены в текстовом формате с кодировкой Windows-1251 (CP1251), что характерно для русскоязычных текстов. Общий объём исходного файла составил более 500 тысяч символов.

Структура словаря предполагает, что каждая строка файла начинается с заглавного слова словарной статьи, после которого следует толкование, примеры употребления и другая лингвистическая информация. Для статистического анализа требовалось извлечь только заглавные слова – основные словарные единицы.

Предварительный анализ данных показал необходимость преобразования кодировки из CP1251 в UTF-8 для корректной обработки кириллических символов в современных версиях Python.

II. Методика обработки данных

Алгоритм обработки данных реализован на языке программирования Python с использованием библиотек re (регулярные выражения), pandas (анализ данных) и matplotlib, seaborn (визуализация) [5].

III. Извлечение словарных единиц

Первый этап обработки данных – извлечение заглавных слов из каждой строки исходного файла. Для этого применяются регулярные выражения, позволяющие находить последовательности буквенных символов русского и латинского алфавитов. Используется шаблон \b[а-яА-Яа-zA-Z]+\b, где \b обозначает границу слова.

Алгоритм последовательно обрабатывает каждую строку исходного файла, извлекает из неё все слова с помощью функции re.findall() и сохраняет первое слово (заглавное слово словарной статьи) в выходной файл. Результатом данного этапа является текстовый файл, содержащий список всех заглавных слов словаря – по одному на каждой строке.

IV. Статистический анализ

На втором этапе выполняется статистический анализ полученного списка слов. Для этого все слова загружаются в память и преобразуются в структуру данных pandas.DataFrame – табличное представление, удобное для анализа.

Для каждого слова вычисляется его длина – количество символов. Создаётся новый столбец таблицы, содержащий длины всех слов. На основе этих данных вычисляются следующие статистические показатели:

- средняя длина слова (среднее арифметическое);
- минимальная и максимальная длина слова;
- распределение частот для каждой длины слова;
- медианная длина и мода распределения.

V. Визуализация результатов

Третий этап – визуализация полученных результатов. С использованием библиотеки pyplot строится гистограмма распределения длин слов. Гистограмма дополняется линией ядерной оценки плотности (KDE), которая позволяет визуализировать распределение в виде непрерывной кривой (см. рис. 1).

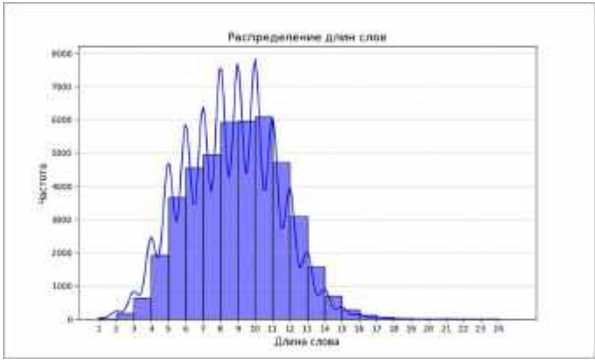


Рис. 1 – Гистограмма распределения длин слов

VI. Результаты исследования

В результате обработки словаря Даля было извлечено 44 691 заглавное слово. Статистический анализ показал следующие результаты:

- средняя длина слова составила 8,49 символа;
- минимальная длина – 1 символ;
- максимальная длина – 24 символа;
- наиболее частая длина слова (мода) – 10 символов (13,61% всех слов);
- медианная длина – 9 символов.

Распределение слов по диапазонам длин представлено в таблице 1.

Таблица 1 – Распределение слов по длине

Диапазон длин	Количество слов	Доля, %
1–3 символа	1 054	2,36
4–6 символов	10 154	22,72
7–9 символов	16 827	37,65
10–12 символов	13 888	31,08
13–15 символов	2 566	5,74
16 и более	202	0,45

Анализ показывает, что наибольшую долю (37,65%) составляют слова длиной от 7 до 9 символов. Слова длиной 10–12 символов занимают второе место (31,08%). Таким образом, около 69% всех слов словаря имеют длину от 7 до 12 символов.

Полученное значение средней длины слова (8,49 символа) существенно превышает среднюю длину слова в современном русском языке, которая, по различным оценкам, составляет от 5,28 до 6 символов [6, 7]. Это объясняется спецификой словаря Даля, который включает большое количество диалектных, устаревших и сложных составных слов.

Среди самых длинных слов в словаре – «понедоброжелательствовал» (24 символа), «беспокровительственный» (22 символа), «переосвидетельствовать» (22 символа).

Распределение слов по первым буквам алфавита показало, что наибольшую долю (24,78%) составляют слова, начинающиеся на букву «П», что связано с большим количеством приставочных образований в русском языке.

Заключение

В работе представлен эффективный алгоритм статистического анализа лексического состава исторического словаря с использованием современных инструментов Data Science. Разработанный метод позволяет извлекать количественные характеристики словарного состава и визуализировать результаты анализа.

Результаты исследования показывают, что словарь Даля характеризуется относительно большой средней длиной слова (8,49 символа), что отражает богатство словообразовательных возможностей русского языка XIX века и широкое представление диалектной и архаичной лексики.

Предложенная методика может быть применена для сравнительного анализа различных словарей и текстовых корпусов, а также для исследования эволюции лексического состава языка.

1. Толковый словарь живого великорусского языка / В. И. Даль. – М.: Русский язык, 1863–1866. – 4 т.
2. Национальный корпус русского языка [Электронный ресурс]. – Режим доступа: `https://ruscorpora.ru`. – Дата доступа: 17.10.2025.
3. Бодуэн де Куртенэ, И. А. О словаре Даля / И. А. Бодуэн де Куртенэ // Толковый словарь живого великорусского языка В. И. Даля. – СПб., 1903. – Т. 1. – С. I–XLII.
4. McKinney, W. Python for Data Analysis / W. McKinney. – O'Reilly Media, 2017. – 544 p.
5. Лутц, М. Изучаем Python / М. Лутц. – СПб.: Символ-Плюс, 2019. – 1280 с.
6. Пиотровский, Р. Г. Математическая лингвистика / Р. Г. Пиотровский, К. Б. Бектаев, А. А. Пиотровская. – М.: Высшая школа, 1977. – 383 с.
7. Гусейн-Заде, С. М. О средней длине слова в русском языке / С. М. Гусейн-Заде // Проблемы структурной лингвистики. – М.: Наука, 1982. – С. 82–88.

ОРГАНИЗАЦИЯ ВВОДА-ВЫВОДА ПРИ ВЫПОЛНЕНИИ ПАРАЛЛЕЛЬНЫХ ВЫЧИСЛЕНИЙ В СИСТЕМЕ FLC2

Логинова И. П.
Объединённый институт проблем информатики Национальной академии наук Беларуси
Минск, Республика Беларусь
E-mail: irilog@mail.ru

Ввод-вывод – важный фактор при проведении высокопроизводительных вычислений в современных многопроцессорных системах. Организация ввода-вывода при параллельных вычислениях является серьезной задачей, как с точки зрения эффективности, так и с точки зрения мобильности и масштабируемости параллельной программы. Рассмотрены некоторые аспекты реализации операций ввода-вывода в системе FLC2 в процессе выполнения процедур логической оптимизации при проведении параллельных вычислений на многоядерной вычислительной системе.

Введение

Функционально операции ввод-вывода в программах можно разделить на такие группы:

1. Явные операторы ввода-вывода, которые необходимы для ввода начальных данных и вывода окончательных результатов.
2. Для задач с большим временем выполнения необходимо периодически запоминать состояние задачи, так называемые контрольные точки.
3. В задачах требуется запоминать текущее состояние вычислений через определенные промежутки времени, например, для использования в подсистемах визуализации.
4. Вычисления с большими массивами данных, расположенными на внешней памяти.

Разные функциональные назначения операций ввода-вывода требуют использования разных подходов к их реализации. Неэффективное управление вводом-выводом при проведении параллельных вычислений может свести на нет любые оптимизации параллельного выполнения задач, если в них имеется интенсивный ввод-вывод.

I. Параллельные вычисления в FLC2

Функционал системы логической оптимизации функционально-структурных описаний дискретных устройств FLC2 [1] позволяет реализовать достаточно простой подход к организации параллельных вычислений на многоядерной вычислительной системе. В основе лежит работа параллельных процессов, каждый из которых выполняется на ядре процессора, представляя работу одного потока в одном процессе [2]. В отдельном процессе производится старт команды, аргументы которой определяют имя внешней программы и исходный объект оптимизации. Исходные объекты исполняемой программы являются независимыми друг от друга. Организовать совокупности независимых объектов для проведения параллельных вычислений с использованием программ технологически независимой оптимизации позволяет среда системы FLC2. Средствами FLC2 можно проводить проектные операции для преобразования исходного объекта в одноранговую сеть независимых объектов и осуществлять параллельные вычисления посредством запуска процессов, в которых выполняется программа для одного элемента сети. Для такой организации параллельных вычислений был использован стандарт OpenMP (версия 3.5 и выше), который специфицирует набор директив компилятора (для языков C, C++ и др.), функции библиотек и переменных окружения [3].

II. Обзор существующих решений

В настоящее время нет общего широко используемого интерфейса параллельного ввода–вывода (далее в/в). Для различных технологий параллельного программирования, существуют такие решения [3], как, например, MPI-IO или OpenMP-IO. Как правило, при реализации различных параллельных программ используются специализированные библиотеки в/в, предназначенные для работы с научными данными различных форматов. Отсутствие общего подхода и стремление к достижению максимальной производительности в/в привело к тому, что разработчики новых языков программирования, вынуждены разрабатывать собственные средства для параллельного в/в. Для приложений, использующих OpenMP, возможности в/в существенно ограничены, следующими схемами организации в/в, когда N процессов пишут в один файл или N процессов пишут в N файлов. Для OpenMP можно использовать следующие подходы. Например, при использовании совместно используемого файла для поддержки согласованности файла с результатом, операции чтения и записи должны выполняться вне параллельных областей [4]. В случае, если несколько процессов обращаются к одному файлу, доступ к дескриптору файла (общему для всех потоков) должен быть защищен, например, с использованием механизма критических секций. Другой подход – позволить каждому процессу использовать отдельный файл, чтобы избежать ошибок типа гонок (race condition) или ошибок синхронизации при доступе к общему дескрипто-

ру файла. Хотя такой подход часто приводит к более высокой производительности, он имеет следующий недостаток: он требует дополнительных (достаточно дорогостоящих) шагов пред- и постобработки, для того чтобы создать необходимое количество входных файлов и объединить выходные файлы, записанные разными процессами. Для повышения производительности операций в/в в OpenMP приложениях разрабатываются интерфейсы параллельного ввода-вывода, который обеспечивает доступ нескольких потоков к одному файлу без необходимости явного использования операций синхронизации [3,4]. Эти спецификации очень близки к спецификациям MPI-IO. Основное отличие от MPI-IO – отсутствие неблокирующих операций ввода-вывода.

III. Схема обработки вызовов ввода-вывода в параллельных процессах

Общепринято, что процессы рассматриваются операционной системой (ОС) как заявки или контейнеры для всех видов ресурсов, кроме процессорного времени. Этот ресурс распределяется ОС между другими единицами работы – потоками, которые представляют собой последовательности команд. В простейшем случае процесс состоит из одного потока. Новый процесс может быть создан по запросу текущего процесса. Ресурсы могут быть выделены процессу на все время его жизни или только на определенный период. За обеспечение процессов необходимыми ресурсами отвечает подсистема управления процессами и потоками. Создание процесса включает загрузку кодов и данных исполняемой программы данного процесса в оперативную память. Когда процесс завершается благодаря одному из событий, например, обычного завершения выполняемой в процессе программы подсистема управления процессами закрывает все файлы, с которыми работал процесс, освобождает области оперативной памяти, отведенные под коды, данные и системные информационные структуры процесса.

Стандартные средства современных ОС не позволяют создать для одного приложения несколько процессов для параллельной работы, но многоядерность и использование технологии OpenMP [4] позволили реализовать подход [2]. При этом подходе процесс, выполняемый в каждом ядре, сам решает, когда ему отдать управление кооперативному (non-preemptive) планировщику задач, чтобы последний выбрал из пула задач готовый к выполнению процесс, используя механизм work stealing. Процесс работает в пользовательском режиме, но когда он обращается к системному вызову в/в, то ввод-вывод осуществляется одновременно с вычислениями в ядрах процессора, которые «отвлекаются» только для выдачи команд контроллерам о начале и завершении ввода-вывода. ОС может выполнять вызовы в/в в синхронном и асинхронном режимах.

В первом случае, как представлено на схеме на рис. 1, процесс, сделавший такой вызов, приостанавливается до тех пор, пока контроллер в/в не выполнит свою работу. Затем планировщик переводит процесс в состояние продолжения вычислений, процесс может воспользоваться результатами вызова в/в. Синхронный в/в выполняется следующим образом. В каждым процессе, выполняемом на ядре многопроцессорной системы, для ввода исходных данных и вывода результата открываются отдельные файлы, что является достаточно удобным средством организации параллельной работы контроллеров ввода-вывода.

Асинхронный вызов в/в не приводит к переводу процесса в режим ожидания и после выполнения некоторых начальных системных действий, например запуска операции ввода-вывода, управление возвращается вызывающему процессу. Асинхронный системный вызов в подходе [2] используется для получения трассы выполнения параллельных процессов и визуализации.

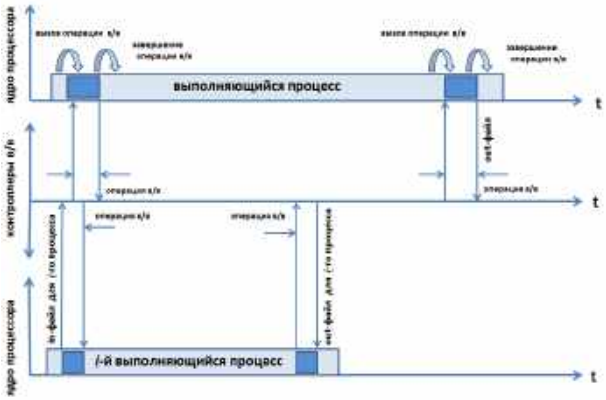


Рис. 1 – Схема синхронного выполнения операций ввода вывода в системе FLC2 при работе параллельных процессов

IV. Список литературы

1. Бибило, П. Н. Система логической оптимизации функционально-структурных описаний цифровых устройств на основе продукционно-фреймовой модели представления знаний /П. Н. Бибило, В. И. Романов // Проблемы разработки перспективных микро- и наноэлектронных систем – (МЭС-2020) : IX Юбилейная Всероссийская научно-техническая конференция : Сборник трудов ; под общей ред. А. Л. Стемпковского. – Москва : ИППМ РАН, 2021 Москва : ИППМ РАН, 2020. – С. 9-16.
2. Бибило, П. Н. Экспериментальное сравнение эффективности программ минимизации систем булевых функций в классе дизъюнктивных нормальных форм / П. Н. Бибило, И. П. Логинова // Информатика. – 2022. – Т. 19, № 2. – С. 26-55.
3. Бахтин, В. А. Отладка параллельных программ в DVM-системе / В. А. Бахтин, Д. А. Захаров, А. А. Ермичев, В. А. Крюков // Электронные библиотеки // – 2020. – Vol. 23, № 4. – С. 866–886.
4. Kshitij Mehta, Edgar Gabriel, Barbara Chapman. Specification and Performance Evaluation of Parallel I/O Interfaces for OpenMP. OpenMP in a Heterogeneous World. Lecture Notes in Computer Science Volume 7312, 2012, pp. 1-14.

ПРОТОТИПИРОВАНИЕ ПОДГОТОВКИ ДАННЫХ ИЗ ВИДЕОРЯДА И СЕГМЕНТАЦИИ ДЛЯ ГЕНЕРАЦИИ 3D МОДЕЛИ ГОЛОВЫ ЧЕЛОВЕКА ПО 2D ИЗОБРАЖЕНИЯМ

Лукашевич М. М., Венгеренко В. В., Воронов А. А.
Кафедра информационных систем управления, Белорусский государственный университет
Объединенный институт проблем информатики Национальной Академии наук Беларуси
Минск, Республика Беларусь
E-mail: lukashevichmm@bsu.by, {vengerenko, voronov}@lsi.bas-net.by

Материал посвящен описанию разработки приложения для автоматического создания 3D модели головы человека на основе изображений, полученных с камеры. Приложение позволит исследовать модель во всех проекциях и найдет применение в медицине и судебной экспертизе.

Введение

Современные подходы к разработке алгоритмов для восстановления 3D модели лица и головы из одиночного или набора изображений позволяют автоматически создавать 3D модель головы человека и исследовать такую модель в различных проекциях. Решение этой задачи требует применения специализированных программных и технических средств, разработка которых – актуальное и перспективное направление научно-технической деятельности. Применяемые алгоритмы восстановления 3D модели, реализованные программно, определяют требования к необходимому аппаратному обеспечению (камерам видеонаблюдения) систем видеоаналитики, которые направлены на обеспечение мониторинга безопасности и полезны в судебной экспертизе или медицине.

I. Алгоритм восстановления 3D модели головы по 2D изображениям

Алгоритм восстановления 3D модели головы по 2D изображениям состоит из следующих шагов [1]:

1. Подготовка данных из видеопотока.
2. Выделение объектов интереса и отделение их от фона, сегментация изображений, дальнейшая обработка, связанная с удалением неинформативных артефактов и определение очертаний объектов интереса. В основе всех известных алгоритмов восстановления 3D модели лежит именно выделение объектов интереса, отделение их от фона и сегментация.
3. Построить трехмерную сетку и вписать в нее полученные изображения.
4. Текстурирование 3D модели.

Рассмотрим подробнее шаги 1–2 представленного алгоритма. Подготовка данных из видеопотока состоит из следующих шагов:

1. Вычисление основных характеристик видеопотока. Ширина, высота изображения и частота кадров (FPS).
2. Масштабирование изображения. Рассчитывается коэффициент уменьшения для приведения кадров к максимальному размеру 720 пикселей по большей стороне.
3. Выбор ключевых кадров. Выполняется анализ кадров по показателю резкости с помощью оператора Лапласа, который применяется к изображению после перевода его в полутон и бинаризации. Для каждой секунды видео определяется самый четкий кадр из диапазона 5–25 кадров. Это позволяет пропускать возможные артефакты в начале секунды.
4. Сохранение результатов. Оригинальные ключевые кадры сохраняются в соответствующий каталог «№секунды.png».
5. Определение пяти ключевых кадров на основе классификационной модели. Ключевые кадры из видеопотока – анфас, вид слева, справа, сзади, сверху.

II. Алгоритмы сегментации изображений головы от фона

Алгоритмы удаления фона на изображениях головы человека используют такие архитектуры глубоких сверточных нейронных сетей, как U-Net, DeepLab-V3+ и Mask R-CNN, которые показывают высокое качество сегментации, а для работы на мобильных устройствах и в реальном времени предложены оптимизированные архитектуры: LiteUNet, MODNet и PortraitNet. Они сочетают в себе приемлемое качество сегментации и высокую скорость обработки, что делает их подходящими для внедрения в мобильные потребительские устройства. Одним из наиболее перспективных направлений является использование Transformer-архитектур, таких как Segmenter, Trans-UNet, Swin-Unet, которые особенно эффективны в обработке [2–8]. В ходе реализации проекта разработаны и экспериментально исследованы алгоритмы сегментации при помощи нейросетевых архитектур UNet, PortraitNet, Segmenter, SAM. Модели U-Net, PortraitNet, Segmenter, SAM были обучены на наборе данных, представленном исходными изображения головы человека в разных ракурсах и эталонными масками сегментации. Модель SAM использовалась без дообучения с предварительно расстановкой опорных точек для сегментации. Ре-

зультаты тестирования показали, что наилучшую точность демонстрирует модель Segmenter (IoU = 0,9739, Dice = 0,9867, Accuracy = 0,9973).

Приведем описание алгоритма для нейросетевой архитектуры Segmenter.

1. Разбиение изображения на патчи. Входное изображение разбивается на фиксированные блоки (патчи) одинакового размера, например, 16x16 пикселей. Каждый патч преобразуется в вектор признаков с помощью линейного слоя.
2. Добавление позиционных эмбеддингов. Каждому визуальному токену (вектору патча) добавляется позиционный эмбеддинг, который указывает на его местоположение в исходном изображении. Это позволяет модели учитывать пространственную структуру изображения.
3. Обработка через Transformer-encoder. Последовательность токенов подается в Transformer-encoder, где каждый токен взаимодействует со всеми остальными через механизм внимания (self-attention). На этом этапе модель выделяет глобальные признаки и устанавливает связи между различными частями изображения.
4. Инициализация масочных токенов. Decoder использует набор обучаемых масочных токенов (learnable mask tokens), каждый из которых отвечает за формирование определённой семантической маски (например, для каждого класса объекта).
5. Восстановление масок через кросс-внимание. Масочные токены взаимодействуют с закодированными признаками из encoder'a через кросс-внимание, что позволяет decoder'у формировать детализированные маски, соответствующие реальной структуре изображения.
6. Генерация финальной карты сегментации. На выходе decoder'a получается набор масок, каждая из которых соответствует определённому объекту или классу. Эти маски объединяются в финальную карту сегментации, совпадающую по размерам с исходным изображением.

Пример результатов сегментации приведен на рис. 1.

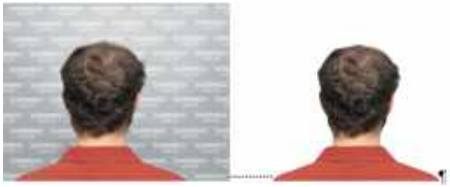


Рис. 1 – Примеры результатов сегментации

Заключение

В ходе работы были разработаны алгоритмы подготовки данных из видеопотока и удаления фона на изображениях головы человека, являющиеся важной частью решения задачи реконструкции 3D модели головы человека по серии фотографий, и выполнено их экспериментальное исследование. Выполнен анализ существующих подходов к 3D реконструкции головы человека: методов фотограмметрии, нейросетевых моделей и ручного моделирования. Установлено, что для условий мобильной съемки наиболее перспективным является использование методов фотограмметрии. Реализовано также извлечение ключевых кадров, их масштабирование с сохранением пропорций и оценка резкости с использованием оператора Лапласа, что позволяет эффективно отбирать кадры для дальнейшей сегментации. Для решения задачи удаления фона (бинарная сегментация) протестированы четыре современные архитектуры нейронных сетей: U-Net, PortraitNet, Segmenter, SAM. Результаты тестирования показали, что наилучшую точность демонстрирует модель Segmenter.

Работа выполнена в рамках договора с БРФФИ Ф25ТУРГ-001 от 05 марта 2025 г., тема "Приложение для генерации компьютерной 3D модели головы человека и ее сравнение с изображениями, полученными с камеры".

Список литературы

1. Schönberger, J. L. Structure-from-Motion Revisited / J. L. Schönberger, J. M. Frahm // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 4104–4113. – DOI: 10.1109/CVPR.2016.445.
2. Wu, C. Towards Linear-Time Incremental Structure from Motion / C. Wu // Proceedings of the International Conference on 3D Vision (3DV). – 2013. – P. 127–134. – DOI: 10.1109/3DV.2013.25.
3. Yao, Y. MVSNet: Depth Inference for Unstructured Multi-View Stereo / Y. Yao, Z. Luo, S. Li, T. Fang, L. Quan // Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – P. 767–783. – DOI: 10.1007/978-3-030-01237-3_47.
4. Gu, X. Cascade Cost Volume for High-Resolution Multi-View Stereo / X. Gu, Z. Fan, S. Zhu, Z. Dai, F. Tan, P. Tan // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2020. – P. 3265–3274. – DOI: 10.1109/CVPR42600.2020.00257.
5. Choy, C. B. 3D-R2N2: A Unified Approach for Single and Multi-view 3D Object Reconstruction / C. B. Choy, D. Xu, J. Gwak, K. Chen, S. Savarese // Proceedings of the European Conference on Computer Vision (ECCV). – 2016. – P. 628–644. – DOI: 10.1007/978-3-319-46484-8_-38.
6. Qi, C. R. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space / C. R. Qi, L. Yi, H. Su, L. J. Guibas // Advances in Neural Information Processing Systems (NIPS). – 2017. – P. 5099–5108.
7. Mildenhall, B. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis / B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng // Proceedings of the European Conference on Computer Vision (ECCV). – 2020. – P. 405–421. – DOI: 10.1007/ 978-3-030-58452-8_24.
8. Saito, S. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization / S. Saito, Z. Huang, R. Natsume, S. Morishima, A. Kanazawa, H. Li // Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). – 2019. – P. 2304–2314. – DOI: 10.1109/ICCV.2019.00239.

АНАЛОГ ШИФРСХЕМЫ ЭЛЬ-ГАМАЛЯ НА ОСНОВЕ ЭЛЛИПТИЧЕСКИХ КРИВЫХ

Матвеев Н. С., Орлов Д. В., Скиба И. Г.
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {n.matveev, i.skiba}@bsuir.by

В работе рассматривается построение криптографической схемы, аналогичной шифрсхеме Эль-Гамаля, на основе эллиптических кривых. Приведены процедуры шифрования и расшифрования сообщений. Описаны особенности применения эллиптических кривых в сравнении с классической схемой, основанной на дискретном логарифмировании в мультипликативной группе.

Введение

Шифрсхема Эль-Гамаля является одной из базовых криптографических схем с открытым ключом. Ее классическая версия основана на задаче дискретного логарифмирования в мультипликативной группе по модулю большого простого числа. При этом для обеспечения криптостойкости требуется использование длинных ключей, что приводит к увеличению вычислительных затрат.

Одним из направлений развития является использование эллиптических кривых, для которых также формулируется задача дискретного логарифмирования. Сложность этой задачи позволяет построить аналогичные по структуре криптографические схемы, но с меньшими параметрами. В данной работе рассматривается алгоритм Эль-Гамаля на эллиптических кривых, описываются его основные этапы и особенности применения.

I. Классическая шифрсхема Эль-Гамаля

Шифрсистема Эль-Гамаля была предложена в 1985 г. и является фактически одним из вариантов метода выработки открытых ключей Диффи-Хеллмана. Криптографическая стойкость данной системы основана на сложности проблемы логарифмирования в мультипликативной группе конечного простого поля.

Задаётся простое число p и примитивный элемент g в мультипликативной группе Z_p^*.

Числа $x \in \{1, \ldots, p-2\}$ и $y = g^x \bmod p$ выступают в качестве открытого и закрытого ключей соответственно.

Пусть сообщение $m \in Z_p^*$. Для шифрования выбирается случайное число $k \in \{1, \ldots, p-2\}$. Вычисляются:

$$c_1 = g^k \bmod p,$$

$$c_2 = m \cdot y^k \bmod p.$$

Шифротекст имеет вид пары (c_1, c_2).

Получатель, обладающий закрытым ключом x, вычисляет:

$$m = c_2 \cdot (c_1^x)^{-1} \bmod p,$$

где $(c_1^x)^{-1}$ обозначает мультипликативный обратный элемент по модулю p.

Для проверки корректности схемы подставим выражения для c_1 и c_2:

$$c_2 \cdot (c_1^x)^{-1} \equiv (m \cdot y^k) \cdot (g^{kx})^{-1} \pmod p.$$

Так как $y = g^x$, то

$$m \cdot g^{kx} \cdot (g^{kx})^{-1} \equiv m \pmod p.$$

Следовательно, исходное сообщение m корректно восстанавливается [1].

II. Основная идея

Классическая схема Эль-Гамаля строится на задаче дискретного логарифмирования в мультипликативной группе Z_p^*. Адаптация заключается в замене этой группы на аддитивную группу точек эллиптической кривой $E(F_p)$.

В обоих случаях принцип работы сохраняется:

- выбор секретного ключа x и вычисление открытого ключа как g^x (в Z_p^*) или xP (в $E(F_p)$);
- использование случайного параметра k для шифрования;
- возможность восстановления исходного сообщения благодаря симметрии операций.

III. Параметры системы

Параметрами схемы на эллиптических кривых являются [2]:

- простое число p — модуль эллиптической кривой;
- эллиптическая кривая $E(F_p)$;
- целое число $m \neq p$ — порядок группы точек эллиптической кривой $E(F_p)$;
- простое число q — порядок циклической подгруппы группы точек эллиптической кривой $E(F_p)$, для которого выполнены следующие условия:

$$m = nq,\ n \in Z,\ n \geq 1.$$

- точка $P \neq O$ эллиптической кривой $E(F_p)$ с координатами (x_p, y_p), удовлетворяющая равенству $qP = O$.

Каждый пользователь схемы цифровой подписи должен обладать личными ключами:

- закрытым ключом – (d, q), удовлетворяющим неравенству $0 < d < q$;
- открытым ключом – (P, q, Q), где точка эллиптической кривой Q с координатами (x_q, y_q), удовлетворяет равенству $dP = Q$.

IV. Шифрование

Для точки $R \in F$ через $X(R)$ обозначим её x-координату.

Процесс шифрования можно описать следующим образом:

- генерируется случайный $k \in \{1, \ldots, n-1\}$;
- вычисляется $C_1 = kP$;
- вычисляется общий секрет $S = kQ$. Берётся $u = X(S)$;
- шифротекст: $C_2 = M + u \pmod q$;
- отправляется пара (C_1, C_2).

Здесь C_1 играет роль аналога $g^k \bmod q$, а C_2 — аналога $m \cdot y^k \bmod p$ из классической схемы.

V. Расшифрование

Для расшифровки сообщения:

- получатель вычисляет

$$S' = dC_1 = d(kP) = k(dP) = kQ;$$

- берёт $u' = X(S')$ и восстанавливает

$$M = C_2 - u' \pmod q.$$

Здесь операция вычитания точек соответствует делению в классической схеме, где сообщение восстанавливалось как $m = c_2 \cdot (c_1^x)^{-1} \bmod p$.

VI. Корректность

Подставим определения:

$$C_2 - u' \bmod q = (M + X(k(Q)) - X(d(kP)) \bmod q.$$

Так как $Q = dP$, имеем:

$$M + X(k(dP)) - X(d(kP)) \bmod q = M.$$

VII. Применение

Ключевое преимущество предложенной адаптации схемы Эль-Гамаля на эллиптических кривых заключается в том, что при сопоставимом уровне криптостойкости алгоритм требует существенно меньших размеров ключей по сравнению с классическими реализациями на основе дискретного логарифмирования в мультипликативной группе простого модуля или на основе RSA [3].

Уменьшенная длина ключей открывает ряд прикладных возможностей:

1. Устройства с ограниченными ресурсами (IoT, смарт-сенсоры). Сокращённый объём ключевых материалов облегчает хранение и управление миллионами ключей в распределённых системах.
2. Мобильные приложения и тонкие клиенты. Компактные ключи упрощают резервирование и синхронизацию ключевых данных в ограниченных пользовательских хранилищах и облачных бэкапах.
3. Протоколы с ограничением размера сообщений. Встроенные системы, блокчейн-транзакции и другие протоколы с жёсткими лимитами на размер сообщений выигрывают за счёт уменьшения длины открытых ключей и подписей [4].
4. Хранение ключей в облаке. При массовом хранении и репликации ключей меньшие размеры существенно экономят место и снижают эксплуатационные расходы.

Таким образом, уменьшение длины ключей обеспечивает не только более компактное хранение и передачу данных, но и расширяет применимость асимметричных схем в условиях, где ранее использование криптографии было затруднено из-за ограничений по объёму данных.

Заключение

Представленная схема шифрования является вариантом адаптации алгоритма Эль-Гамаля на эллиптических кривых.

Ключевая особенность данной схемы заключается в том, что её безопасность опирается на задачу дискретного логарифмирования на эллиптических кривых, которая в настоящее время считается более трудной по сравнению с классическим дискретным логарифмированием. Это позволяет достигать сопоставимой стойкости при меньших размерах ключей.

Практическая значимость метода определяется именно сокращением длины ключей, что упрощает хранение, распространение и использование криптографических данных в прикладных системах, где критичны ограничения по размеру ключей и сообщений.

1. Paar, C. Understanding Cryptography: From Established Symmetric and Asymmetric Ciphers to Post-Quantum Algorithms / C. Paar, J. Pelzl, T. Güneysu. – Cham : Springer, 2024. – 543 p.
2. Государственный стандарт Российской Федерации. ГОСТ Р 34.10-2012. Информационная технология. Криптографическая защита информации. Процессы формирования и проверки электронной цифровой подписи / – М. : Стандартинформ, 2012. – 36 с.
3. National Institute of Standards and Technology. Recommendation for Key Management, Part 1. / – Gaithersburg, MD : NIST, 2020. – 142 p.
4. Jiang, S. Key-and-Signature Compact Multi-Signatures for Blockchain: A Compiler with Realizations / S. Jiang, D. Alhadidi, H. F. Khojir. // Cryptology ePrint Archive, Paper 2023/061. – 2023. – 35 p.

СРАВНЕНИЕ СУБД ДЛЯ ХРАНЕНИЯ СТРУКТУР ДРЕВОВИДНЫХ ДАННЫХ

Мельник А. А.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: a.melnik@bsuir.by

В статье рассматривается проблема эффективного хранения древовидных структур данных. Проводится сравнительный анализ методов решения этой задачи в реляционных (PostgreSQL) и нереляционных (MongoDB, Neo4j) базах данных. На основе практических замеров скорости выполнения запросов делается вывод, что графовая СУБД Neo4j является наиболее производительным решением благодаря своей природе, идеально соответствующей работе со связанными данными.

Введение

Современные информационные системы повсеместно используют древовидные структуры для организации иерархических данных. Эти структуры позволяют эффективно представлять отношения вложенности в таких областях, как бухгалтерский учет, управление организационными структурами, построение категорий товаров и социальных графов. С увеличением сложности и объемов данных задача оптимального хранения и обработки древовидных структур становится особенно актуальной.

Традиционные реляционные подходы к хранению иерархических данных часто сталкиваются с проблемами производительности при работе с глубокими древовидными структурами. Это стимулирует поиск альтернативных решений среди нереляционных баз данных, которые предлагают специализированные модели хранения и эффективные алгоритмы обхода. Каждый из подходов имеет свои преимущества и ограничения, требующие тщательного анализа.

Проведение сравнительного исследования различных методов хранения древовидных структур представляется важной задачей. Такой анализ должен охватывать как реляционные, так и нереляционные подходы и учитывать требования современных приложений к производительности и масштабируемости. Результаты исследования позволят выработать практические рекомендации по выбору оптимальной стратегии хранения иерархических данных.

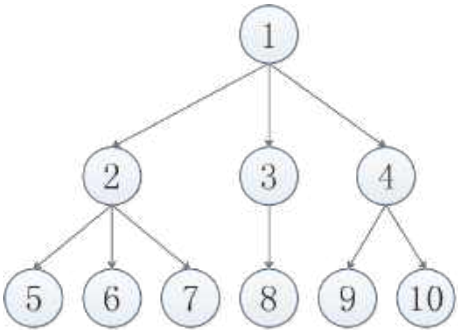


Рис. 1 – Пример дерева

I. Хранение в реляционных БД

Классические реляционные подходы могут сталкиваться с проблемами производительности при работе с глубокими иерархиями из-за необходимости использования множественных операций JOIN, что особенно заметно в системах без эффективного горизонтального масштабирования [1].

В реляционных БД можно выделить четыре способа хранения древовидной структуры данных: Список смежности (Adjacency List), Таблица связей (Closure Table), Вложенные множества (Nested sets), Материализованный путь (Materialized Path).

Список смежности позволяет хранить родительский узел для каждой вершины дерева. Для этого в таблицу добавляется еще одно поле, в котором хранится id родителя. Основная идея способа «Таблица связей» состоит в создании дополнительной таблицы для хранения id родителя и потомка. Основной недостаток состоит в том, что чем глубже в дереве находится узел, тем больше записей в таблице потребуется для описания связей, однако это позволяет обеспечить быстрый поиск. Материализованный путь позволяет хранить полный путь от корня дерева до нужного узла. Чтобы найти все дочерние элементы определенного узла, необходимо использовать LIKE с %, который работает медленнее, чем =, но в большинстве случаев этого достаточно. В схеме вложенных множеств каждый узел имеет присвоенные левые и правые значения, указывающие на его место в дереве.

II. Хранение в нереляционных БД

В нереляционных БД чаще всего используют: Array of Ancestors – сохраняет ссылки на «родительские» узлы и массив, в котором хранятся все предки [2], Materialized Path, Parent References – организует документы в древовидную структуру путем хранения ссылок на «родительские» узлы в «дочерних» узлах, Child References – организует документы в древовидную структуру путём

хранения ссылок на «дочерние» узлы в «родительских» узлах., Nested Sets.

Примерами систем управления нереляционными БД являются MongoDB и Neo4j. Neo4j – это графовая система управления базами данных с открытым исходным кодом, реализованная на Java. Она является ведущей графовой СУБД в мире. Для базы данных Neo4j был создан декларативный язык запросов Cypher, обеспечивающий эффективное чтение и запись данных в Neo4j. MongoDB – система управления базами данных, которая работает с документоориентированной моделью данных. В отличие от реляционных СУБД, MongoDB использует гибкую документо-ориентированную модель, не требующую строгой схемы, и обладает собственным богатым языком запросов.

III. Практическое исследование

В результате замеров можно увидеть, что наиболее эффективной и удобной СУБД для хранения древовидных данных является Neo4j – графовая СУБД. Это обусловлено тем, что объекты в такой базе данных связаны между собой как узлы в графах [3]. MongoDB же является документоориентированной базой данных, которая не была специально разработана для работы с древовидными структурами данных [4].

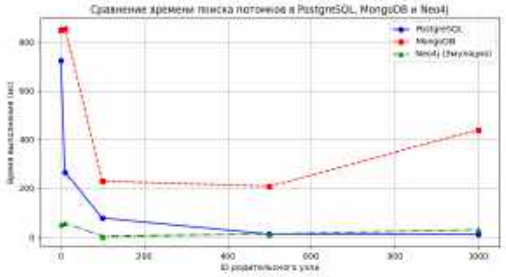


Рис. 2 – Результат замера скорости выполнения запросов поиска в PostgreSQL, MongoDB, Neo4j

Анализ результатов производительности выявил четкую закономерность: с увеличением глубины дерева и сложности запросов преимущество Neo4j становится более выраженным. В то время как реляционные подходы в PostgreSQL демонстрировали прогрессивное замедление при обработке многоуровневых иерархий [5], а MongoDB показывала средние результаты из-за необходимости выполнения дополнительных операций агрегации, графовая модель Neo4j обеспечивала стабильно высокую скорость выполнения даже для сложных рекурсивных запросов. Это подтверждает теоретическое положение о том, что специализированные графовые базы данных наиболее адекватно отвечают требованиям работы с древовидными структурами.

Заключение

Проведенное исследование позволяет сделать однозначный вывод о том, что выбор системы управления базами данных для работы с древовидными структурами напрямую определяет производительность и масштабируемость приложения.

Сравнительный анализ реляционных и нереляционных подходов показал, что классические реляционные СУБД, такие как PostgreSQL, несмотря на множество схем организации иерархических данных (список смежности, вложенные множества и др.), сталкиваются с фундаментальными ограничениями. К ним относятся низкая производительность операций JOIN, сложности с горизонтальным масштабированием и высокая нагрузка на сервер при работе с глубокими иерархиями.

Практические замеры скорости выполнения запросов подтвердили, что специализированные графовые базы данных демонстрируют превосходство в данной задаче. Neo4j показала себя как наиболее эффективное решение благодаря своей нативной модели данных, где связи между узлами являются объектами первого класса. Это позволяет ей напрямую и быстро обходить древовидные структуры, избегая дорогостоящих операций, присущих реляционным подходам.

Таким образом, для проектов, где работа с иерархическими и связанными данными является ключевой, графовые СУБД, и в частности Neo4j, следует рассматривать как оптимальный выбор, обеспечивающий высокую производительность и удобство разработки. В то время как документоориентированные СУБД, подобные MongoDB, предлагают гибкость, их использование для сложных древовидных структур может быть менее эффективным по сравнению со специализированными графовыми системами.

1. Maddipatla S. P. et al. Using databases to implement algorithms: Estimation of allan variance using b+-tree data structure //2024 American Control Conference (ACC). – IEEE, 2024. – С. 2164-2169.
2. Pan J. J., Wang J., Li G. Survey of vector database management systems // The VLDB Journal. – 2024. – Т. 33. – № 5. – С. 1591-1615.
3. Hodler A. E., Needham M. Graph data science using Neo4j //Massive Graph Analytics. – Chapman and Hall/CRC, 2022. – С. 433-457.
4. Dhanagari M. R. MongoDB and data consistency: Bridging the gap between performance and reliability // Journal of Computer Science and Technology Studies. – 2024. – Т. 6. – №. 2. – С. 183-198.
5. Makris A. et al. MongoDB Vs PostgreSQL: A comparative study on performance aspects // GeoInformatica. – 2021. – Т. 25. – № 2. – С. 243-268.

ИНФОРМАЦИОННАЯ СИСТЕМА ДЛЯ СБОРА, АНАЛИЗА И ВИЗУАЛИЗАЦИИ ДАННЫХ О НАБЛЮДЕНИЯХ ЗА БЕЗОПАСНОСТЬЮ НА МОРСКИХ ОБЪЕКТАХ НЕФТЕГАЗОВОЙ ИНДУСТРИИ

Нестеренков С. Н., Иванов А. А.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by, andreyvorobey.play@gmail.com

Разработана концепция информационной системы для сбора, анализа и визуализации данных, связанных с наблюдениями за безопасностью на морских объектах нефтегазовой индустрии. Система предназначена для автоматизации процесса регистрации наблюдений, выявления отклонений, проведения статистического анализа и представления информации в виде наглядных интерактивных визуализаций. Рассмотрены архитектура, функциональные модули, методы интеллектуального анализа и подходы к обеспечению надежности и защищённости информации в распределённых условиях эксплуатации.

Введение

Современная нефтегазовая индустрия характеризуется высокой степенью технологической сложности и повышенным уровнем риска при эксплуатации морских объектов. Безопасность персонала и предотвращение аварийных ситуаций требуют систематического сбора и анализа информации о потенциальных угрозах, происшествиях и наблюдениях за поведением персонала в производственной среде [1].

В последние годы большое внимание уделяется созданию цифровых платформ, которые позволяют автоматизировать процессы мониторинга и анализа данных о безопасности. Однако большинство существующих решений ориентированы на отдельные аспекты промышленной безопасности и не обеспечивают комплексной поддержки жизненного цикла наблюдений – от регистрации до аналитической обработки и визуализации.

Целью данной работы является разработка архитектуры и алгоритмов функционирования информационной системы для сбора, анализа и визуализации данных о наблюдениях за безопасностью на морских объектах нефтегазовой отрасли.

I. Постановка задачи

Основными задачами, решаемыми проектируемой системой, являются:

- организация централизованного сбора и хранения данных наблюдений, поступающих от сотрудников и автоматических сенсорных устройств;
- реализация инструментов интеллектуального анализа данных с целью выявления закономерностей и потенциальных рисков;
- визуализация информации в интерактивной форме принятия управленческих решений;
- обеспечение информационной безопасности и отказоустойчивости системы при эксплуатации в распределённой сетевой среде.

Система должна функционировать в условиях ограниченного канального ресурса и повышенных требований к защищённости данных. Кроме того, необходимо обеспечить возможность интеграции с корпоративными системами управления производством (ERP, SCADA, MES и др.), что позволит повысить уровень прозрачности производственных процессов и своевременно выявлять отклонения.

II. Архитектура и функциональные модули системы

Разработанная архитектура включает три основных уровня: уровень сбора данных, уровень аналитической обработки и уровень представления результатов пользователю.

На уровне сбора данных осуществляется агрегация информации из различных источников: отчётов операторов, журналов событий, датчиков вибрации, температуры, давления и внешних баз данных. Для унификации форматов используется модуль предварительной обработки, выполняющий очистку, нормализацию и временную синхронизацию данных.

Аналитический уровень реализует алгоритмы статистического и интеллектуального анализа. В частности, применяется кластеризация инцидентов по типам, выявление тенденций с использованием временных рядов, а также оценка факторов риска на основе многомерных моделей [2]. Для повышения точности прогнозирования используется ансамблевый подход, сочетающий регрессионные методы и алгоритмы машинного обучения (градиентный бустинг, деревья решений, случайный лес). Эти методы позволяют оценивать вероятность возникновения событий и предлагать профилактические меры.

Уровень визуализации предоставляет интерфейсы для отображения карт инцидентов, динамических графиков и дашбордов. Визуальные средства реализованы с применением веб-

технологий (JavaScript, D3.js, Plotly), что позволяет использовать систему как на локальных терминалах, так и в облачной среде. Пользователь может настраивать набор отображаемых показателей и фильтров в зависимости от категории объекта, временного диапазона и уровня допуска.

III. Анализ и визуализация данных

Особое внимание уделено модулю визуализации, который обеспечивает наглядное представление результатов анализа. Используются интерактивные графические элементы: тепловые карты распределения событий, диаграммы трендов, сетевые графы взаимосвязей между объектами и факторами риска.

Для повышения информативности визуализации реализованы функции фильтрации данных по временным и географическим признакам, а также возможность формирования отчётов в форматах PDF, Excel и HTML.

На этапе анализа данных реализованы методы корреляционного анализа и факторного моделирования, позволяющие определять наиболее значимые причины нарушений безопасности. Для выявления аномалий используется алгоритм локальной факторной плотности (LOF), что позволяет обнаруживать редкие, но критичные события [3].

Перспективным направлением является внедрение методов глубокого обучения для автоматической классификации инцидентов по текстовым и визуальным описаниям. Применение сверточных и рекуррентных нейронных сетей позволит повысить точность распознавания типов событий, а также выявлять скрытые закономерности, недоступные при традиционном анализе. Кроме того, предусмотрена возможность интеграции системы с внешними источниками данных – например, метеорологическими сервисами или системами контроля судоходства, что позволяет учитывать внешние факторы при моделировании риска.

Кроме того, система поддерживает механизм прогнозирования показателей безопасности на основе моделей временных рядов ARIMA и Prophet. Это обеспечивает возможность раннего предупреждения потенциальных инцидентов и оценки эффективности проводимых мероприятий.

Отдельное внимание уделено вопросам обеспечения надёжности и защиты данных. Используется механизм ролевого доступа, шифрование на уровне транспортного протокола (TLS 1.3) и резервное копирование с использованием распределённого хранилища. Для повышения устойчивости система разворачивается в контейнерной инфраструктуре (Docker, Kubernetes), что позволяет гибко масштабировать вычислительные ресурсы.

Заключение

Практическая реализация предложенного подхода открывает возможности для создания единого цифрового пространства промышленной безопасности, объединяющего данные различных предприятий и объектов морской инфраструктуры. Такое пространство позволит обеспечить более высокую степень прозрачности производственных процессов, упростить обмен аналитическими данными и стандартизировать процедуры регистрации наблюдений и оценки рисков. В перспективе планируется развитие интеллектуальных подсистем самообучения, которые смогут адаптировать алгоритмы анализа под особенности конкретных производственных процессов. Эти подсистемы будут использовать механизмы непрерывного накопления опыта и корректировки моделей на основе поступающих данных, что обеспечит повышение точности прогнозов и снижение вероятности ложных срабатываний. Это позволит перейти от реактивного к превентивному управлению рисками и существенно повысить общую устойчивость нефтегазового комплекса. Кроме того, внедрение системы в рамках корпоративных стандартов HSE (Health, Safety, Environment) создаст основу для интеграции с международными инициативами в области устойчивого развития и цифровой трансформации отрасли. В долгосрочной перспективе подобные решения могут стать основой для построения международных центров аналитики безопасности, способных координировать действия различных операторов в режиме реального времени, формируя единую экосистему управления рисками в морской нефтегазовой индустрии.

1. Дмитриев, В. И. Система управления безопасностью судоходных компаний и судов / В. И. Дмитриев. – 2023. – 271 с.
2. Alhadi, D. A. Redesigning Online HSE Observation Card: Improving Adoption Rate & Reducing Error Rate — A UX Case Study / Dony A. Alhadi. – 2023. – Дата публикации: 3 июня 2023 г.
3. Ramon Meris. Conducting Safety Observations for Improving Workplace Safety Management – Ramon Meris – 2023. – Дата публикации: 25 сентября 2025 г.
4. Зиновьева, А. А. Количественные методы оценки рисков нефтегазовых предприятий / А. А. Зиновьева // Молодёжь и наука: материалы междунар. науч.-практ. конф. старшеклассников, студентов и аспирантов (27 мая 2022 г.): в 2 т. – Нижний Тагил: НТИ (филиал) УрФУ, 2022. – Т. 2. – С. 442–444.
5. Голубович, Ю. И. Искусственный интеллект в управлении рисками в финансовой сфере / Ю. И. Голубович, С. Н. Нестеренков, С. А. Байчик // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics: сб. науч. ст. X Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13 марта 2024 г.). В 2 ч. Ч. 1 / редкол.: В. А. Богуш [и др.]. – Минск: БГУИР, 2024. – С. 345–350.

АНАЛИЗ ТЕКСТОВОГО СОДЕРЖИМОГО ГЛАВНОЙ СТРАНИЦЫ ВЕБ-САЙТА ДЛЯ ОПРЕДЕЛЕНИЯ ЕГО ТЕМАТИКИ ПРИ ПОМОЩИ LLM

Нестеренков С. Н., Лазук И. С.
Кафедра программного обеспечения информационных технологий,
кафедра электронных вычислительных машин,
Белорусский Государственный Университет Информатики и Радиоэлектроники
Минск, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by, i.lazuk@bk.ru

Работа посвящена задаче автоматического определения тематики веб-сайтов по тексту их главной страницы. Предложена методика, включающая сбор HTML-контента, извлечение основного текста, предобработку, применение больших языковых моделей (LLM) в режимах zero/few-shot и приёмы стабилизации ответов. Представлены экспериментально-гипотетические результаты сравнения различных LLM (GPT-4 от OpenAI, YandexGPT, DeepSeek) и режимов их использования на главных страницах веб-ресурсов. Обсуждаются ограничения, связанные с вёрсткой, мультимедийностью, использованием нескольких языков, предлагаются направления дальнейших исследований.

Введение

Автоматическое определение тематики веб-страниц является важной прикладной задачей для поисковых систем, рекомендательных сервисов, каталогов, мониторинга СМИ, рекламных платформ и систем модерации. Точная категоризация главной страницы сайта ускоряет индексацию, повышает релевантность поиска, управляет показом рекламы, выявляет нежелательный контент и приоритезирует обход ресурсов. В условиях стремительного роста и высокой динамичности веб-ресурсов ручная разметка тематики непрактична [1].

Большие языковые модели (LLM) продемонстрировали способность к обобщению и эффективному дообучению, что делает их перспективным инструментом для семантического анализа веб-контента и решения задач классификации в zero- и few-shot режимах [2,3]. Помимо прямой генеративной классификации, LLM поддерживают вспомогательные стратегии рассуждений и самосогласованности, повышающие устойчивость и объяснимость результата [3]. В работе описывается методика определения тематики сайта по тексту главной страницы с использованием LLM и проводится сравнительный анализ различных моделей и режимов их применения (GPT-4, YandexGPT, DeepSeek) [4–6].

I. Методы сбора и извлечения текста

Использованный конвейер обработки включает в себя этапы сбора, извлечения информации, предобработки и классификации.

Главная страница загружается по протоколу HTTP(S) с учётом `robots.txt`. Для сайтов с клиентской отрисовкой используется браузер для исполнения JavaScript и получения финального дерева DOM. Извлечение основного текста страницы выполняется с помощью различных эвристик выделения текста из HTML. Дополнительно выделяются заголовок страницы, теги `<h1>/<h2>` и текст якорей ключевых ссылок.

Предобработка включает нормализацию кодировок текста, удаление скриптов и стилей, декодирование HTML-сущностей, разбиение на абзацы и предложения, устранение лишних пробелов, дедупликацию строк. Определяется язык текста и доли языков при смешении: для многоязычных страниц применяется взвешивание сегментов по вероятности языка.

Подходы к использованию LLM:

- Прямая генеративная классификация по заранее заданному набору тематических категорий в zero-shot режиме с предписанным форматом ответа (метка, краткое обоснование, оценка уверенности) [2,3].
- Few-shot режим с 1–3 типовыми примерами на класс для адаптации подсказки [2].
- Двухшаговая схема «резюме -> классификация» для длинных страниц: сначала краткое содержательное резюме скачанной страницы, затем классификация по полученному резюме.
- Самосогласованность: несколько независимых прогонов с агрегацией ответов для снижения вариативности [3].

Набор тематических категорий из 20 направлений: новости и СМИ, электронная коммерция, финансы, образование, медицина и здоровье, развлечения, технологии, путешествия и туризм и др. Качество оценивается точностью предсказанной категории. Дополнительно анализируются медианное время отклика модели LLM на одну страницу.

Набор данных: 100 главных страниц русскоязычных и англоязычных сайтов ручной разметки (баланс по категориям, коммерческие и некоммерческие домены). Страницы с преобладанием мультимедиа без текста исключены, но потенци-

альное направление развития для таких страниц отмечены в анализе результатов.

II. Результаты исследования

Сравнивались различные современные LLM и режимы их применения. Рассматривались GPT-4 (OpenAI), YandexGPT и DeepSeek. Для каждой модели оценивались zero-shot и few-shot режимы (3 примера на класс) и измерялось время отклика.

Таблица 1 – Сравнение LLM по определению тематики главной страницы

Модель	Точность, zero-shot	Точность, few-shot (3 примера)	Медианное время отклика, с
GPT-4 (OpenAI)	0.88	0.92	1.7
YandexGPT	0.84	0.88	1.5
DeepSeek	0.85	0.89	1.6

Ключевые наблюдения:

- Few-shot режим стабильно повышает точность на 3–5 п.п. относительно zero-shot. Эффект особенно заметен для редких тематических категорий и страниц с малым содержанием текста [2].
- Различия между моделями объясняются качеством предоставленной инструкции для модели и объёмом контекста: модель класса GPT-4 демонстрируют наилучшее сочетание точности.
- Самосогласованность даёт дополнительный прирост 0.8–1.5 п.п. точности при увеличении времени отклика примерно в 2–3 раза [3].
- Предварительное резюмирование длинных страниц снижает долю ошибок на 1.0–1.6 п.п. за счёт подавления навигационного и рекламного шума.

III. Анализ результатов и возможные улучшения

В ходе проведения исследования были выделены следующие ограничения:

- Зависимость от вёрстки и динамической отрисовки: баннеры, всплывающие окна и персонализация могут искажать основной текст – необходимы устойчивые процедуры извлечения контента.
- Мультимедийность и малотекстовые страницы: сайты, где смысл задаётся изображениями/видео, хуже поддаются чисто текстовому анализу – перспективны использование VLM моделей.
- Многоязычие и смешение языков: на мультиязычных сайтах снижается качество – многоязычные конфигурации помогают, но не устраняют проблему полностью.
- Стоимость и время отклика: LLM требуют больше ресурсов – целесообразны каскады с ранней остановкой.
- Объяснимость и воспроизводимость: генеративные ответы вариативны – полезны фиксированные шаблоны вывода, понижение температуры и строгое протоколирование подсказок.

Для улучшения результатов перспективно выглядят мультимодальные LLM с учётом изображений, активное обучение для краевых случаев, а также тонкая настройка инструкций (промптов) под разные домены. Переносимость решений на многоязычные и низкоресурсные домены требует исследований устойчивости к языковому смещению.

Заключение

Проведённый анализ показывает, что большие языковые модели обеспечивают высокое качество определения тематики веб-сайтов по тексту главной страницы. На срезе из 100 страниц наблюдается устойчивое преимущество few-shot над zero-shot режимом и заметные различия между моделями (GPT-4, YandexGPT, DeepSeek) в зависимости от масштаба и качества инструктажной донастройки. Наилучшим образом себя показывает модель GPT-4 от OpenAI. Ключевыми факторами эффективности являются устойчивое извлечение основного текста, корректная предобработка, проектирование подсказок и осмысленная комбинация режимов LLM (резюмирование, самосогласованность). Дальнейшая работа связана с мультимодальными расширениями и иерархическими наборами категорий.

IV. Список литературы

1. Баяк, Е. И. Автоматизация ответов службы технической поддержки с использованием нейронных сетей на примере учреждения высшего образования / Е. И. Баяк, В. А. Быстрова, С. Н. Нестеренков // Информационные технологии и системы 2024 (ИТС 2024) = Information Technologies and Systems 2024 (ITS 2024) : материалы международной научной конференции, Минск, 20 ноября 2024 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2024. – С. 143–144.
2. Kojima, T. Large language models are zero-shot reasoners / T. Kojima, S. S. Saha, R. Schubert, Y. Li, J. Cho // Advances in Neural Information Processing Systems (NeurIPS). – 2022. – Vol. 35. – P. 22199–22213.
3. Wang, X. Self-consistency improves chain of thought reasoning in language models / X. Wang, J. Wei, D. Schuurmans, Q. Le [и др.] // Proc. of the International Conference on Learning Representations (ICLR). – 2023.
4. OpenAI. GPT-4 Technical Report [Электронный ресурс] / OpenAI Research. – 2023. – Режим доступа: https://arxiv.org/abs/2303.08774. – Дата доступа: 10.10.2025.
5. DeepSeek. DeepSeek-V2: A Strong, Efficient Mixture-of-Experts Language Model [Электронный ресурс] / DeepSeek-AI Research. – 2024. – Режим доступа: https://arxiv.org/abs/2405.04434. – Дата доступа: 10.10.2025.
6. Yandex Cloud. YandexGPT: документация по API [Электронный ресурс]. – 2025. – Режим доступа: https://yandex.cloud/ru/services/yandexgpt. – Дата доступа: 10.10.2025.

АНАЛИЗ МЕТОДОВ РЕШЕНИЯ ЗАДАЧИ РАЗМЕЩЕНИЯ КАМЕР НАБЛЮДЕНИЯ В ПОМЕЩЕНИЯХ С ПЕРЕГОРОДКАМИ

Новыш И. М., Герман Ю. О.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: inovysh@gmail.com, jgerman@bsuir.by

В данной работе рассматриваются методы решения задачи оптимального размещения камер видеонаблюдения в помещениях со сложной планировкой и перегородками. Цель исследования заключается в сравнении эффективности различных алгоритмов оптимизации для максимального покрытия наблюдаемой области при минимальном количестве камер.

Введение

Задача оптимального размещения камер наблюдения относится к классу NP-трудных задач комбинаторной оптимизации. Актуальность исследования обусловлена ростом интереса к автоматизации систем видеонаблюдения в целях обеспечения безопасности и интеллектуального мониторинга объектов. Проблема оптимального размещения камер является ключевой при проектировании систем наблюдения в помещениях со сложной архитектурой и наличием перегородок, так как напрямую влияет на эффективность системы, экономию оборудования и вычислительных ресурсов. Практическая значимость работы заключается в возможности адаптации рассмотренных алгоритмов для решения задач в робототехнических системах, системах «умного дома» и распределённых сетях датчиков [1, 2].

I. Рассматриваемые методы и алгоритмы для решения задачи о покрытии

Для решения задачи оптимального размещения камер видеонаблюдения в помещениях с перегородками в работе были рассмотрены различные алгоритмы оптимизации, отличающиеся как принципами построения, так и характером поиска решений. Их выбор обусловлен стремлением охватить как детерминированные, так и стохастические подходы, что позволяет провести более полное сравнение эффективности методов в различных условиях.

Одним из базовых подходов является жадный алгоритм [3], который реализует пошаговую стратегию выбора позиций камер. На каждом этапе алгоритм добавляет ту камеру, которая обеспечивает максимальное дополнительное покрытие неохваченных областей помещения. Такой метод прост в реализации, обладает высокой скоростью работы и хорошо масштабируется на небольшие задачи. Однако его недостаток заключается в том, что он не учитывает глобальные зависимости между выбранными позициями, из-за чего может приводить к локально оптимальным, но не глобально наилучшим решениям.

В противоположность жадным методам, алгоритм муравьиной колонии [4] представляет собой стохастическую метаэвристику, вдохновлённую коллективным поведением насекомых. Он моделирует процесс поиска пути муравьями, при котором особи оставляют феромонные следы, отражающие успешность найденных решений. С течением итераций феромоны на более перспективных маршрутах усиливаются, направляя поиск к потенциально лучшим вариантам размещения камер. Достоинством данного подхода является способность избегать застревания в локальных минимумах, однако он требует значительных вычислительных ресурсов и тщательной настройки параметров, таких как интенсивность феромона и коэффициент испарения. В противном случае алгоритм может сходиться слишком медленно или не обеспечивать полного покрытия наблюдаемой области.

Другим мощным стохастическим методом является имитация отжига. Данный алгоритм основан на физической аналогии процесса кристаллизации металлов, где система постепенно охлаждается, переходя из хаотического состояния к более устойчивому. В контексте рассматриваемой задачи это означает постепенное «улучшение» конфигурации размещения камер, при котором на ранних этапах допускаются менее выгодные перемещения, что помогает избежать локальных минимумов, а затем — постепенное «замораживание» поиска для стабилизации найденного решения. Метод имитации отжига часто показывает высокую устойчивость и гибкость при увеличении размера сетки или усложнении конфигурации перегородок, хотя требует подбора оптимального режима охлаждения и числа итераций.

Отдельный интерес представляет метод групповых резолюций [5], использующий идею коллективного анализа множественных локальных решений. Его суть заключается в разделении исходной задачи на подзадачи и их последовательном решении с последующим объединением результатов. Такой подход позволяет повысить вероятность нахождения глобального оптимума,

особенно при высокой размерности задачи. Метод хорошо сочетается с другими эвристиками и может использоваться как дополнительный этап уточнения решений, полученных стохастическими алгоритмами.

Наконец, генетический алгоритм [6] опирается на принципы естественного отбора и генетической эволюции. Каждое возможное размещение камер рассматривается как особь, а совокупность решений как популяция. С помощью операций кроссовера, мутации и отбора происходит постепенное улучшение качества решений в каждом поколении. Этот метод обладает высокой универсальностью и способен находить близкие к оптимальным решения даже в задачах со сложными ограничениями. Однако эффективность алгоритма во многом зависит от параметров — размера популяции, вероятности мутации и стратегии селекции, – что требует предварительной настройки или адаптивных механизмов.

Таким образом, рассмотренные методы охватывают широкий спектр подходов – от быстрых детерминированных алгоритмов до гибких стохастических метаэвристик. Их совместное использование и комбинирование в гибридных схемах позволяет повысить качество решений при сохранении приемлемого времени вычислений.

II. Сравнительный анализ алгоритмов для решения задачи

Проведённые вычислительные эксперименты показали, что эффективность алгоритмов значительно зависит от размерности сетки и конфигурации перегородок. Жадный алгоритм продемонстрировал минимальное время решения задачи и обеспечил до 95–98 % покрытия наблюдаемой области, что делает его эффективным инструментом при решении задач малой размерности [3]. Алгоритм муравьиной колонии [4] в ряде экспериментов не обеспечивал полного покрытия сетки или требовал значительно большего времени для сходимости, что ограничивает его применение при больших объёмах данных. Метод имитации отжига [5] показал результаты, сопоставимые с жадным алгоритмом, однако на сетках большего размера проявил более высокую устойчивость и лучшее качество покрытия. Генетический алгоритм обеспечил близкое к оптимальному покрытие, но его эффективность во многом зависела от настройки параметров мутации и кроссовера [6]. Метод групповых резолюций позволил повысить качество решения за счёт объединения результатов нескольких локальных поисков, но также требовал значительных вычислительных затрат.

Таким образом, выбор метода зависит от характера задачи: для небольших сеток оптимален жадный алгоритм, а для задач с большим числом преград и увеличенной размерностью предпочтителен метод имитации отжига.

Ниже приведена сравнительная таблица с результатами проведенных опытов.

Таблица 1 – Работа алгоритмов для сетки 30x30

Алгоритм	Число камер	Время работы,с	Покрытие, %
Жадный	111	19.2	100
Генетический	111	1222.2	100
МГР	152	724.4	100
Имитация отжига	111	20.8	100
Метод муравья	92	1090.7	82

Заключение

Полученные результаты показывают, что жадный алгоритм является наиболее эффективным по времени выполнения при приемлемом качестве покрытия, в то время как метод имитации отжига обеспечивают более устойчивое поведение при увеличении сложности задачи. Алгоритм муравьиной колонии, несмотря на способность избегать локальных минимумов, требует длительного времени вычислений.

Перспективным направлением дальнейших исследований является разработка гибридных алгоритмов, комбинирующих преимущества детерминированных и вероятностных подходов [7], а также применение методов машинного обучения для адаптивного выбора стратегии размещения камер в зависимости от геометрии помещения и ограничений вычислительных ресурсов.

1. Зайченко, Ю. П. Современные методы оптимизации. – Киев: Слово, 2019. – С. 15-20.
2. Blum, C., Roli, A. Metaheuristics in Combinatorial Optimization: Overview and Conceptual Comparison. ACM Computing Surveys, 2003. – p. 268–276.
3. Cormen, T. H. et al. Introduction to Algorithms. MIT Press, 2022. – p. 116-120.
4. Dorigo, M., Stützle, T. Ant Colony Optimization. MIT Press, 2004. – p. 43-65.
5. Баранов, Г. А. Метод групповых учётов и резолюций в задачах комбинаторной оптимизации. Вестник МГТУ, 2018. – №4. – С. 98-104.
6. Holland, J. H. Adaptation in Natural and Artificial Systems. University of Michigan Press, 1975. – p. 45–67.
7. Eiben, A. E., Smith, J. E. Introduction to Evolutionary Computing. Springer, 2015. – p. 89–95.

КРИПТОГРАФИЧЕСКИЙ ФУНКЦИОНАЛ МЕССЕНДЖЕРА PEREGRINE НА БАЗЕ НАЦИОНАЛЬНЫХ СТАНДАРТОВ ШИФРОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Петров С. Н., Пакуль Е. С., Корчинский А. А.
Кафедра защиты информации,
Белорусский государственный университет информатики и радиоэлектороники
Национальный детский технопарк
Минск, Республика Беларусь
E-mail: petrov@bsuir.by

Разработан мессенджер Peregrine на языке программирования Rust, имеющий клиент-серверную архитектуру, в которой клиенты (устройства пользователей) взаимодействуют с центральным сервером, пересылающим сообщения от отправителя к получателю, выполняющим аутентификацию пользователей и управление доступом. Центральным звеном стала разработка библиотеки bee2-rs на языке программирования Rust. Библиотека создана для вызовов функций библиотеки Bee2 (от официального разработчика), написанной на языке программирования C. Приведено частичное описание криптографического функционала мессенджера.

Введение

Приложения для обмена мгновенными сообщениями (мессенджеры) стали важной частью современной жизни. Мессенджеры стали универсальным инструментом для коммуникации и совместной работы. Однако в связи с актуализацией проблематики обеспечения безопасности данных и возможными ограничениями (например, запретом на использование) становятся актуальными вопросы разработки альтернативных, безопасных мессенджеров.

Конфиденциальность информации является одной из уязвимых областей в системах мгновенных сообщений. Информация, передаваемая через системы обмена сообщениями, проходит через несколько точек раскрытия до достижения получателя. Сообщения проходят через серверы и сети, которые могут быть небезопасными, такк как нет гарантии того, что отправленная информация защищена. Ситуация усложняется тем, что службы обмена мгновенными сообщениями часто содержат программы обмена файлами, которые могут оставлять конфиденциальную информацию на сервере после завершения сеанса. Серверная инфраструктура систем обмена мгновенными сообщениями является критическим компонентом, который также подвержен различным типам атак, например, атак типа «отказ в обслуживании», которые могут нарушить работу серверов. Такие атаки часто осуществляются путем перегрузки серверов запросами, что приводит к исчерпанию ресурсов и неспособности обрабатывать законные запросы. Другой формой атаки на серверную инфраструктуру является несанкционированный доступ к серверам, где хранятся сообщения и метаданные пользователей, что может привести к утечке конфиденциальной информации. Особую опасность представляют атаки на централизованные серверы, где хранятся ключи шифрования или где происходит обработка сообщений перед их отправкой получателям.

I. Мессенджер Peregrine

Авторами разработан мессенджер с защитой пользовательских данных и end-to-end (сквозным) шифрованием на основе национальных криптографических стандартов. Мессенджер имеет клиент-серверную архитектуру, в которой клиенты (устройства пользователей) взаимодействуют с центральным сервером, пересылающим сообщения от отправителя к получателю, выполняющим аутентификацию пользователей и управление доступом. Определены функциональные возможности серверной и пользовательских частей. Мессенджер назван Peregrine.

Актуальность разработки нового мессенджера обусловлена следующими факторами:

- рост спроса на приватность и активный переход на альтернативные платформы после скандалов с утечками данных пользователей WhatsApp;
- доминирование Telegram и Viber создает зависимость от зарубежных сервисов, что повышает риски санкционных ограничений;
- государственные и коммерческие организации заинтересованы в решениях с серверами внутри страны.

Основные компоненты системы:

- Клиентское приложение (коммуникация с сервером происходит по API с использованием HTTP POST запросов).
- Сервер, выступаюший в роли посредника между клиентами, управляет аутентификацией пользователей, хранением данных в базе данных. Сервер использует протокол без сохранения состояния (stateless protocol),то есть общение с сервером состоит из независимых пар запрос-ответ. Код сервера состоит из модуля базы данных, модуля хранили-

ща данных, где хранятся файлы, отправленные пользователями, а также модуля API.
- База данных MySQL (хранение пользовательских данных, токенов сессии и т.п.).
- Балансировщик нагрузки.

II. Криптографический функционал мессенджера

Основной акцент в проекте сделан на безопасность. Центральным звеном стала разработка библиотеки bee2-rs на языке программирования Rust. Библиотека создана для вызовов функций библиотеки Bee2. Официальный разработчик Bee2– НИИ прикладных проблем математики и информатики БГУ [1]. В библиотеке реализованы алгоритмы и протоколы, определенные в стандартах СТБ 34.101.31, (belt): шифрование, имитозащита, хэширование, управление ключами; СТБ 34.101.45 (bign): ЭЦП и транспорт ключа (шифрование с открытым ключом); СТБ 34.101.47 (brng): имитозащита, ГПСЧ, одноразовые пароли (HOTP, TOTP, OCRA); СТБ 34.101.60 (bels): разделение секрета; СТБ 34.101.66 (bake): протоколы формирования общего ключа (BMQV, BSTS, BPACE);СТБ 34.101.77 (bash): хэширование). Однако библиотека Bee2 написана на языке C, а мессенджер на языке Rust и напрямую использовать Bee2 в проекте невозможно. Для решения данной проблемы использован механизм binding для языка Rust (Rust Foreign Function Interface, FFI).

Были сделаны абстракции нулевой стоимости для повышения удобства и уменьшения вероятности неправильного использования библиотеки (например, в некоторые функции нельзя передать данные неправильного размера, иначе программа не скомпилируется). Также для блочного шифра Belt было реализовано дополнение байтов в конец изначальных данных, чтобы длина результирующих зашифрованных данных была кратна n (по умолчанию – 16, $1 < n < 256$) для невозможности определить оригинальную длину сообщения по длине зашифрованных данных.

Для проверки правильности функций и соответствия их функциям библиотеки Bee2 были написаны unit-тесты. Для Bash было написано 3 теста по 52 проверок в каждом, для Belt – 7 тестов с общим числом проверок 76, для Bign – 4 теста счислом проверок 311, для Brng написано 3 теста с 23 проверками.

Реализован расширенный тройной протокол Диффи-Хеллмана (X3DH) для обмена ключами симметричного алгоритма шифрования. X3DH устанавливает общий секретный ключ между двумя сторонами, которые взаимно аутентифицируют друг друга на основе открытых ключей. Реализован режим belt-dwp алгоритма шифрования Belt (аутентификационное шифрование с ассоциированными данными). Для предотвращения кибератак типа bruteforce реализован следующий алгоритм аутентификации – клиент генерирует пару ключей для асимметричного алгоритма шифрования по паролю с использованием PBKDF2, после чего подписывает приватным ключом текущую временную метку (timestamp), а также период валидности сигнатуры до и после этого timestamp для принятия запроса даже в случае рассинхронизации времени между сервером и клиентом функции для протокола X3DH.

В структуре мессенджера был реализован функционал для использования криптографических стандартов отличных от национальных. На данный момент поддерживается только AES-GCM как альтернатива Belt. Открытый исходный код мессенджера позволяет добавить альтернативную библиотеку. При необходимости можно использовать разные криптографические библиотеки для разных чатов. В будущем планируется реализация поддержки следующих криптографических библиотек: chacha20poly1305, x25519-dalek, ed25519-dalek, k256, pbkdf2, sha3. У большинства из указанных ранее библиотек был проведен аудит безопасности [2], который либо не выявил серьезных уязвимостей, либо они были исправлены.

Исходный код разработанной библиотеки bee2-rs доступен на GitHub [3], а также опубликован в реестр библиотек (crates) сообщества Rust. Исходный код мессенджера был опубликован на платформае GitHub [4].

III. Заключение

Мессенджер Peregrine, использующий end-to-end шифрование на основе национальных стандартов, призван решить проблемы, обусловленные растущим спросом на приватность, стремлением снизить зависимость от зарубежных сервисов и необходимостью создания отечественных решений для государственных и коммерческих организаций. Проект успешно решает сложную задачу интеграции национальных криптографических стандартов в современную, безопасную среду Rust, делая основной акцент на предотвращении ошибок разработчика (через абстракции) и использовании надежных протоколов (X3DH, AEAD).

1. Криптографические стандарты Республики Беларусь [Электронный ресурс] //НИИ прикладных проблем математики и информатики БГУ – Режим доступа: `https://apmi.bsu.by/resources/std.html`. – Дата доступа: 01.10.2025.
2. bee2-rs [Электронный ресурс] / Е.С. Пакуль – Режим доступа: `https://github.com/tpyauheni/bee2-rs`. – Дата доступа: 01.10.2025.
3. Peregrine [Электронный ресурс] / Е.С. Пакуль – Режим доступа: `https://github.com/tpyauheni/peregrine`. – Дата доступа: 01.10.2025.

ПРИНЦИПЫ ПОСТРОЕНИЯ И МЕТОДЫ БАЛАНСИРОВКИ НАГРУЗКИ В ВЫСОКОПРОИЗВОДИТЕЛЬНЫХ СЕТЯХ ЦЕНТРОВ ОБРАБОТКИ ДАННЫХ

Почебут А. С.
Кафедра проектирования информационно-компьютерных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: a.pochebut03@gmail.com

В работе рассматриваются современные архитектурные решения и методы балансировки нагрузки высоконагруженных сетей в центрах обработки данных. Основное внимание уделено анализу теоретических подходов к проектированию масштабируемых и отказоустойчивых сетевых архитектур, обеспечивающих равномерное распределение трафика. Приведён обзор существующих методов балансировки, их классификация, ограничения и сравнительный анализ применимости в условиях высокой сетевой нагрузки. Рассмотрены способы повышения масштабируемости и отказоустойчивости инфраструктуры при сохранении высокой производительности.Проанализированы подходы к автоматизации балансировки нагрузки на основе централизованного управления трафиком и анализа потоков данных.

Введение

Современные центры обработки данных представляют собой масштабные вычислительные комплексы, объединяющие тысячи серверов, систем хранения и сетевых устройств. Рост объема данных и усложнение топологии сетей приводят к необходимости совершенствования методов балансировки нагрузки и разработке архитектур, способных эффективно управлять распределением нагрузки между узлами сети. Постоянное увеличение числа соединений между узлами делает традиционные методы маршрутизации и балансировки неэффективными в высоконагруженных условиях.

В таких системах балансировка нагрузки становится не только средством повышения производительности, но и ключевым механизмом обеспечения устойчивости и надёжности сетевой инфраструктуры [1]. От того, насколько эффективно распределяются сетевые потоки между уровнями и сегментами, напрямую зависят показатели задержки, потерь пакетов и доступности сервисов. Таким образом, проблема балансировки трафика требует сочетания теоретических принципов сетевого моделирования с инженерными решениями.

I. Цель и задачи исследования

Целью настоящей работы является анализ архитектурных подходов и методов балансировки нагрузки с целью оптимизации функционирования сетей в центрах обработки данных. В ходе исследования ставятся следующие задачи:

- провести сравнительный анализ существующих архитектур и технологий балансировки трафика с точки зрения масштабируемости, отказоустойчивости и экономичности;
- систематизировать методы распределения сетевой нагрузки по уровням применения (физический, транспортный, прикладной) и типам алгоритмов (детерминированные, вероятностные, адаптивные);
- исследовать влияние характеристик архитектуры сети на производительность и устойчивость при различных сценариях трафика (равномерный, неравномерный, пиковый).

Дополнительно анализируются тенденции развития технологий балансировки в контексте перехода к автоматизированным и самоуправляемым сетям. Рассматривается возможность использования статистических и эвристических методов для прогнозирования сетевой загрузки и предварительного перераспределения потоков, так как эти методы позволяют снизить нагрузку на отдельные узлы и повысить общую эффективность работы системы без необходимости модернизации физической инфраструктуры.

II. Методы исследования

В работе применяются методы системного анализа, теории графов и моделирования потоков данных. Анализ архитектур выполнялся с использованием структурных моделей сетей на основе графов, где узлы соответствуют коммутаторам и маршрутизаторам, а рёбра – каналам связи с ограниченными пропускными способностями. Балансировка рассматривалась как задача оптимального распределения потоков в графе с учётом ограничений по пропускной способности, задержке и приоритетам. Такой подход позволил получить аналитические оценки эффективности архитектурных решений и определить условия, при которых равномерное распределение нагрузки становится невозможным.

Для анализа алгоритмов применялось сравнение детерминированных схем (Weighted Round Robin, Least Connections) и адаптивных моделей, основанных на текущей загрузке каналов. Также

рассматривались теоретические аспекты использования SDN-контроллеров для централизованного управления маршрутами [2]. Результаты сопоставлялись с опубликованными данными лабораторных экспериментов и модельных симуляций, что позволило подтвердить корректность выбранного подхода и определить возможные направления практического применения теоретических выводов.

III. Результаты исследований

Исследование показало, что архитектуры Fat Tree и Spine-Leaf [3] обладают наилучшей масштабируемостью и сбалансированным распределением нагрузки за счёт высокой плотности связей и симметричной структуры соединений (см. рис. 1). В условиях высокой нагрузки вероятность перегрузки отдельных каналов в таких архитектурах снижается на 15-20% по сравнению с классическими трёхуровневыми моделями. Это подтверждает их преимущество при проектировании современных центров обработки данных, где требуется высокая пропускная способность и предсказуемая производительность.

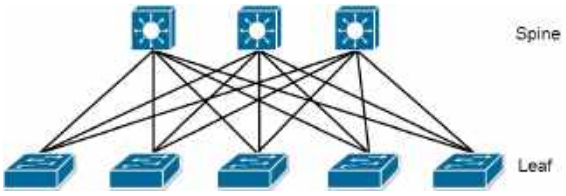


Рис. 1 – Архитектура Spine-Leaf

Среди методов балансировки особый интерес представляют динамические алгоритмы, использующие информацию о текущей загрузке каналов. Они позволяют в реальном времени перераспределять трафик между каналами, снижая вероятность возникновения узких мест и уменьшая задержки. Теоретические оценки показывают, что при применении адаптивных механизмов пропускная способность может увеличиваться на 20-25%, а устойчивость сети к пиковым нагрузкам возрастает на 10-12%. При этом эффективность таких решений во многом зависит от скорости обновления статистики о состоянии каналов и точности прогнозов загрузки.

Анализ комбинированных схем показал, что использование нескольких алгоритмов балансировки одновременно (например, детерминированного и вероятностного) может повысить устойчивость сети к аномальным изменениям трафика [4]. Такие схемы позволяют системе адаптироваться без полной реконфигурации и обеспечивают гибкость при перераспределении потоков. Важным элементом становится применение интеллектуальных модулей, способных принимать решения о маршрутизации на основе текущего контекста сети и статистики её загрузки.

Также выявлено, что избыточность соединений, заложенная в архитектуре, не всегда приводит к росту производительности, если не сопровождается эффективными механизмами балансировки [5]. Для достижения высокой отказоустойчивости необходимо сочетать физическую резервируемость каналов с интеллектуальным управлением маршрутами, что обеспечивает равномерное использование ресурсов и минимизацию времени отклика при отказах.

Заключение

Анализ архитектурных подходов и методов балансировки нагрузки подтверждает, что архитектуры определенных типов являются наиболее перспективными и эффективными при построении высокопроизводительных сетей в центрах обработки данных благодаря своей масштабируемости и устойчивости к сбоям. Применение динамических и комбинированных методов балансировки позволяет значительно повысить эффективность распределения трафика, снизить задержки и улучшить показатели использования ресурсов при минимальных изменениях в инфраструктуре.

В дальнейшем перспективным направлением исследований является разработка аналитических моделей, связывающих параметры архитектуры сети с характеристиками трафика и алгоритмами балансировки. Это позволит формировать рекомендации по выбору оптимальных архитектурных решений и механизмов управления нагрузкой для конкретных сценариев эксплуатации центров обработки данных.

1. Al-Fares, M. A Scalable, Commodity Data Center Network Architecture / M. Al-Fares, A. Loukissas, A. Vahdat // ACM SIGCOMM Computer Communication Review. – 2008. – Vol. 38, № 4. – P. 63–74.
2. Kreutz, D. Software-Defined Networking: A Comprehensive Survey / D. Kreutz, F. Ramos, P. Verissimo // Proceedings of the IEEE. – 2014. – Vol. 103, № 1. – P. 14–76.
3. Докучаев, В. А. Архитектура центров обработки данных : учеб. пособие / В. А. Докучаев, А. А. Кальфа, Д. В. Гадасин. – Москва: МТУСИ, 2018. – 155 с.
4. Cisco Systems. Cisco Data Center Design Guide [Electronic resource]. – Mode of access: https://www.cisco.com/c/en/us/solutions/data-center-virtualization/index.html. – Date of access: 18.10.2025.
5. Greenberg, A. The Cost of a Cloud: Research Problems in Data Center Networks / A. Greenberg, J. Hamilton, D. Maltz // ACM SIGCOMM Computer Communication Review. – 2009. – Vol. 39, № 1. – P. 68–73.

МЕТОДЫ ОПТИМИЗАЦИИ МАРШРУТИЗАЦИИ В ВЫСОКОНАГРУЖЕННЫХ СЕТЯХ В ЦЕНТРАХ ОБРАБОТКИ ДАННЫХ

Почебут А. С.
Кафедра проектирования информационно-компьютерных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: a.pochebut03@gmail.com

В работе представлены результаты исследования методов оптимизации маршрутизации в высоконагруженных сетях в центрах обработки данных. Рассмотрены современные подходы к организации маршрутов, включая использование программно-определяемых сетей (SDN) для оптимизации передачи данных и повышения эффективности использования сетевых ресурсов. Проанализированы методы, основанные на комбинировании централизованных и распределённых механизмов принятия решений, которые позволяют повысить эффективность использования сетевых ресурсов и снизить время отклика систем. Проведена оценка влияния параметров маршрутизации на пропускную способность, задержки передачи данных и устойчивость системы при росте объема трафика.

Введение

Рост объема данных, генерируемых современными цифровыми сервисами, приводит к существенному увеличению нагрузки на инфраструктуру центров обработки данных и требует совершенствования механизмов маршрутизации, обеспечивающих рациональное использование сетевых ресурсов. В условиях масштабных распределённых систем классические методы маршрутизации, основанные на статических таблицах маршрутов, становятся недостаточно гибкими и не способны эффективно реагировать на изменение состояния сети в реальном времени. Возникает необходимость в оптимизации маршрутов на основе анализа текущего состояния сети и прогнозирования нагрузки.

Проблема оптимизации маршрутов особенно актуальна там, где требования к задержке и пропускной способности крайне высоки [1]. Неправильное распределение потоков может привести к деградации производительности и снижению качества обслуживания. В современных условиях маршрутизация должна не только обеспечивать доставку пакетов по кратчайшему пути, но и учитывать параметры текущей загрузки каналов, приоритеты трафика и доступность ресурсов. Следовательно, оптимизация маршрутизации становится комплексной задачей, объединяющей принципы сетевой теории, анализа данных и адаптивного управления.

I. Цель и задачи исследования

Целью работы является анализ методов маршрутизации и определение подходов к их оптимизации в центрах обработки данных в условиях высокой нагрузки. В работе решаются следующие задачи:

- проанализировать существующие протоколы маршрутизации и выявить их ограничения в условиях высокой динамичности сетевого трафика;
- исследовать особенности применения классических алгоритмов маршрутизации (Дейкстры, Беллмана–Форда) в архитектурах центров обработки данных и определить целесообразность их применимости при увеличении числа соединений;
- оценить возможности интеграции программно-определяемых сетей (SDN) с традиционными механизмами маршрутизации и балансировки;
- сравнить эффективность распределения потоков при различных стратегиях выбора маршрутов и архитектурных конфигурациях сети.

Кроме того, рассматривается влияние сетевой топологии и параметров каналов связи на эффективность маршрутизации. Проводится систематизация факторов, определяющих задержку распространения сигналов, устойчивость к отказам и эффективность восстановления маршрутов после сбоев. Такой подход позволяет комплексно оценить применимость теоретических методов к практическим условиям эксплуатации центра обработки данных.

II. Методы исследования

Исследование основано на анализе алгоритмов маршрутизации и моделей сетевых топологий с использованием методов теории графов, комбинаторной оптимизации и анализа потоков в сетях. Сеть рассматривается как ориентированный граф, в котором вершины соответствуют маршрутизаторам, а рёбра соответствуют каналам связи с определенной пропускной способностью и стоимостью передачи. Задача маршрутизации формулируется как нахождение оптимального множества путей, минимизирующих общие задержки при соблюдении ограничений пропускной способности.

В рамках анализа проведено сравнение основных алгоритмов маршрутизации по критериям устойчивости, адаптивности к изменениям состояния сети и сложности реализации [2]. Особое внимание уделено рассмотрению алгоритмов, обеспечивающих балансировку нагрузки, например, ECMP (Equal Cost Multi-Path), который позволяет распределять потоки данных по нескольким эквивалентным маршрутам. Также анализируется использование SDN-контроллеров для централизованного управления маршрутами, при котором такой контроллер может принимать решения на основе глобальной информации о состоянии сети. Этот подход обеспечивает более точную оптимизацию, но требует дополнительных ресурсов для мониторинга и анализа данных.

III. Результаты исследований

Анализ показал, что классические алгоритмы динамической маршрутизации, такие как алгоритм Дейкстры и алгоритм Беллмана-Форда обеспечивают стабильность маршрутов и предсказуемое поведение сети, но не обладают достаточной гибкостью и адаптивностью в динамически изменяющихся условиях высоконагруженных центров обработки данных [3]. Применение протоколов OSPF и ECMP позволяет улучшить балансировку нагрузки, однако требует значительных вычислительных ресурсов при масштабировании, так как их вычислительная сложность растёт пропорционально числу узлов. В теоретических моделях наблюдается зависимость времени перерасчёта таблиц маршрутизации от числа маршрутизаторов, что при больших масштабах сети становится критическим фактором производительности.

Использование программно-определяемых сетей (SDN) [4], напротив, позволяет централизованно контролировать маршруты и управлять трафиком в зависимости от текущего состояния сети. Такой подход обеспечивает гибкость и возможность внедрения интеллектуальных механизмов принятия решений, но создаёт новые риски. Зависимость от контроллера и увеличение задержек при обновлении маршрутов может негативно сказаться на работе всей сети.

Изучение сочетания приницпов распределённой и централизованной логики маршрутизации позволяет сформулировать рекомендации по применению гибридных схем управления в современных центрах обработки данных.

Была также рассмотрена адаптивная маршрутизация с использованием метрик пропускной способности, времени отклика, длины пути и стоимости маршрута. Исследование показало, что гибридные схемы, сочетающие статическую маршрутизацию для устойчивых сегментов и динамическую для магистральных каналов, позволяют повысить общую эффективность сети на 10-15% и снизить вероятность перегрузок. Такие решения обеспечивают компромисс между скоростью реакции и стабильностью маршрутов, что особенно важно для сетей с высокой вариативностью нагрузки.

Кроме того, исследованы вопросы устойчивости маршрутов при отказе узлов. Использование резервных таблиц маршрутизации, протоколов быстрой конвергенции (например, BFD – Bidirectional Forwarding Detection) и алгоритмов быстрого перестроения путей позволяет повысить доступность сервисов и снизить время восстановления соединений в теоретической модели на 30-40%. Это подтверждает важность применения многоуровневых схем маршрутизации, обеспечивающих дублирование и интеллектуальное управление маршрутами для повышения надёжности и отказоустойчивости сети.

Заключение

Результаты проведённого исследования показывают, что оптимизация маршрутизации в высоконагруженных сетях центров обработки данных должна основываться на рациональном сочетании централизованных и распределённых подходов. Применение SDN-технологий позволяет повысить гибкость и управляемость сети, но требует внедрения механизмов контроля и резервирования, снижающих риски отказов. В процессе анализа подтвердилась эффективность гибридных решений, при которых маршруты формируются с учётом состояния каналов и прогнозируемой нагрузки.

В перспективе развитие технологий маршрутизации будет связано с внедрением интеллектуальных систем анализа сетевого трафика и предиктивных моделей, основанных на машинном обучении [5]. Это позволит не только оптимизировать текущие маршруты, но и адаптировать инфраструктуру центров обработки данных к изменениям нагрузки в реальном времени, обеспечивая более высокий уровень устойчивости и эффективности работы сетей.

1. Shabanov, B. M. Building the Software Defined Data Center / B. M. Shabanov, O. I. Samovarov // Proceedings of ISP RAS. – 2018. – Vol. 30, № 6. – P. 7–24.
2. Таненбаум Э. Компьютерные сети / Э. Таненбаум, Д. Уэзеролл. – Изд. 6-е. – СПб.:Питер, 2023. – 816 с.
3. Medhi D. Network Routing: Algorithms, Protocols, and Architectures / D. Medhi, K. Ramasamy. – San Francisco:Morgan Kaufmann Publishers, 2007. – 957 с.
4. Kreutz, D. Software-Defined Networking: A Comprehensive Survey / D. Kreutz, F. Ramos, P. Verissimo // Proceedings of the IEEE. – 2014. – Vol. 103, № 1. – P. 14–76.
5. Feamster, N. Why (and How) Networks Should Run Themselves [Electronic resource] / N. Feamster, J. Rexford // arXiv preprint. – Mode of access: https://arxiv.org/abs/1710.11583. – Date of access: 19.10.2025.

ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА ПРИНЯТИЯ УПРАВЛЕНЧЕСКИХ РЕШЕНИЙ В СФЕРЕ КАДРОВОГО ДЕЛОПРОИЗВОДСТВА

Прудникова И.В.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: irena_prudnikova@mail.ru

Актуальность поддержки принятия управленческих решений неуклонно растет в условиях современной динамичной и неопределенной бизнес-среды. Всё больше компаний сталкиваются с необходимостью оперативно обрабатывать огромные объемы данных, предсказывать будущие тренды и принимать обоснованные решения в условиях ограниченных ресурсов и времени.

Одним из трендов в области управления персоналом является автоматизация HR-процессов. Выделяются основные процессы в области управления персоналом, которые все чаще автоматизируются: найм и отбор персонала; системы обучения и развития; системы мотивации сотрудников; системы адаптации [1]. В связи с этим, системы поддержки принятия решений в сфере управления персоналом приобретают все большую актуальность, позволяя автоматизировать рутинные задачи, анализировать большие объемы данных и принимать обоснованные стратегические решения.

Согласно рассмотренных результатов исследования о преимуществах цифровой трансформации в сфере управления персоналом, изучается эффективность современных технологий, которые могут быть использованы для цифровизации HR-процессов: использование облачных сервисов, система электронного документооборота, искусственный интеллект, чат-бот, VR (виртуальная реальность) и AR (дополненная реальность), роботизированная автоматизация процессов (RPA), HRM-системы [2].

Многие компании используют неформализованные или слабоформализованные подходы к управлению персоналом, часто полагаясь на ручные методы, такие как таблицы Excel и бумажные документы (приказы, докладные записки и т.д.), вместо автоматизированных систем. Применение инновационных технологий в области автоматизации отвечает идее гипераватоматизации (термин, введенный аналитиками Gartner)[3], согласно которой все, что можно автоматизировать в организации, должно быть автоматизировано. Г. И. Алиева и В. Д. Семина в своей статье выделяют зоны риска автоматизации HR-процессов, такие как: некорректная постановка задачи для разработчика, ошибки при написании технического задания; переоценка результатов от внедрения HR-автоматизации (завышенные ожидания и некорректная оценка эффективности); избыточное упрощение процессов. Исключение и/или минимизация наступления рисков автоматизации HR-процессов невозможно без четкой структуризации и формализации бизнес-процессов [4]. Автоматизация этих процессов, обеспечивающая доступ к актуальной информации, позволяет принимать обоснованные решения, основанные на данных и прогнозной аналитике.

Изучив функциональные возможности HRM-систем, представленных на рынке Республики Беларусь, выявлено ряд недостатков, которые препятствуют владельцу процесса (HR-менеджеру) или руководителю компании принимать быстрые и эффективные решения в области управления персоналом:

- основной функционал представленных HRM-систем сфокусирован на рекрутинге, который заключается в подборе релевантного резюме и размещения в базе кандидатов компании, но нет функционала перевода кандидата в статус сотрудника без ручного ввода необходимой информации специалиста по кадрам;
- функционал адаптации персонала базируется на напоминании руководителю и работнику об этапах адаптационного периода, но нет оценки адаптации и рекомендаций по разработке планов развития;
- функционал оценки персонала строится на опросах, которые разрабатывают и вводят HR-менеджеры, но нет прогнозной аналитики на основании полученных результатов;
- все системы работают на основе информации, введенной вручную HR-менеджером (за исключением отбора релевантных резюме);
- ни одна из перечисленных систем не предлагает комплексные решения по HR-процессам, что вынуждает компании приобретать модульные решения и интегрировать в свои процессы.

Проанализировав научные исследования и функционал HRM-систем, можно выделить несколько перспективных направлений развития автоматизации HR-процессов.

Во-первых, это интеграция ИИ-алгоритмов для прогнозной аналитики.

Во-вторых, распространение RPA-решений для автоматизации рутинных HR-операций.

В-третьих, развитие HRM-систем с расширенными возможностями самообслуживания сотрудников.

Следует отметить, что все составляющие рассматриваемых HRM-систем оперируют одними и теми же данными о сотрудниках, обеспечивая целостность и непротиворечивость информации на всех этапах работы с персоналом. Это достигается благодаря централизованному хранению и обработке данных, что позволяет избежать дублирования и повышает эффективность работы системы в целом. Автоматизация процессов управления персоналом позволяет ускорить процессы принятия решений и снизить риски ошибок, связанных с человеческим фактором. В результате компании могут значительно сократить время и затраты на управление персоналом, а также повысить качество работы кадровых служб, становясь более стратегическими и ориентированными на результат, способствуя достижению конкурентного преимущества организации.

Применение ИТ-решения таких как RPA дает возможность взаимодействовать с любыми информационными системами, сайтами и выполнять ровно то, что необходимо для реализации бизнес-процесса. Программные роботы могут извлекать, переносить, изменять информацию и данные, как если бы это делал человек за компьютером, только в несколько раз быстрее и точнее не допуская ошибок. RPA востребована во многих областях среднего и крупного бизнеса, начиная с банковского сектора, здравоохранения, логистики, страхования, производства и строительства. Вместе с тем, роботизация процессов активно применяется не только в сферах, где важную роль играет служба клиентской поддержки, но и внутренних процессах компании. Одним из таких внутренних процессов является – система управление персоналом (HR-система).

Пример процессов управления персоналом, которые могут быть роботизированы:

1. Ввод данных о сотрудниках: Автоматическое внесение информации о новых сотрудниках в HR-систему, включая личные данные, контактную информацию, данные о трудовом стаже и т.д. Это может включать в себя импорт данных из различных источников, таких как сканированные документы или электронные таблицы.
2. Обработка отпусков и больничных листов: Автоматическое создание и обработка заявок на отпуск, больничные листы и другие виды отсутствий сотрудников. Робот может проверять соответствие заявки правилам компании и автоматически обновлять данные в системе учета рабочего времени.
3. Расчет заработной платы: RPA может значительно упростить отдельные этапы этого процесса, например, автоматический сбор данных о заработной плате и премиях, автоматическое начисление налогов и других вычетов.
4. Подготовка кадровых документов: Автоматическое создание и оформление различных кадровых документов, таких как приказы о приеме на работу, увольнении, изменении должности и т.д.
5. Управление обучением и развитием: Автоматизация процесса регистрации сотрудников на тренинги и курсы, рассылка напоминаний о мероприятиях, сбор оценок и отзывов.
6. Управление рабочим временем: Автоматический сбор и анализ данных о рабочем времени сотрудников, контроль соблюдения правил и норм.

Применения RPA оптимально подходит для быстрого достижения измеримых результатов автоматизации без масштабных изменений бизнес-процессов.

Актуальность темы роботизации в системе управления персоналом обусловлена необходимостью повышения эффективности работы HR-системы, снижения затрат и освобождения времени HR-специалистов для выполнения более стратегических задач. Также актуальность RPA в системе управления персоналом обусловлена постоянным ростом объема данных, повышением требований к эффективности и качеству работ, нехваткой квалифицированных специалистов и развития новых HR-технологий. В рамках дальнейшего исследования планируется изучить систему HR-процессов, степень их автоматизации, в результате чего предложить алгоритм интеллектуальной поддержки принятия управленческих решений в сфере кадрового делопроизводства.

1. Гоненко А. А., Грудинина Е. А. Автоматизация процессов управления персоналом: тренды и изменения, 2024 г.
2. Горев А. В., Тонких Н. В., Автоматизация процессов управления удаленным персоналом с помощью цифровых технологий: опыт и рекомендации, 2025 г.
3. Hyperautomation [Электронный ресурс]. – Режим доступа: `https://www.gartner.com/en/information-technology/glossary/hyperautomation`. – Дата доступа: 05.10.2025.
4. Алиева Г. И.,Семина В. Д., Риски при автоматизации работы hr-специалиста, 2023 г.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СОВРЕМЕННЫХ МЕТОДОВ РЕШЕНИЯ ЛОГИКО-КОМБИНАТОРНЫХ ЗАДАЧ

Семижон Е. А., Герман Ю. О., Герман О. В.
Факультет информационных технологий и управления,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: juliagerman@gmail.com

Приведен сравнительный анализ современных методов решения логико-комбинаторных задач, которые могут использоваться для интеграции в практически эффективные решатели этих задач.

Введение

При отсутствии изначально метода решения задачи возможны следующие технологии поиска решения (парадигмы) [1,2]: P1. Построение поискового дерева в духе метода A* Нильсона. P2. Решение задачи на основе слабого (приближенного, эвристического) метода с последующим отсечением (части области поиска) и повторное (итерационное) использование слабого метода до исчерпания области поиска или срабатывания критерия получения точного решения; P3. Реализация ограниченного направленного перебора, генетических алгоритмов . P4. Использование нейросетей, классифицирующих деревьев. P5. Использование статистических методов, например, на основе теоремы Байеса, методов типа ограниченной выборки из популяции, метода главных компонентов, факторного анализа и др. P6. Использование жадных алгоритмов, методов многокритериального выбора альтернатив, эвристических методов (например, метода отжига, метода муравья и подобных). P7. Применение логических методов решения задач, экспертных систем, индуктивных логических методов. P8. Использование аналогий и моделирования (например, на основе метода Монте-Карло).

I. Анализ подходов

Парадигма P1 связана с построением поискового дерева (дерева решения, гипотез). Узлы дерева представляют состояния задачи. С каждым узлом связывается количественная оценка, на основании которой производится выбор очередного более перспективного узла для раскрытия и продвижения по дереву. Кроме того, приписываемые узлам оценки должны позволять отсекать бесперспективные узлы. Отметим недостатки алгоритма A* Нильсона, которые могут ограничивать его применение в некоторых ситуациях.

1. Высокие требования к памяти. Алгоритм хранит все открытые и закрытые узлы, что может привести к значительным затратам по памяти, особенно в больших графах. Это может стать критическим фактором при работе с ограниченными ресурсами.
2. В графах с большим количеством узлов и рёбер, особенно если они плохо структурированы, A* может работать медленно, так как исследует много узлов. Эффективность A* зависит от качества используемой эвристической функции. Если эвристика неадекватна (например, не является допустимой), это может привести к неэффективному поиску или даже к невозможности нахождения оптимального решения.
3. Неоптимальность в динамических средах A* не подходит для динамических сред, где граф может изменяться во время выполнения алгоритма. В таких случаях алгоритм может потребовать значительных затрат на пересчет путей. Как альтернативу можно указать алгоритм D* для динамически изменяющегося дерева состояний.
4. Алгоритм требует предварительной информации о графе, что не всегда возможно. Например, в случаях, когда граф представляется в реальном времени (например, для навигации), требуется постоянное обновление данных.

Парадигма P2 хорошо иллюстрируется симплексным методом решения задачи линейного программирования. В симплексном методе выполняется пошаговый переход в очередной узел многогранника решений, в котором значение целевой функции лучше (не хуже, во всяком случае), чем в предыдущем узле. Этим гарантируется отсечение ранее найденных решений и сужение области поиска до ее полного исчерпания. Проблема связана с тем, что возможны случаи, когда количество просматриваемых узлов экспоненциально растет с линейно растущими размерами задачи, то есть симплексный алгоритм теоретически имеет экспоненциальную сложность, хотя на практике в среднем ведет себя как полиномиальный метод. Кроме того, симплексный алгоритм может потенциально зацикливаться при переходе в узел, на котором значение целевой функции не изменяется. Полиномиально эффективное решение задаче линейного программирования доставляет метод Л. Хачияна, известный как метод стягивающих эллипсоидов. Суть его состоит в последовательном отсечении частей описанного вокруг симплекса эллипсоида с сохранением точки оптимума. По духу, к методу эллипсоидов близок метод Н.

Кармаркара – также полиномиальной сложности, который даже лучше в этом отношении метода Л. Хачияна. Методы типа отсечений (например алгоритм Гомори для задач целочисленной оптимизации) по-прежнему сохраняют экспоненциальную вычислительную сложность и имеют следующие общие недостатки

1. Сложность реализации. Методы отсечений могут требовать сложной реализации, особенно при разработке эффективных эвристик для генерации отсечений. Это может потребовать значительных усилий и экспертизы.
2. Неоптимальность отсечений. Не все отсечения являются оптимальными. Плохо подобранные отсечения могут не уменьшить пространство поиска эффективно, что может привести к увеличению времени выполнения.
3. Потребление памяти. При использовании методов отсечений может потребоваться значительное количество памяти для хранения сгенерированных отсечений, что может быть проблемой в больших задачах.
4. Комбинаторный взрыв. В некоторых случаях, особенно с увеличением размера входных данных, количество возможных отсечений может увеличиваться экспоненциально, что усложняет управление.

Парадигма P3 наиболее ярко реализована в генетических алгоритмах. Направленный перебор реализуется размножающейся популяцией объектов-решений, которые порождают новые объекты-решения примерно как это делается в эволюционной схеме. Во многих случаях генетические алгоритмы дают отличные результаты. Однако их эффективность заметно снижается при наличии ограничений. Отыскиваемое генетическим алгоритмами решение не является доказательно оптимальным.

Парадигма P4 вообще связана с представлением о грядущей сингулярности цивилизационного масштаба. Нейросети приходят на выручку тогда, когда модель задачи построить крайне сложно или невозможно. Аналитик в этом случае готовит данные для обучения нейросети. Если данные подготовлены качественно и полно, то результат может быть почти «ошеломляющим». Распознавание лиц, речи, перевод языков, постановка диагнозов, выявление лжи и другие подобные вещи хорошо решают именно нейросети, с которыми связали ожидаемый всплеск искусственного интеллекта. Тем не менее, сложные комбинаторные задачи с предположительно экспоненциальной сложностью вряд ли смогут быть решены нейросетями за полиномиальное время, если, разумеется, гипотеза о несовпадении классов P и NP верна, что принимается в общем случае и при этом ищется точное решение задачи. Как и в случае генетических алгоритмов, решения, доставляемые нейросетями в общем случае не являются доказательно оптимальными. Парадигма P5 ориентирована на задачи с неполными, недостоверными знаниями. Как и в предыдущем случае, решение ищется без построения модели либо строится статистическая модель (например, на основе регрессии), которая схватывает суть приближенно (отсюда и необходимость проверки статистической адекватности модели). Тем не менее, статистический подход дает практически приемлемые результаты и имеет большое значение в решении задач. Качество решений оценивается с помощью статистических критериев.

Группа задач с парадигмой P6 ориентирована на эвристические методы, например, жадные алгоритмы. Имеются теоретически обоснованные случаи, когда жадные алгоритмы доставляют точное решение задачи (это относится к задачам со структурой матроидов). Интерес представляют различного рода усиления эвристических методов.

Логические методы в решении задач (парадигма P7) имеют самостоятельное значение. Логика напрямую связана с булевой алгеброй, поэтому логические формулы могут быть переведены на язык алгебры, например, задача логического вывода может быть переформулирована как нелинейная оптимизационная алгебраическая задача. Особый статус логике дает теорема С. Кука об NP-полноте задачи Выполнимость. В силу этого результата любая комбинаторная задача из NP эффективно сводится к задаче Выполнимость. Однако полиномиально эффективного алгоритма для задачи Выполнимость не найдено, и можно полагать, что его вообще нет. Большой интерес к логике связан с индуктивным и нечетким выводом как инструментом, пользующимся практическим спросом.

Аналогии – это не столько парадигма, сколько то общее, что присуще другим методам. Например, распознавание образов и постановка диагнозов основаны на выявлении аналогий. Язык аналогий нечеткий, поэтому нечеткая математика является теоретической платформой для этой парадигмы. Как вывод отметим, что методы, ориентированные на парадигмы P2, P3, P4, P6, P7 следует рассматривать под углом зрения их интеграции в «точные» решатели, как средство усиления вычислительной мощности таких решателей.

II. Заключение

Наибольший интерес связан с интеграцией (объединением) приведенных механизмов, например, с нейронными сетями, которые могут использоваться как «полноценные» решатели, так и в составе подобных решателей, использующих другие механизмы (например, алгоритм A* Нильсона на дереве состояний задачи).

1. Кормен, Т. Алгоритмы.Построение и анализ / Т. Кормен, Ч.Лейзерсон. // Wiley. M, 2013, 1328с.
2. Luke, S. Essentials of metaheuristics / S. Luke // Lulu publisher, Usa. 2009, 320p.

АНАЛИЗ СТЕГАНОГРАФИЧЕСКИХ МЕТОДОВ

Серкевич Д. С., Герман Ю. О.
Кафедра защиты информации,
кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: sdasha695@gmail.com, jgerman@bsuir.by

В данной работе проведена классификация методов стеганографии для изображений, а также проанализированы их сильные и слабые стороны. Представлена актуальность развития методов стеганографии в сфере кибербезопасности. Понимание принципов функционирования стеганографических методов и их практического применения способствует построению эффективных систем защиты данных от несанкционированного доступа и сохранию конфиденциальности личной информации.

Введение

С ростом распространенности кибератак потребность в безопасных методах передачи информации стала первостепенной. Основная цель стеганографии – скрыть сам факт передачи сообщения, в отличие от криптографии, которая шифрует содержание, но не скрывает саму передачу.

Стеганография в изображениях стала более популярной в последние годы, чем другие виды стеганографии, возможно из-за большого потока изображений в электронном виде, доступного с появлением цифровых камер и высокоскоростной интернет-передачи. Поэтому она играет важную роль в обеспечении информационной безопасности, защите авторских прав и передачи конфиденциальных данных [1].

I. Стеганографическая система

Стеганография – это метод организации связи (передачи сообщений), при котором скрывается само наличие связи.

Стеганографическая система (см. рис. 1) в общем виде состоит из двух компонентов: вставки и извлечения сообщения.

Рис. 1 – Стеганографическая система

Процесс обработки стеганографического сообщения включает следующие объекты:

- сообщение – данные любого типа;
- контейнер – любая информация, пригодная для сокрытия в ней сообщений;
- стегоконтейнер – контейнер, содержащий скрытое сообщение;
- ключ (стегоключ) – секретный ключ, необходимый для шифрования (расшифровки) сообщения с целью усиления защиты.

По аналогии с криптографией, по типу стегоключа стегосистемы можно подразделить на два типа:

1. с секретным ключом;
2. с открытым ключом [2].

II. Методы для временной области

1. *Псевдослучайные перестановки.* Метод вставляет биты сообщения с изменением порядка их появления в сообщении на основе генератора псевдослучайных чисел. Это делает внедрение менее предсказуемым и затрудняет обнаружение скрытых данных.
   Преимущества:
   - метод вставляет биты сообщения, изменяя порядок их следования, что усложняет процесс обнаружения и расшифровки информации;
   - биты скрываемого сообщения равномерно распределены по всему контейнеру.

   Недостатки:
   - бит сообщения, вставленный случайным образом в младший бит контейнера, может быть поврежден;
   - требуется синхронизация ключей между отправителем и получателем [3].
2. *Метод с использованием патчей.* Этот метод является статистическим кодированием информации путем изменения некоторых статистических свойств контейнера и использует проверку гипотез при извлечении сообщения. Секретный ключ применяется к случайно выбранному подмножеству пикселов из изображения, затем подмножество разделяется по двум патчам – A и B.
   Преимущества:
   - устойчивость к атакам;
   - гибкость маскировки;
   - сообщение распределено по всему изображению.

   Недостатки метода:
   - сложность реализации;
   - ограниченная емкость, в один патч инкапсулируется только один бит;

– зависимость от качества сегментации.

3. *Метод расширения спектра.*Модификация метода расширения спектра заключается в изменении базисной функции, которая является произведением гармонического сигнала на псевдослучайную последовательность.Частотная характеристика сообщения обладает гораздо меньшей энергией, чем энергия контейнера.
   Преимущества:
   – высокая устойчивость к шуму и атакам;
   – низкая вероятность обнаружения;
   – применимость к различным типам носителей.
   Недостатки:
   – сложность реализации;
   – низкая емкость;
   – зависимость от качества генерации псевдослучайной последовательности [4].
4. *Метод наименее значимого бита (LSB).* Метод заключается в замене последнего значащего бита в исходном файле на биты сообщения, которое надо скрыть. В результате подмены полученный файл должен быть не отличим от оригинала человеческими органами восприятия.
   Преимущества:
   – простота реализации;
   – высокая емкость;
   – быстрая обработка, подходит для реального времени и маломощных устройств.
   Недостатки:
   – низкая устойчивость к атакам;
   – чувствительность к преобразованиям;
   – при неправильном выборе носителя или чрезмерной нагрузке данные становятся заметными [5].

III. Методы преобразования области определения

1. *Дискретное косинус-преобразование (ДКП).* Данный метод преобразует блоки изображения из пространственной области в частотную, что делает их более надежными по сравнению с методами во временной области, включая сжатие, обрезку и некоторые алгоритмы обработки изображений.
   Преимущества:
   – устойчивость к сжатию, метод совместим с JPEG;
   – гибкость внедрения.
   Недостатки:
   – сложность реализации;
   – ограниченная емкость, в каждый блок можно внедрить лишь небольшое количество информации;
   – уязвимость к фильтрации.
2. *Дискретное вейвлет-преобразование (ДВП).* Инкапсуляция сообщения с помощью ДВП дает хорошие результаты, которые превосходят методы ДКП. Вейвлет-преобразование разлагает изображение на поддиапазоны частот: низкие частоты по горизонтали и вертикали (основная структура изображения) и высокочастотные компоненты, содержащие детали и шум.
   Преимущества:
   – высокая скрытность ;
   – устойчивость к сжатию и фильтрации;
   – совместимость с кодами коррекции ошибок;
   Недостатки:
   – сложность реализации;
   – ограниченная емкость, внедрение больших объёмов данных может повлиять на качество изображения;
   – чувствительность к обработке [3].

IV. Заключение

В данной статье были рассмотрены некоторые из основных методов стеганографии изображений для временной области и преобразования области определения. Исходя из проведенного анализа следует отметить, что различные форматы графических файлов имеют свои методы для сокрытия информации, которые в свою очередь имеют сильные и слабые стороны.

Выбор метода стеганографии зависит от конкретных требований и условий применения. Таким образом, если важна простота и скорость, можно использовать LSB метод, для усложнения извлечения без изменения структуры – псевдослучайные перестановки, для изображений с высокой степенью защиты и устойчивостью к атакам – метод ДКП или ДВП. Преобразование ДКП или ДВП более надежно, потому что позволяет скрыть сообщение в области частот. Данная область менее подвержена зрению человека. С точки зрения развития методов стеганографии, интересен подход, основанный на перестановке битов, замене старших и младших разрядов.

V. Список литературы

1. Грибунин, В. Г. Стеганография, цифровые водные знаки и стегоанализ / В. Г. Грибунин // – М.: Вузовская книга, 2018. – 110 с.
2. Грибунин, В. Г. Цифровая стеганография. Аспекты защиты / В. Г. Грибунин, И. Н. Оков, И. В. Туринцев // – М.: Солон-Пресс, 2009. – 264 с.
3. Пономарев, И. В. Стеганографические методы встраивания и обнаружения сокрытых сообщений, использованых GIF-изображения в качестве файлов-контейнеров/ И. В. Пономарев, Д. И. Строкин // Известия Алтайского государственного университета. –2022. –№1(123).– с. 112-115.
4. Земцов, А. Методы цифровой стеганографии для защиты авторских прав / А. Земцов, // – М.: LAP Lambert Academic Publishing, 2019. – 148 с.
5. Рассел, Дж. Стеганография / Дж. Рассел, // – М.: VSD, 2019. – 193 с.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ ДАННЫХ: РЕЛЯЦИОННАЯ, ГРАФОВАЯ И СЕТЕВАЯ МОДЕЛИ

Славинский Г. А., Боброва Н. Л.
Кафедра информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: gr.slavin.sci@mail.ru, bobrova@bsuir.by

Целью статьи является проведение сравнительного анализа трех моделей данных: реляционной, графовой и сетевой. В ходе исследования были определены критерии сравнения моделей, выполнен их анализ, рассмотрены области применения каждой модели, выявлены их сильные и слабые стороны. В статье выдвинуто предположение о возможном развитии и будущем каждой из моделей.

Введение

Модель данных – логическая структура, описывающая организацию, хранение и взаимосвязи данных в базе. Она служит основой для построения и управления базами данных и поддерживается системами управления базами данных (СУБД) [1].

Эволюция моделей баз данных тесно связана с развитием вычислительной техники и изменяющимися требованиями к организации и обработке информации. Первые СУБД основывались на иерархической и сетевой моделях, которые появились в 1960-х годах и стали первыми попытками формализовать хранение и доступ к данным. Недостатки иерархической и сетевой моделей оказали огромное влияние на развитие баз данных, сформировав требования к новым моделям.

В начале 1970-х годов Эдгар Кодд[2] предложил реляционную модель данных, которая благодаря своей математической строгости и удобству работы через язык запросов SQL быстро стала стандартом в индустрии. Реляционные базы данных обеспечили более простое моделирование данных и эффективное прикладное программирование, что значительно упростило разработку информационных систем.

С ростом объемов данных и появлением новых задач, связанных с Big Data и распределенными системами, традиционные реляционные модели начали испытывать ограничения по масштабируемости и гибкости структуры данных. Это привело к появлению и развитию нереляционных моделей баз данных, которые предлагают альтернативные подходы к организации информации, включая документные, ключ-значение, графовые и колоночные модели [3].

Идея графовых баз данных, в основе которых лежит модель хранения данных в виде графов, появилась ещё в 1980-х годах [4]. В 2007 году СУБД Neo4j стала первой коммерчески успешной и широко используемой реализацией этой концепции [5].

В данной статье рассматриваются реляционная, графовая и сетевая модели. Каждая из них отражает важные парадигмы хранения и обработки данных, применяемые в СУБД. Понимание этих моделей позволяет выбрать оптимальный подход к проектированию и реализации информационных систем.

I. Критерии оценки

Сравнительный анализ – это метод исследования, используемый для изучения сходств и различий между некоторым множеством объектов, явлений или подходов. Данный метод предполагает последовательную оценку объектов по заранее выбранным критериям с целью выявления их преимуществ, недостатков, особенностей.

Критерии оценки представляют собой параметры или признаки, по которым оцениваются и сопоставляются объекты, явления или подходы. Правильный выбор критериев обеспечивает объективность сравнительного анализа.

Для сравнения указанных моделей данных были выбраны следующие критерии, которые важны при выборе и оценке модели данных [6]:

- Структура данных – способ организации данных;
- Гибкость – адаптивность к изменениям в структуре данных;
- Масштабируемость – возможность увеличения объема данных и числа пользователей, вертикальная – увеличение мощности одного сервера, горизонтальная – распределение нагрузки;
- Производительность – скорость обработки запросов и операций с данными;
- Целостность и согласованность данных – способность обеспечения корректной работы и непротиворечивости информации;
- Актуальность – востребованность сегодня;
- Область применения – насколько модель подходит для конкретных задач и типов данных.

Выбранные параметры предполагают определение как возможностей, так и ограничений рассматриваемых моделей. Глубокое понимание вопроса способствует более обоснованному выбору модели для решения конкретных задач, что

минимизирует риск ошибок на начальной стадии проекта.

II. Результаты проведенного анализа

Результаты проведенного анализа представлены в таблице 1.

Таблица 1 – Результаты сравнительного анализа

Критерий	Реляционная	Графовая	Сетевая
Структура данных	Двумерные таблицы, ключи	Узлы и ребра графа	Связанные наборы записей
Гибкость	Средняя	Высокая	Низкая
Масштабируемость	Хорошо масштабируется вертикально	Гибридный подход	Крайне ограничена
Производительность	Высокая для табличных запросов	Высокая для связей	Хорошая для навигации
Целостность и согласованность	Сильная через транзакции и ограничения	Зависит от реализации	Крайне ограничена
Актуальность	Широко применяется	Рост интереса	Редко используется
Область применения	Универсальная	Соц. сети, рекомендации	Телекоммуникации, инженерия

На основе данных в таблице 1, были сделаны следующие выводы о пригодности моделей:

- универсальное решение с сильной поддержкой целостности данных и хорошей производительностью – реляционная;
- оптимальное решение для задач со сложными взаимосвязями – графовая;
- решение с ограниченной гибкостью и масштабируемостью, но с хорошей производительностью для навигации по сложным связям – сетевая.

Выбор подходящей модели данных должен основываться на специфике конкретной задачи, что минимизирует риски ошибок и повышает вероятность успешной реализации проекта.

III. Детализация преимуществ и недостатков моделей

Реляционная модель характеризуется высокой надежностью благодаря строгой структуре и поддержке ACID-транзакций. Однако может испытывать снижение производительности при работе со сложными связями, сложнее масштабируется горизонтально.

Графовая модель специализирована на эффективной работе с сильно связанными данными[4]. Ее достоинства – естественное представление связей, высокая производительность и гибкость схемы. Графовые СУБД могут уступать в скорости простых агрегационных запросов, требуют освоения специфичных языков запросов.

Сетевая модель ранее использовалась для моделирования сложных связей, но сегодня считается устаревшей, среди ее недостатков:низкая гибкость, отсутствие современных средств обеспечения целостности данных.

Каждая модель имеет свои сильные и слабые стороны. Реляционная модель является наиболее популярным и универсальным решением, графовая модель имеет свои сильные стороны, способствующие росту интереса к ней. Сетевая модель применяется преимущественно в специализированных системах, но не является массовым решением для большинства бизнес-задач.

IV. Заключение

Таким образом, с учетом текущих тенденций, реляционная модель останется фундаментом традиционных бизнес-приложений. Графовые базы данных продолжат набирать популярность и расширять сферу применения с акцентом на те области, где требуется анализ сложных сетевых структур. Применение сетевой модели в некоторых специализированных системах способно сохраняться и в дальнейшем.

Будущее по прежнему за масштабируемыми и гибкими моделями, способными эффективно справляться со сложными и взаимосвязанными данными, а также адаптированными к работе с большими объемами данных и распределенными системами. Реляционные и графовые модели останутся востребованными инструментами разработки, дополняя друг друга в разных сферах применения.

V. Список литературы

1. Database Systems: The Complete Book / H. Garcia-Molina, J. D. Ullman, J. Widom. – NJ: Pearson, 2008. – 1248 p.
2. Edgar F. Codd / IBM [Electronic resource]. – Mode of access: `https://www.ibm.com/`.
3. Новиков, Б. А. Сравнительный анализ производительности SQL и NoSQL СУБД / Б. А. Новиков, М. Ю. Левин // Компьютерные инструменты в образовании. – 2017. – № 4. – С. 48–63.
4. Graph Databases: New Opportunities for Connected Data / I. Robinson, J. Webber, E. Eifrem. – CA: O'Reilly Media, 2015. – 236 p.
5. Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL / Дж. Р. Уилсон [и др.]. – М.: ДМК Пресс, 2013. – 386 с.
6. Введение в системы баз данных / К. Дж. Дейт. – М.: Диалектика, 2024. – 648 с.

СРАВНЕНИЕ АЛГОРИТМОВ ВЗАИМОДЕЙСТВИЯ КОМПОНЕНТОВ В СУЩЕСТВУЮЩИХ ФРЕЙМВОРКАХ И РАЗРАБОТКА ДВУХСВЯЗНОЙ МОДЕЛИ ВЗАИМОДЕЙСТВИЯ НА СТОРОНЕ КЛИЕНТА

Струкова А. А., Шилин Л. Ю.
кафедры информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: alinastrukovaa@gmail.com, shilin@bsuir.by

В статье проводится сравнительный анализ алгоритмов взаимодействия компонентов в популярных фреймворках Angular и React. На основе выявленных недостатков предлагается новый алгоритм взаимодействия компонентов на стороне клиента, основанный на двухсвязной модели обмена данными.

Введение

В последние годы наблюдается устойчивый рост популярности компонентных архитектур при создании клиентских приложений. Разделение интерфейса на изолированные модули упрощает разработку, тестирование и поддержку систем, однако при этом усложняется процесс взаимодействия между компонентами и синхронизации данных между ними.

Наиболее известные фреймворки Angular и React решают эту задачу по-разному. Angular использует строгую модель связывания данных и централизованное управление изменениями, в то время как React делает ставку на простоту и оптимизацию через виртуальный DOM.

Обе модели доказали свою эффективность, но обладают определёнными ограничениями: Angular страдает от избыточных проходов по дереву компонентов, а React требует сложных инструментов для синхронизации состояний.

В связи с этим актуальной задачей становится разработка алгоритма взаимодействия компонентов, который объединяет преимущества существующих подходов, минимизируя их недостатки.

Архитектурные принципы

Angular реализует архитектуру MVVM (Model–View–ViewModel), где каждый компонент связан с шаблоном, отражающим текущее состояние данных. Передача информации между компонентами осуществляется через входные параметры, обеспечивающие поток данных сверху вниз, выходные события, формирующие поток снизу вверх, и реактивные потоки Observable библиотеки RxJS.

Каждое изменение состояния инициирует механизм Change Detection, проходящий по дереву компонентов и проверяющий, произошло ли изменение значений, используемых в шаблоне.

Это обеспечивает предсказуемость поведения, однако в крупных приложениях вызывает значительные накладные расходы при рендеринге и синхронизации состояний.

React опирается на концепцию однонаправленного потока данных: родитель передаёт свойства дочерним компонентам, а те реагируют на изменения состояния с помощью хуков (useState, useEffect, useMemo).

Основное отличие React – использование виртуального DOM, который сравнивает текущее и новое состояние дерева компонентов. Рендеринг выполняется только для реально изменившихся элементов:

$$\Delta DOM = DOM_{new} - DOM_{old}$$

Это делает React более производительным, особенно при частых обновлениях интерфейса. Для сложных взаимодействий между множеством компонентов часто требуются дополнительные библиотеки для глобального управления состоянием (Redux, MobX, Zustand), что увеличивает когнитивную сложность системы.

Проблемы существующих подходов

Сравнение архитектур фреймворков Angular и React позволяет выделить следующие ключевые проблемы:

1. Angular: чрезмерная нагрузка на систему обнаружения изменений, особенно при больших деревьях компонентов.
2. React: отсутствие встроенной двухсторонней синхронизации данных между компонентами, необходимость использовать сторонние хранилища.
3. Обе системы: избыточная передача данных по цепочкам компонентов, даже когда изменения касаются локальных областей.

Следовательно, требуется алгоритм, который минимизирует количество перерендеров, сохраняет реактивность при изменениях данных и обеспечивает синхронизацию между зависимыми компонентами без промежуточных прослоек.

Двухсвязная модель взаимодействия компонентов

Предлагаемый алгоритм основан на двухсвязной модели (Bidirectional Component Linking Model, BCLM), в которой каждый компонент может как публиковать изменения своего состояния, так и подписываться на изменения состояния других компонентов напрямую, минуя посредников.

Пусть C_i и C_j – два компонента с локальными состояниями $S_i(t)$ и $S_j(t)$. Связь между ними определяется как двунаправленное отношение:

$$L_{i,j} = (C_i \leftrightarrow C_j)$$

Каждое изменение состояния δS_i автоматически вызывает событие синхронизации:

$$\tau_{i,j}(t) = h(S_i(t))$$

$$S_j(t+1) = f(S_j(t), \tau_{i,j}(t))$$

Функция h определяет, какие именно данные компонента C_i подлежат публикации, а функция f – как эти данные применяются у подписчика. Таким образом, состояние каждого компонента может влиять на другие в обе стороны, формируя адаптивную реактивную сеть.

Этапы работы алгоритма

Алгоритм взаимодействия компонентов на основе двухсвязной модели включает следующие шаги:

1. Инициализация связей: каждый компонент регистрирует свои исходящие и входящие связи $L_{i,j}$;
2. Обнаружение изменений: если $\delta S_i \neq 0$, запускается событие обновления;
3. Передача данных подписчикам:

$$\tau_{i,j} = h(S_i)$$

4. Применение изменений у подписчиков:

$$S_j(t+1) = f(S_j(t), \tau_{i,j})$$

5. Оптимизация рендеринга: обновление интерфейса выполняется только при фактическом изменении состояния:

$$R_j = [S_j(t+1) \neq S_j(t)]$$

В отличие от Angular, система не требует глобального обхода дерева компонентов, поскольку обновления распространяются только между реально связанными узлами. В отличие от React, состояние может обновляться двунаправленно, что исключает необходимость в централизованных стореджах и промежуточных прослойках.

Преимущества двухсвязной модели

Предложенный алгоритм имеет ряд преимуществ. Он предоставляет изоляцию изменений, так как обновления распространяются только на те компоненты, которые связаны логически, без обхода всего дерева. Отсутствие глобального механизма change detection снижает нагрузку на процессор. Также алгоритм снижает когнитивную сложность, так как нет необходимости использовать внешние библиотеки для глобального состояния.

Двухсвязная модель позволяет как прямую, так и обратную передачу данных без нарушения реактивности, это делает алгоритм взаимодействия компонентов на стороне клиента гибким. Алгоритм может быть интегрирован в существующие архитектуры Angular и React как промежуточный слой взаимодействия компонентов, поэтому одним из преимуществ является совместимость разработанного решения с существующими системами на базе существующих фреймворков.

Заключение

Предложенный алгоритм взаимодействия компонентов на стороне клиента, основанный на двухсвязной модели, сочетает в себе преимущества архитектур Angular и React, устраняя их основные недостатки. Данный подход позволяет достичь баланса между структурированностью (характерной для Angular) и производительностью (характерной для React). Он снижает количество ненужных обновлений, повышает согласованность данных и обеспечивает гибкость при проектировании сложных SPA-приложений.

I. Список литературы

1. Грубер И. Ю. Современные JavaScript-фреймворки. – Минск: ITBooks, 2022.
2. Google Developers. *Angular Change Detection Explained.* – 2023.
3. Dan Abramov. *Inside React Fiber: Reconciliation Algorithm.* – React Blog, 2022.
4. Preact Team. *Signals: Fine-Grained Reactivity for the Modern Web.* – 2023.

АНАЛИЗ КВАНТИЛЬНОГО И ГАУССОВСКОГО ПОДХОДОВ В МОДЕЛИРОВАНИИ РЕДАКЦИОННЫХ ПРОЦЕССОВ

Сьянов Д. А.
Кафедра веб-технологий и компьютерного моделирования,
Белорусский государственный университет
Минск, Республика Беларусь
E-mail: syanov@bsu.by

В работе рассматривается задача вероятностного прогнозирования длительностей процессов в редакционных рабочих потоках научных публикаций. Для решения используется модель на основе архитектуры Transformer с двумя вариантами выходных подсистем: квантильной (с логарифмическим масштабированием целевой переменной) и гауссовской (предсказывающей параметры нормального распределения). Проведено сравнительное исследование эффективности обеих модификаций на реальных данных после глубокой очистки и обогащения признаков.

I. Введение

Прогнозирование сроков прохождения рукописи через редакционные этапы является одной из задач управления научными журналами. От продолжительности каждого этапа зависят планирование редакционной нагрузки, прозрачность коммуникации с авторами и общая эффективность публикационного цикла. Однако временные интервалы между событиями в редакционном процессе характеризуются высокой вариативностью, асимметрией и зависимостью от контекста, что затрудняет применение классических точечных моделей регрессии [1,2].

Рассматривается возможность использования данных о динамике поступления и обработке рукописей в редакционном процессе как аналога интенсивности поступления и интенсивности потока обслуживания заявок в классических системах массового обслуживания. Такая постановка задачи позволяет рассматривать редакционную систему как стохастический процесс с переменной нагрузкой и конечной пропускной способностью, что делает релевантным применение вероятностных моделей.

Для обработки данных используется архитектура Adaptive Probabilistic Sequence-to-Vector Transformer (APS2V-T), предназначенная для вероятностного прогнозирования длительностей этапов редакционного процесса. Модель основана на трансформерном энкодере [3], обрабатывающем последовательности событий и контекстные признаки, и включает две стохастические подсистемы – квантильную и гауссовскую, описывающие распределения длительностей.

II. Формирование датасета

Исходным источником данных выступает система для организации рецензируемых научных изданий Open Journal Systems (OJS), содержащая данные о времени наступления всех редакционных событиях. Из «сырых» логов восстанавливались последовательности событий для каждого идентификатора рукописи, начиная с момента подачи и до публикации.

Процесс подготовки данных включал нормализацию типов событий и унификацию меток редакторских решений с восстановлением пропущенных публикационных событий на основе метаданных, дополнение выборки признаками нагрузки (числом активных материалов, количеством рецензентов и интервалами активности пользователей), а также построение унифицированных последовательностей с фиксированным набором стадий и их выравниванием по общей временной шкале.

Целевые значения рассчитывались как абсолютные длительности (в часах) от даты подачи до наступления последующих стадий. Маскирование целей исключает использование данных о будущих событиях при формировании признаков для прогноза [4]. Для обеспечения корректности моделирования применялись фильтры качества: удалены записи с отсутствующими или противоречивыми временными метками, случаи нулевой или отрицательной задержки публикации. После очистки итоговый корпус составил около 4 тыс. документов и более 30 тыс. событий. Основные статистические характеристики приведены в табл. 1.

Таблица 1 – Распределение длительностей по основным стадиям публикационного процесса

Показатель	Ср.(ч)	мин.	макс.	средн.
нач.рец.	648	4	1956	423
зав.рец	1812	288	2592	595
реш.ред	2430	375	3383	502
нач. лит.ред.	3640	1697	5943	811
публик.	4210	2454	6168	890

Распределения длительностей характеризуются выраженной асимметрией с удлинённым «хвостом» на поздних стадиях, что отражает неоднородность редакционных циклов и различия в динамике работы подразделений. Этот дисбаланс

был учтён при стратификации данных и формировании батчей в процессе обучения модели APS2V-T.

III. Результаты обучения

Обучение модели APS2V-T проводилось отдельно для двух подсистем — квантильной и гауссовской. Основной целью было сравнение способности каждой подсистемы адекватно моделировать распределение временных интервалов ключевых событий редакционного процесса.

По полученным метрикам (табл. 2, рис. 1) видно, что:

- Квантильная аппроксимация демонстрирует существенное превосходство над гауссовской моделью по метрике CRPS [5], что подтверждает эффективность использования дискретной сетки квантилей при моделировании асимметричных и гетероскедастичных временных интервалов;
- Значение MAE для медианного квантиля (q50) в квантильной модели оказывается ниже, чем для точечной оценки в гауссовской, что свидетельствует о более высокой точности воспроизведения центральной тенденции распределения;
- Гауссовская аппроксимация проявляет ограниченную способность к корректному описанию асимметричных распределений длительностей, однако сохраняет ценность в качестве компактного параметрического представления прогноза.

Таблица 2 – Сравнение по ключевым метрикам

Аппрокс.	Epoch	CRPS	NLL	MAE
Квантильная	162	527.742	948.112	24.3
Гауссовская	300	3222.190	4900.509	31.7

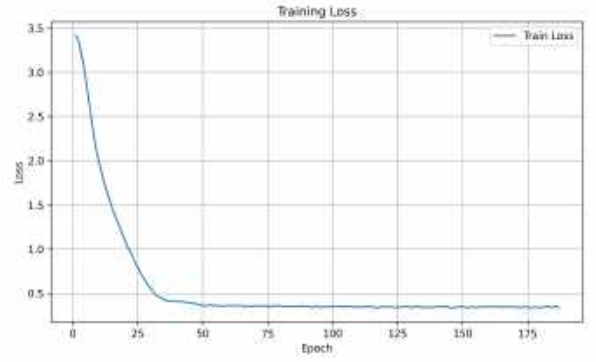


Рис. 1 – train loss квантильная

Качественный анализ предсказаний модели показывает явные различия между квантильной и гауссовской подсистемами APS2V-T. Квантильная аппроксимация корректно отражает вариативность событий на всех этапах редакционного процесса: предсказанные значения $q10$ и $q90$ дают интервал вероятных сроков, позволяя оценить как более ранние, так и более поздние сценарии наступления ключевых событий. Оверлей кривых обучения, представленный на рисунке (рис. 2), демонстрирует, что квантильная аппроксимация сходится быстрее и обеспечивает более стабильное поведение CRPS на валидационном сете, тогда как гауссовская аппроксимация демонстрирует замедленное снижение значения CRPS, что указывает на более медленную сходимость и меньшую устойчивость прогноза.

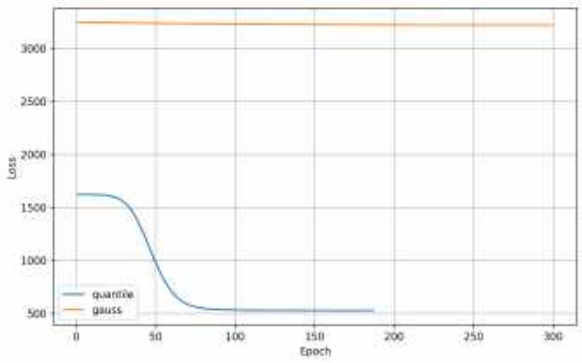


Рис. 2 – CRPS

IV. Заключение

Проверка на отдельных кандидатах, особенно для рукописей с длинными задержками, подтверждает преимущества квантильной подсистемы: она сохраняет реалистичные прогнозы для редких, но значимых длинных хвостов распределения, в то время как гауссовская подсистема систематически недооценивает длительные задержки. Таким образом, дискретная квантильная аппроксимация позволяет точнее прогнозировать медиану и разброс сроков, обеспечивая надёжное представление полной вариативности временных интервалов в редакционном процессе.

1. Wasserman, L. All of Statistics: A Concise Course in Statistical Inference [Electronic resource] : Springer NY, 2013. – Mode of access: `https://doi.org/10.1007/978-0-387-21736-9`. – Date of access: 23.10.2025.
2. Harrell, F. E. Jr. Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis [Electronic resource] : Springer Cham, 2015. – Mode of access: `https://doi.org/10.1007/978-3-319-19425-7` – Date of access: 23.10.2025.
3. Vaswani, A. et al. Attention Is All You Need [Electronic resource] : NIPS, 2017. – Mode of access: `https://doi.org/10.48550/arXiv.1706.03762`. – Date of access: 23.10.2025.
4. Salinas, D., Flunkert, V., Gasthaus, J. , Januschowski, T. DeepAR: Probabilistic forecasting with autoregressive recurrent networks [Electronic resource] : International Journal of Forecasting. – July–September 2020. – Volume 36. – Issue 3. – P. 1181–1191. – Mode of access: `https://doi.org/10.1016/j.ijforecast.2019.07.001` – Date of access: 23.10.2025.
5. Gneiting, T., Raftery, A. E. Strictly Proper Scoring Rules, Prediction, and Estimation [Electronic resource] : Journal of the American Statistical Association. – March, 2007. – Vol.102. – No. 477. – P. 359–378. – Mode of access: `https://doi.org/10.1198/016214506000001437` – Date of access: 23.10.2025.

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА МЕТОДА ADVERSARIAL TRAINING НА ЗАДАЧЕ БИНАРНОЙ КЛАССИФИКАЦИИ

Хаджинова Н. В., Агель А. А., Пашковец М. В.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: {agel, matvey.pashkovez01}@gmail.com, khajynova@bsuir.by

Быстрое распространение систем, основанных на машинном обучении (ML), обострило проблему их безопасности и устойчивости к целенаправленным атакам. Это поставило под угрозу достоверность обработки информации. В частности, атаки обхода (evasion attacks) с использованием состязательных примеров (adversarial examples) представляют серьезную угрозу для систем классификации данных, включая их способность корректно интерпретировать входные данные. В настоящей работе рассматривается и оценивается метод Adversarial Training (Состязательное обучение) как ключевая стратегия обеспечения надежности и достоверности классификационных моделей. В качестве механизма генерации состязательных примеров использован итерационный метод Projected Gradient Descent (PGD). Целью работы является разработка и экспериментальная оценка метода повышения устойчивости классификаторов к атакам обхода для обеспечения защиты систем обработки данных.

Введение

Автоматизированные системы, основанные на машинном обучении, стали неотъемлемым компонентом информационных инфраструктур. Однако их широкое внедрение обострило проблему защиты от целевых атак, направленных на компрометацию целостности и достоверности данных [1]. Проблема обеспечения безопасности ML-систем решается различными путями, включая как создание робастных моделей, так и разработку интеллектуальных агентов, способных активно противодействовать атакам [2]. В настоящей работе мы фокусируемся на первом подходе. Особую угрозу представляют состязательные атаки (adversarial attacks), которые позволяют злоумышленнику манипулировать входными данными, чтобы заставить модель принимать заранее определенные неверные решения. Такие атаки, как отравление (poisoning) данных на этапе обучения и обход (evasion) на этапе эксплуатации, используют добавление к данным малозаметных возмущений (в рамках заданной ℓ_p-нормы), что подрывает саму основу доверия к автоматизированным системам [3].

Проблема обеспечения робастности ML-агентов к этим атакам является ключевой задачей в области безопасности искусственного интеллекта. Для агентов, выполняющих задачи мониторинга, классификации или принятия решений, обеспечение устойчивости к состязательным возмущениям является важным условием для сохранения целостности информационных систем и противодействия целевым кибератакам.

I. Adversarial Training как метод защиты

Adversarial Training (AT) является наиболее эффективным и широко используемым методом повышения робастности моделей к состязательным примерам. Суть метода заключается в включении состязательных примеров в обучающий набор данных. Модель обучается не только на чистых данных, минимизируя обычную функцию потерь $L(\theta, x, y)$, но и на данных с максимально возможным возмущением δ в пределах допустимой ℓ_p-нормы ε, которые максимизируют функцию потерь:

$$\min_{\theta} \mathbb{E}_{(\theta,x,y)\sim D}\left[L(\theta, x, y) + \min_{\|\delta\|_p \leq \varepsilon} L(\theta, x + \delta, y)\right],$$

где θ — параметры модели, x – входные данные, y — истинный класс, и D — распределение данных [3-5]. Этот процесс можно интерпретировать как обучение модели, которое минимизирует функцию потерь в наихудшем случае, обеспечивая таким образом своего рода регуляризацию. Компромисс между робастностью и точностью, исследованный в [5], является важным аспектом при применении данного метода.

II. Экспериментальная часть

В данной работе принципы обеспечения робастности и метод состязательного обучения демонстрируются на примере двумерного классификатора, решающего синтетическую задачу бинарной классификации (Класс 0 и Класс 1). Моделирование атаки демонстрирует целенаправленное искажение входных признаков (x_1, x_2) с минимальной величиной возмущения [4]. Эксперимент проводился с использованием языка Java для чистого воспроизведения алгоритмов.

Для исследования выбрана модель линейного классификатора (логистическая регрессия). Модель имеет два веса и смещение, использует функцию активации Сигмоида. Обучение проводится с использованием функции потерь Mean

Squared Error. Выбор MSE обусловлен необходимостью упрощения демонстрации принципов обеспечения достоверности обработки информации.

Для демонстрации использован синтетический 2D-набор данных из 100 точек. Входные признаки (x_1, x_2) генерируются случайным образом в диапазоне [0,1] и являются линейно разделимыми.

Для обучения модели применялся двухэтапный подход. На первом этапе модель обучалась в течение 1000 эпох с использованием стандартного метода градиентного спуска исключительно на чистых данных, формируя базовую модель. На втором этапе проводилось состязательное обучение в течение 1000 эпох. На каждой итерации для каждого исходного образца x с помощью метода FGSM генерировался состязательный аналог x_{adv}, используя текущие веса модели и возмущение с ℓ_∞-нормой ε=0.2, применяемое однократно. Затем модель обучалась на модифицированных образцах, минимизируя функцию потерь.

Для оценки надежности обработки информации использовались два типа входных данных:

1. Исходные данные (x): Немодифицированные входные образцы из набора данных. Точность обработки этих данных (Accclean) характеризует стандартную производительность системы в штатных условиях;
2. Модифицированные данные (xadv): Образцы, целенаправленно искаженные путем добавления малозаметного возмущения (δ). Возмущение генерировалось с помощью одношагового метода FGSM с максимальной допустимой ℓ_∞-нормой ε=0.2.

Совместное использование этих двух типов данных позволяет количественно оценить уязвимость системы обработки информации: высокая точность на исходных данных при значительном снижении на модифицированных образцах свидетельствует о недостаточной надежности обработки и уязвимости к целенаправленным искажениям.

В эксперименте сравнивались две модели, каждая из которых обучалась в течение 1000 эпох: Базовая Модель, обученная только на чистых данных, и Устойчивая Модель, обученная на состязательных примерах (x_{adv}), генерируемых на каждой итерации. Оценка проводилась на полном наборе данных с ε=0.2.

Таблица 1 – Итоговые результаты робастности

Модель	Acc_{clean}(Чистые данные)	Acc_{adv}(Состязательные данные)
Базовая Модель	95,00 %	30 %
Устойчивая Модель	62,00 %	62,00 %

Заключение

В настоящей работе разработан и экспериментально проверен метод повышения устойчивости алгоритмов классификации в автоматизированных системах к целенаправленным искажениям входных данных. Ключевым результатом является создание методики защиты на основе состязательного обучения (Adversarial Training), которая позволяет повысить робастность модели к атакам с обходом защиты. Разработанный протокол проверки целостности данных и механизм генерации состязательных примеров формируют основу для встраивания модуля обнаружения аномалий в конвейер обработки информации. В качестве дальнейшего развития данного исследования представляет интерес интеграция предложенного метода в агентные системы, способные к активной защите, как это показано в [2].

Экспериментальные результаты подтвердили эффективность предложенного метода, продемонстрировав увеличение устойчивости классификатора. Полученные метрики работы системы в нормальных и атакованных условиях позволяют оптимально настроить баланс между производительностью и надежностью автоматизированных систем обработки информации.

1. Goodfellow, I. J. Explaining and Harnessing Adversarial Examples / I. J. Goodfellow, J. Shlens, C. Szegedy // International Conference on Learning Representations (ICLR). – 2015.
2. Khajynava, N. Adaptation of adversarial machine learning for training agents to counter data attacks / N. Khajynava, Z. Mutero, A. Adam // Технические средства защиты информации. – Минск, 2025. – С. 385–387.
3. Surekha, M. A Comprehensive Analysis of Poisoning Attack and Defence Strategies in Machine Learning Techniques / M. Surekha, A. K. Sagar // 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT). – 2024.
4. Madry, A. Towards deep learning models resistant to adversarial attacks / A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu // International Conference on Learning Representations (ICLR). – 2017.
5. Zhang, H. Theoretically principled trade-off between robustness and accuracy / H. Zhang, Y. Yu, J. Jiao, E. P. Xing, L. El Ghaoui, M. I. Jordan // International Conference on Machine Learning (ICML). – 2019.

КОГНИТИВНО-ДОВЕРИТЕЛЬНАЯ МОДЕЛЬ ГИБРИДНЫХ СИСТЕМ «ЧЕЛОВЕК–АГЕНТ»

Хаджинова Н. В., Кадлубай В. А.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: khajynova@bsuir.by, tylyla2004@gmail.com

Разработана модель информационных потоков в системе "человек–агент оптимизирующая обработку данных в условиях когнитивных ограничений пользователя. Предложены алгоритмы адаптивной фильтрации и представления информации, снижающие когнитивную нагрузку на оператора на 28%. Разработаны принципы семантического согласования ментальных моделей пользователя и агента для эффективной совместной обработки информации. На основе эксперимента подтверждена эффективность предложенных решений.

Введение

Современные автономные системы и искусственные агенты становятся неотъемлемой частью профессиональной деятельности операторов сложных систем в таких областях, как управление технологическими процессами, мониторинг критических объектов и обработка больших данных. Однако полная автоматизация часто невозможна или нежелательна, что приводит к необходимости создания гибридных систем, в которых человек и автономный агент совместно обрабатывают информацию и принимают решения.

Ключевыми проблемами таких систем являются формирование адекватного уровня доверия оператора к агенту и управление когнитивной нагрузкой в процессе информационного взаимодействия. Недостаточное доверие приводит к неэффективному использованию возможностей системы, тогда как избыточное доверие может вызвать снижение критичности оператора и ухудшение ситуационной осведомленности.

I. Многоуровневая модель информационного взаимодействия

Информационное взаимодействие в системах «человек–агент» представляет собой сложный многокомпонентный процесс. На основе анализа архитектур автоматизированных систем предложена трехуровневая модель доверия:

- *Инструментальное доверие* – основано на оценке надежности и точности обработки информации агентом;
- *Процедурное доверие* – формируется через понимание алгоритмов работы системы;
- *Нормативное доверие* – связано с соответствием ментальных моделей оператора и агента.

Ключевыми факторами доверия являются: прозрачность обработки данных агентом, предсказуемость его поведения, возможность контроля со стороны оператора и историческая надежность. Проведенные исследования[3] показывают, что прозрачность системы увеличивает доверие на 25–40% по сравнению с системами типа черные ящики.

II. Классификация когнитивных нагрузок в информационных системах

Совместная обработка информации в гибридных системах создает специфические когнитивные нагрузки, которые можно классифицировать следующим образом:

Таблица 1 – Типы когнитивных нагрузок в гибридных системах

Тип нагрузки	Источник	Влияние на эффективность
Информационная	Мониторинг потоков данных	Снижение на 15–30%
Процедурная	Принятие алгоритмов обработки информации	Снижение на 10–25%
Коммуникационная	Анализ результатов обработки данных	Снижение на 20–35%

Эксперимент[3] с участием 45 операторов сложных систем показал, что неправильное распределение когнитивной нагрузки приводит к снижению точности обработки информации на 22% и увеличению времени выполнения задач на 35%.

III. Методы оценки и оптимизации информационных потоков

Для оценки когнитивной нагрузки применялся комплекс методов, включающий субъективные шкалы NASA-TLX, объективные физиологические показатели (частота сердечных сокращений, кожно-гальваническая реакция) и поведенческие метрики обработки информации.

Разработана модель оценки когнитивной нагрузки при работе с информационными система-

ми:

$$CL_{\text{total}} = \alpha \cdot \sum_{i=1}^{n} I_i W_i + \beta \cdot \int_0^T P(t)dt + \gamma \cdot \log(1+C), \tag{2}$$

где CL_{total} – общая когнитивная нагрузка; I_i – информационные потоки; $P(t)$ – процедурная сложность; C – коммуникационные затраты; α, β, γ – весовые коэффициенты, определяемые экспериментально.

Предложены следующие архитектурные решения для снижения когнитивной нагрузки:

1. Адаптивная визуализация – динамическое изменение уровня детализации информации в зависимости от контекста задачи;
2. Прогрессивное раскрытие – поэтапное представление сложных данных;
3. Семантическое согласование – обеспечение концептуальной совместимости ментальных моделей пользователя и агента.

Внедрение предложенных методов в систему обработки информации позволило снизить общую когнитивную нагрузку на 28% и повысить точность совместных решений на 18%.

IV. Экспериментальная проверка модели доверия

Для верификации предложенной модели доверия проведен эксперимент[3]. В исследовании участвовали 45 оператора различной квалификации.

Основные результаты эксперимента:

$$T = 0.45 \cdot C + 0.32 \cdot P + 0.23 \cdot R, \tag{1}$$

где T – общий уровень доверия, C – оценка компетентности, P – прозрачность решений, R – надежность.

Статистический анализ данных подтвердил значимость всех коэффициентов модели ($p < 0.01$). Корреляция между предсказанным и измеренным уровнем доверия составила $r = 0.84$.

V. Принципы проектирования человеко-машинных интерфейсов

На основе проведенных исследований [3] сформулированы практические рекомендации по проектированию гибридных систем обработки информации:

- Обеспечивать постепенное увеличение автономности агента по мере роста доверия пользователя;
- Реализовывать механизмы объяснения решений, адаптированные к уровню квалификации пользователя;
- Предоставлять возможности тонкой настройки уровня автоматизации обработки данных;
- Использовать единообразные интерфейсные метафоры для снижения процедурной нагрузки.

Заключение

Разработанная многоуровневая модель информационного взаимодействия и методы управления когнитивной нагрузкой позволяют существенно повысить эффективность гибридных систем «человек–агент». Подтверждено, что учет предложенных принципов при проектировании систем обработки информации способствует формированию адекватного уровня доверия и оптимизации когнитивных нагрузок.

Перспективными направлениями дальнейших исследований являются разработка адаптивных интерфейсов, динамически подстраивающихся под когнитивное состояние оператора, и создание методов прогнозирования доверия на основе анализа информационных потоков.

VI. Список литературы

1. Hoff, K. A. Trust in automation: integrating empirical evidence on factors that influence trust / K. A. Hoff, M. Bashir // Human Factors. – 2015. – Vol. 57, № 3. – P. 407–434.
2. Светличный, А. М. Когнитивные нагрузки в человеко–машинных системах / А. М. Светличный, Е. В. Попова – М.: Наука, 2018. – 256 с.
3. Chen, J. Y. Human–autonomy teeming: a review and analysis of the empirical literature / J. Y. Chen, E. J. de Visser // Human Factors. – 2021. – P. 1–45.
4. Lyons, J. B. Being transparent about transparency: a model for human–robot interaction / J. B. Lyons // Trust and autonomous systems: papers from the AAAI Spring Symposium. – 2013. – P. 45–52.
5. Lee, J. D. Trust in automation: designing for appropriate reliance / J. D. Lee, K. A. See // Human Factors. – 2004. – Vol. 46, № 1. – P. 50–80.

МИНИМИЗАЦИЯ ЗАДЕРЖЕК В РАСПРЕДЕЛЕННЫХ МУЛЬТИАГЕНТНЫХ СИСТЕМАХ С ИСПОЛЬЗОВАНИЕМ ГРАНИЧНЫХ ВЫЧИСЛЕНИЙ

Хаджинова Н. В., Михнюк А. И., Засим А. О.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: khajynova@bsuir.by, {alexander.mikhniuk, alexeyzasim5}@gmail.com

Исследована интеграция граничных вычислений (Edge Computing) и мультиагентных систем (MAS) для снижения временных задержек в умных сетях (Smart Grid) и телемедицине. Проанализированы временные преимущества в отклике локальной обработки данных на периферии сети совместно с децентрализованным управлением агентами по сравнению с облачными решениями. Рассмотрены такие механизмы, как федеративное обучение, гибридная архитектура (Kubernetes, MEC) и протоколы (MQTT, gRPC), оптимизированные для децентрализованных сред. Результаты подтверждают эффективность гибридных систем, объединяющих скорость Edge и автономию MAS.

Введение

Современные критические приложения, такие как интеллектуальные энергетические сети и телемедицина, предъявляют экстремальные требования к скорости обработки данных. Традиционные облачные вычисления сталкиваются с фундаментальными ограничениями: задержками передачи и рисками перегрузки сетей [1, 2]. Интеграция граничных вычислений (Edge Computing) и мультиагентных систем (MAS) открывает путь к минимизации задержек, перенося вычисления ближе к источникам данных и обеспечивая децентрализованную координацию.

Новизна данного исследования заключается в системном анализе роли Edge Computing как катализатора эффективности MAS для задач обработки данных со сверхнизкими задержками. Практическая значимость работы заключается в предложении конкретных механизмов, таких как локальная предобработка данных и динамическое перераспределение задач между агентами. Последующие разделы раскрывают архитектурные принципы Edge-MAS и демонстрируют их эффективность на примерах из энергетики и здравоохранения.

I. Основы технологий

Граничные вычисления (Edge Computing) переносят обработку данных на периферию сети, ближе к источникам информации, таким как IoT-устройства и датчики [2]. Это позволяет выполнять предобработку и запуск легковесных алгоритмов локально, кардинально сокращая объем передаваемых в облако данных и снижая задержки.

Распределенные мультиагентные системы (MAS) обеспечивают децентрализованное управление через сеть автономных агентов, способных к кооперации. Их интеграция с Edge Computing создает синергию: Edge-устройства поставляют агентам предобработанные данные, а те, в свою очередь, динамически перераспределяют вычислительные задачи, оптимизируя ресурсы. Например, в умных сетях гибридные архитектуры на базе Kubernetes на граничных узлах позволяют сократить задержку до 5–10 мс [2].

Ключевую роль в интеграции играют протоколы, оптимизированные для децентрализованных сред. MQTT и gRPC обеспечивают эффективную коммуникацию, Raft ускоряет достижение консенсуса, а Kafka служит надёжным брокером сообщений [1]. Таким образом, комбинация Edge и MAS формирует основу для высокопроизводительных систем в энергетике, телемедицине и других областях с жесткими требованиями к реальному времени.

II. Механизмы снижения задержки

Снижение временной задержки в системах, объединяющих граничные вычисления и мультиагентные системы, достигается за счет глубокой интеграции вычислительных ресурсов и адаптивных стратегий взаимодействия. Одним из перспективных методов является предиктивная обработка данных, при которой Edge-устройства не только фильтруют информацию, но и прогнозируют возможные сценарии на основе локальных моделей машинного обучения [3]. Например, в системах управления автономным транспортом сенсоры лидара предварительно анализируют траектории движения, отправляя в облако только отклонения от нормы, что сокращает объем передаваемых данных и время реакции на препятствия.

Важную роль играет оптимизация сетевых протоколов для децентрализованных сред. Технологии, такие как QUIC (Quick UDP Internet Connections) [1], заменяющие традиционный TCP, минимизируют задержки за счет уменьшения числа «рукопожатий» между устройствами. В сочетании с мультиагентными системами это позволяет организовать параллельную обработку запросов: так, в умных фабриках роботы-агенты, управля-

ющие конвейером, координируют действия через одноранговые соединения, исключая зависимость от центрального сервера.

Динамическое кэширование часто используемых данных на граничных узлах – еще один механизм, снижающий нагрузку на каналы связи. Например, в системах телемедицины результаты анализа типичных случаев (например, стандартных ЭКГ) хранятся локально на Edge-серверах больниц, что ускоряет диагностику повторяющихся сценариев. Мультиагентные системы при этом управляют актуальностью кэша, автоматически обновляя данные на основе изменений в медицинских протоколах.

Особый интерес представляют аппаратно-программные оптимизации для Edge-устройств [4]. Использование специализированных микросхем (ASIC) или нейропроцессоров (NPU) позволяет выполнять ресурсоемкие задачи, такие как обработка видео в реальном времени для систем видеонаблюдения, непосредственно на периферии. Агенты, встроенные в камеры, анализируют потоки данных с помощью оптимизированных алгоритмов компьютерного зрения, сокращая задержку до 2–3 мс, что критично для идентификации угроз в режиме реального времени.

Перспективным направлением остается гибридизация подходов, где предиктивная аналитика на Edge сочетается с глобальной координацией через MAS. Например, в энергосистемах агенты, распределенные по подстанциям, используют локальные прогнозы потребления для балансировки нагрузки, а MAS синхронизирует их действия через механизмы консенсуса, предотвращая конфликты. Это устраняет необходимость постоянного обращения к облаку, сокращая задержку на 40 – 60% в сравнении с чисто централизованными решениями [2, 4] .

Таким образом, снижение временной задержки достигается не за счет отдельных технологий, а через их синергию, где Edge Computing обеспечивает скорость, а MAS – интеллектуальное распределение ресурсов. Это формирует основу для систем, где каждый миллисекундный выигрыш трансформируется в повышение безопасности, надежности и эффективности.

III. Вызовы и ограничения

Внедрение гибридных систем Edge-MAS сталкивается с комплексом взаимосвязанных проблем.

Технические и ресурсные ограничения. Ограниченная вычислительная мощность и энергопотребление Edge-устройств препятствуют выполнению сложных алгоритмов в реальном времени, создавая парадокс между автономностью и функциональностью [1, 4].

Проблемы безопасности. Децентрализованная архитектура увеличивает поверхность для кибератак, а отсутствие единого центра контроля затрудняет мониторинг и реагирование. Применение методов шифрования для защиты данных может нивелировать преимущества в виде низкой задержки [2].

Организационные и нормативные барьеры. Интеграция с унаследованными системами требует значительной модернизации инфраструктуры и пересмотра протоколов [1]. Автономность агентов также порождает этико-юридические дилеммы, связанные с размыванием ответственности за их решения.

Таким образом, успешное внедрение зависит не только от технологического прогресса, но и от преодоления организационного сопротивления и развития нормативной базы.

Заключение

В результате исследования разработана гибридная архитектура обработки информации, сочетающая граничные вычисления и мультиагентные системы (Edge-MAS). В этой архитектуре узлы отвечают за предварительную локальную обработку данных, облачные ресурсы нагружены глубокой аналитикой предобработанных данных, а MAS отвечает за интеллектуальное и отказоустойчивое распределение ресурсов. Такой архитектурный подход позволяет достичь значительного сокращения задержки (40–60%).

При анализе вызовов и ограничений выявлено, что несмотря на потенциал, внедрение Edge-MAS сдерживается ограничениями периферийных устройств, проблемами безопасности и сложностями интеграции с унаследованными системами. Перспективы дальнейших исследований видятся в разработке энергоэффективных процессоров, унифицированных протоколов и гибридных архитектур.

Таким образом, комбинация Edge и MAS представляет собой новую парадигму для построения высокопроизводительных систем реального времени. Последующие работы должны быть направлены на создание адаптируемых межотраслевых платформ для преодоления существующих барьеров и массового внедрения.

1. Khan M.A., Ahmadon M.A. Trends and Challenges in Mobile Edge Computing for the Next Generation Massive Internet of Things // Evolution of Information, Communication and Computing System, 2023. Vol. 4. P. 28–42.
2. Liang B. et al. Multi-access Edge Computing Fundamentals, Services, Enablers and Challenges: A Complete Survey // Journal of Network and Computer Applications, 2022. Vol. 199.
3. Аль-Тамими М., Хассан М. Б., Аббас С. А. Федеративное обучение (FL)–Обзор //Изв. СПбГЭТУ «ЛЭТИ. – 2024. – Т. 17. – № 5. – С. 74-82.
4. Bebortta S. et al. Performance Analysis of Multi-Access Edge Computing Networks for Heterogeneous IoT Systems // Materials Today: Proceedings, 2022. Vol. 58. P. 267-272.

АЛГОРИТМ ЛОКАЛЬНОЙ РЕОПТИМИЗАЦИИ МАРШРУТОВ НА ДИНАМИЧЕСКИХ ГРАФАХ

Хаджинова Н. В., Служалик В. Ю., Хаджинова К. А.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {khajynova, sluzhalik}@bsuir.by, kseniyakhadzhynava@gmail.com

Предложен алгоритм локальной реоптимизации маршрутов в динамических графах с мультикритериальными метриками. Экспериментально показано повышение устойчивости на 30–35% и сокращение времени вычислений почти в 3 раза по сравнению с алгоритмами Дейкстры и A.*

Введение

Транспортные и вычислительные системы требуют поиска маршрутов в условиях динамически изменяющихся параметров. Классические алгоритмы (Дейкстра, A*, Беллмана–Форда), рассчитанные на статические графы, теряют эффективность при изменении весов и топологии [1]. Необходим алгоритм, минимизирующий затраты и устойчиво реагирующий на локальные изменения [2].

I. Математическая постановка задачи

Подходы к оценке устойчивости маршрутов на динамических графах были предложены в работах, где введены количественные характеристики стабильности кратчайших путей при изменении параметров сети [3].

Рассмотрим динамический граф $G(t) = (V, E(t))$, где V – множество вершин, а $E(t)$ – множество рёбер с параметрами, изменяющимися во времени. Ребро $e \in E(t)$ характеризуется весом $w_e(t)$, вероятностью сохранения $P_e(t)$ и качеством $Q_e(t)$.

Оптимальный маршрут $R(t)$, минимизирующий суммарные затраты:

$$C(R(t)) = \sum_{i=1}^{k-1} w_{(v_i, v_{i+1})}(t),$$

где $w_{(v_i,v_{i+1})}(t)$ – вес ребра между вершинами v_i и v_{i+1} в момент t.

Для оценки устойчивости маршрута к изменениям графа вводится метрика:

$$R_{\text{robustness}}(R, R') = 1 - \frac{d(R, R')}{|R|},$$

где R и R' – маршруты до и после изменений, $d(R, R')$ – количество рёбер, отличающихся в двух маршрутах, а $|R|$ – длина начального маршрута.

Для комплексной оценки эффективности маршрута используется функция:

$$E(R) = \alpha D(R) + \beta L(R) - \gamma R_{\text{robustness}}(R, R'),$$

где $D(R)$ – задержка на маршруте, $L(R)$ – длина маршрута, α, β, γ – весовые коэффициенты, регулирующие вклад каждого компонента в итоговую эффективность.

Критерий реоптимизации – сравнение текущего значения эффекта $E(R)$ с показателями нового маршрута R', полученного локальной реоптимизацией. Если локальная реоптимизация улучшает (уменьшает) значение функции E, то происходит обновление маршрута; иначе запускается полный пересчёт маршрута.

Данная постановка обеспечивает баланс между скоростью адаптации маршрута и его устойчивостью к динамическим изменениям параметров графа [4].

Псевдокод алгоритма

```
func robust_update(G_old, G_new, route, thresh):
 (s,e) := find_seg(route, G_old, G_new)
 seg := local_reopt(G_new, route, s, e)
 r := calc_robust(route, seg)
 if r < thresh then
  try:
   route := full_recalc(G_new, route[0], route[-1])
   if no_path then
    route := route
   else:
    route := route[:s] + seg + route[e+1:]
 return route, r
```

Алгоритм обеспечивает быструю адаптацию маршрута при изменениях графа, сохраняя баланс между скоростью и устойчивостью.

II. Анализ результатов

Проведён эксперимент на 100 динамических графах (100–500 вершин, плотность 0.2–0.5, частота изменений 0.1–0.3). Алгоритм сравнивался с Дейкстрой и A* по ключевым метрикам.

На рис. 1 видно, что локальная реоптимизация поддерживает стоимость на стабильном уровне около 5.738, свидетельствуя о быстрой и эффективной адаптации алгоритма к изменениям.

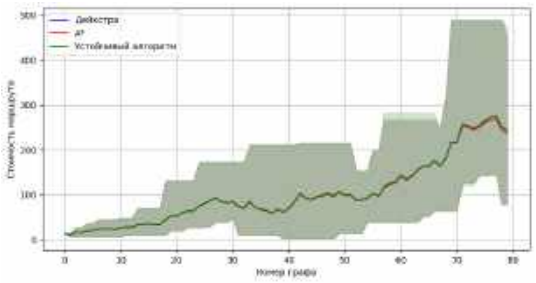


Рис. 1 – Средняя стоимость

Демонстрирация узкого симметричного распределения (см. рис. 2), подтверждает предсказуемость и высокую устойчивость алгоритма.

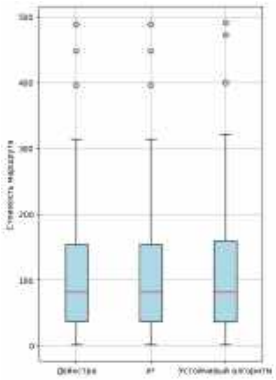


Рис. 2 – Распределение стоимости

Изменение стоимости маршрута во времени (см. рис. 3) отражает плавность и стабильность работы алгоритма при адаптации к динамическим изменениям.

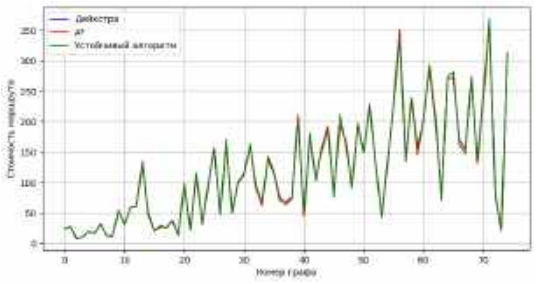


Рис. 3 – Изменение стоимости во времени

Иллюстрация оптимальности путей (см. рис. 4), показывает, как алгоритмы варьируют длину маршрутов на разных графах.

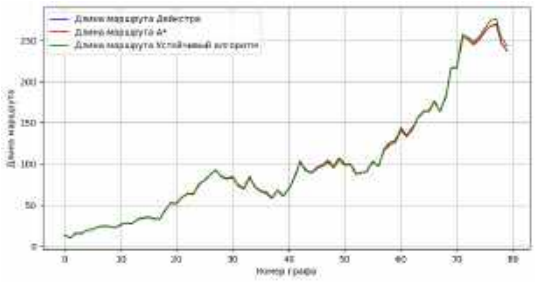


Рис. 4 – Длина маршрутов по графам

На рис. 5 показана производительность, демонстрирующая скорость вычисления оптимального маршрута на выборке динамич. графов.

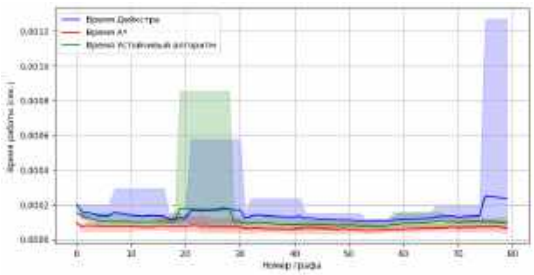


Рис. 5 – Время работы алгоритмов

Метрика устойчивости маршрута стабильно составляет около 0.7409, что на 30-35% выше по сравнению с классическими алгоритмами Дейкстры и A*. Время локальной реоптимизации сокращено почти в 3 раза (около 65%), что говорит о высокой адаптивности и эффективности алгоритма.

В ходе исследований проведено сравнительное тестирование разработанного алгоритма и классических алгоритмов Дейкстры и A* по основным метрикам (см. таблица 1).

Таблица 1 – Производительность алгоритмов

Метрика	Дейкстра	A*	Созданный
Среднее время выполнения (мс)	120	95	40
Средняя длина маршрута	15.8	15.6	15.9
Метрика устойчивости	0.62	0.65	0.85
Успешность реоптимизации (%)	70	75	90

Заключение

Алгоритм обеспечивает сокращение времени обработки до 65% и рост устойчивости на 30–35% без потери качества маршрутов. Небольшое увеличение длины компенсируется стабильностью. Реоптимизация эффективна по скорости, устойчива к изменениям графа и экономит ресурсы, что подтверждено экспериментами.

1. Alwan, Z., et al. A survey of routing in dynamic graphs: challenges and applications // IEEE Communications Surveys & Tutorials. – 2021. – Vol. 23, No. 1. – P. 104–132.
2. Delling, D., Goldberg, A. V., Pajor, T., Werneck, R. F. Customizable Route Planning // Transportation Science. – 2015. – Vol. 49, No. 3. – P. 591–604.
3. Ревотюк, М. П. Оценка устойчивости кратчайших путей на динамически определяемых графах / Ревотюк М. П., Бебех А. В., Хаджинова Н. В. // Информационные технологии и системы 2021 (ИТС 2021) = [translate: Information Technologies and Systems] 2021 (ITS 2021) : материалы международной научной конференции, Минск, 24 ноября 2021 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол. : Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 222–223.
4. Demetrescu, C., Italiano, G. F. A new approach to dynamic all pairs shortest paths // Journal of the ACM. – 2004. – Vol. 51, No. 6. – P. 968–992.

ИММУННЫЙ ПОДХОД К ОБНАРУЖЕНИЮ КОМПРОМЕТАЦИИ В МУЛЬТИАГЕНТНЫХ СИСТЕМАХ

Хаджинова Н. В., Сманцер Н. С., Соркин В. О.
Кафедра информационных технологий автоматизированных систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: khajynova@bsuir.by, {natalya.smantser, sorkindev}@gmail.com

Предложен подход к созданию саморегулируемой мультиагентной системы в целях обнаружения и изоляции скомпрометированных агентов с применением иммунных алгоритмов. В основе подхода лежат принципы искусственных иммунных систем – активно развивающейся области информатики и математического моделирования. Рассмотрены основные принципы работы и механизмы иммунной системы, роль Т-клеток и антигенов, описан алгоритм негативного отбора, предложена структура иммунной подсистемы управления инцидентами в мультиагентной системе.

Введение

Мультиагентные системы (МАС) дают возможность создавать сложные, распределенные и адаптивные решения для таких задач, как управление умными сетями, координация роботов и организация логистических цепочек. Их ключевое преимущество заключается в децентрализованной координации автономных агентов, способных к коллективному решению проблем. Однако эта же децентрализация является источником уязвимости: отказ или злонамеренная компрометация даже одного агента может дестабилизировать работу всей системы. Таким образом, критически важной задачей становится обеспечение саморегулирования МАС, одним из ключевых аспектов которой является своевременное обнаружение и изоляция скомпрометированных агентов.

Решение данной задачи может быть основано на принципе работы иммунной системы человека, которая эффективно решает проблему поиска и нейтрализации чужеродных организмов в распределенной и самоорганизующейся манере. Алгоритмы, основанные на этих принципах, – искусственные иммунные системы (ИИС) – уже успешно зарекомендовали себя в задачах информационной безопасности, таких как обнаружение сетевых вторжений и аномалий [1].

I. Саморегулируемые МАС и применение ИИС

Саморегулирование – это способность системы поддерживать свою работоспособность и целостность в изменяющихся условиях. Саморегулирование включает самодиагностику, адаптацию, самовосстановление. То есть постоянный мониторинг собственного состояния, изменение стратегии поведения при появлении угроз и возврат в стабильное состояние после атаки или сбоя.

Можно провести аналогию между саморегулируемой системой и адаптивной иммунной системой, которая также должна обеспечивать нормальное функционирование организма, одновременно вызывая иммунный ответ на аномальные объекты (антигены). Экспериментальные данные показывают, что толерантность, проявляемая иммунной системой, является результатом динамики и взаимодействия между специфическими регуляторными и эффекторными Т-клетками. Децентрализованный характер этих взаимодействий придает высокую степень устойчивости толерантности и обнаружению неисправностей, без необходимости генетически закрепленной записи о том, как должны выглядеть нормальные ткани [2].

Результаты исследования [1] подтверждают высокую эффективность применения ИИС в составе МАС для обнаружения неисправностей. Использование иммунных алгоритмов, таких как отрицательный отбор, теория опасности, клонирование и иммунная сеть, позволяет системам быстро адаптироваться к новым типам угроз, повышая точность обнаружения. Основные преимущества данного подхода включают способность к самонастройке, автономности и эффективной коммуникации агентов, что обеспечивает устойчивую работу в распределенных средах. Иммунный подход позволяет создавать надежные МАС для, например, управления умными электросетями, где агенты-производители и потребители автономно поддерживают баланс нагрузки и противодействуют кибератакам, или для координации роев дронов, обеспечивая их коллективное поведение и отказоустойчивость.

Данное исследование ограничивается рассмотрением основной способности иммунной системы: распознавать как своих или чужих большое количество антигенов с дальнейшей их классификацией и стимуляцией соответствующих защитных механизмов [3].

II. Алгоритм негативного отбора

В 1994 году Стефани Форрест предложила вычислительную модель способную распознавать «своих» и «чужих», получившую название алгоритм негативного отбора. Целью алгоритма негативного отбора является покрытие чужеродного пространства соответствующим множеством

детекторов. В работе алгоритма есть два этапа: «создание детекторов» и «распознавание чужеродного». На первом этапе создается множество детекторов с применением процесса, использующего в качестве входных данных множество «своих». Детекторы, которые определяют любой из «своих» образцов, исключаются из кандидатов. Правило соответствия обычно смоделировано с использованием модели аффинности (способность распознования) Т-клеток и антигенов. На этапе обнаружения сохраненные детекторы применяются для проверки новых входящих примеров на соответствие «своим» и чужеродным. Если поступивший пример совпадает с детектором, то он идентифицируется как чужеродный [4].

III. Структура ИИС управления инцидентами

Можно предложить структуру подсистемы обнаружения и изоляции аномалий саморегулируемой МАС на основе иммунного алгоритма (см. рис. 1).

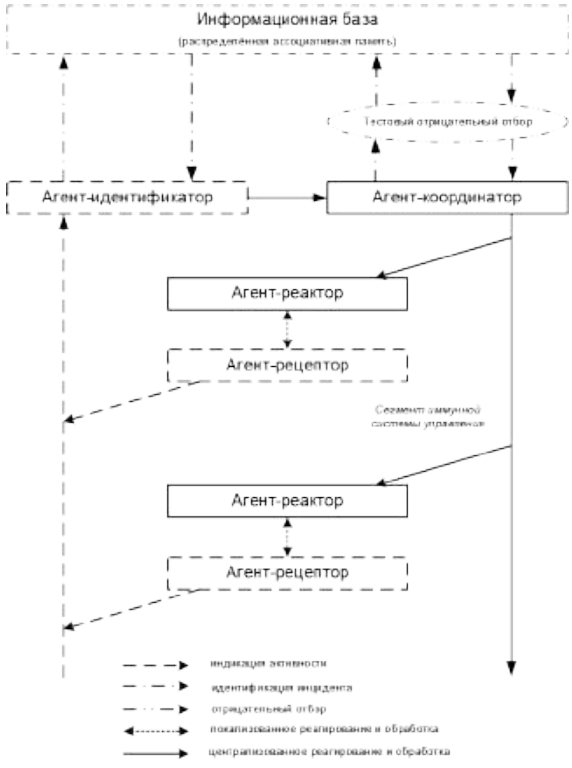


Рис. 1 – Структура ИИС управления инцидентами

Структура включает в себя четыре класса агентов: агенты-рецепторы; агенты-идентификаторы; агенты-координаторы; агенты-реакторы. Агенты-рецепторы соответствуют макрофагам и другим антиген-презентирующим клеткам, которые выставляют частицы антигена на своей поверхности, привлекая внимание В-лимфоцитов для распознавания. Агенты-идентификаторы (детекторы) соответствуют В-лимфоцитам, которые распознают антиген и заранее подвергались «отрицательному отбору» в тимусе. Агенты-координаторы соответствуют лимфокинам, выделяемым Т-лимфоцитами для активации В-лимфоцитов. Агенты-реакторы соответствуют фагоцитам, имеющим антитела для уничтожения антигена.

Заключение

В результате исследования методов построения ИИС предложен вариант структуры иммунной подсистемы управления инцидентами в МАС. В основе предложенной структуры лежат основные принципы работы и механизмы иммунной системы, использована роль Т-клеток и антигенов, механизм отрицательного отбора для генерации агентов-идентификаторов, которые отвечают за распознавание аномалий, «пойманных» агентами-рецепторами.

Таким образом, предложенная иммуноинспирированная архитектура может стать отправной точкой к дальнейшему созданию саморегулируемых МАС для управления сложными объектами, такими как интеллектуальные энергосети или автономные логистические системы. Внедрение таких систем позволит достичь качественно нового уровня отказоустойчивости и автономности, обеспечивая проактивное обнаружение и локализацию аномалий, а также минимизацию каскадных сбоев, что критически важно для функционирования систем в условиях неопределённости и динамически меняющейся среды.

1. Seresht, N. A. MAIS-IDS: A distributed intrusion detection system using multi-agent AIS approach / N. A. Seresht, R. Azmi // Engineering Applications of Artificial Intelligence. – 2014. – Vol. 35. – P. 286–298.
2. Tarapore, D. Abnormality detection in multiagent systems inspired by the adaptive immune system / D. Tarapore [and et al] // Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems. – 2013. – P. 23–30.
3. Гладыш, С. В. Иммунокомпьютинг в управлении инцидентами информационной безопасности / С. В. Гладыш // Штучний інтелект. – 2008. – №1. – С. 12–129.
4. Частикова, В. А. Искусственные иммунные системы: основные подходы и особенности их реализации / В. А. Частикова, Д. А. Картамышев // Научные труды КубГТУ. – 2016. – № 8. – С. 193–208.

АЛГОРИТМЫ СЕГМЕНТАЦИИ ПРИ ОБРАБОТКЕ СПУТНИКОВЫХ СНИМКОВ

Хатеневич А. А., Перцев Д. Ю.
Кафедра электронных вычислительных машин,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: khatenevicha@gmail.com

Задача анализа и обработки спутниковых изображений довольно специфична и узконаправлена, но не менее востребована в условиях активного развития космических технологий. В работе выполнен анализ результатов сегментации изображений несколькими алгоритмами, затем сравнение результатов по выбранной метрике: точность соответствия границ объектов на исходном изображении и на изображении после обработки. Результаты экспериментов выявляют наиболее подходящие алгоритмы сегментации для различных начальных условий при заданных стартовых данных.

Введение

Сегментация в контексте обработки изображений – это процесс разделения цифрового изображения на несколько сегментов, так называемых суперпикселей (наборов пикселей). Цель – изменить изображение таким образом, чтобы упростить процесс анализа и сделать его более эффективным. Метод сегментации также используется для выделения объектов, границ, линий и кривых на изображениях. Результатом является набор сегментов, покрывающих всё изображение, или набор контуров, извлечённых из изображения. Отличительным свойством сегмента является сходство пикселей по одной или нескольким определённым характеристикам (например, цвету, яркости, текстуре). Соседние наборы пикселей, в свою очередь, имеют существенные различия по этим характеристикам.

Разработан ряд алгоритмов и методов для реализации процесса сегментации изображений. Поскольку универсального решения данной задачи не существует, при работе в конкретной предметной области необходимо анализировать существующие методы и понимать специфику данных в этой области, на основе которой разрабатывается решение задачи. Цель данной работы – проанализировать существующие методы в контексте решения задачи сегментации спутниковых изображений, выбрать одну или несколько метрик, которые покажут эффективность алгоритмов, провести сравнительные тесты на реальных данных и затем сравнить алгоритмы по выбранным метрикам.

I. Описание эксперимента

В качестве тестовых данных использовались реальные спутниковые снимки, полученные с помощью оборудования с разрешением 2 метра (1 пиксель соответствует 2 метрам земной поверхности). В эксперименте задействовано 100 изображений различных участков поверхности. Исходные изображения были сформированы в оттенках серого, разрядность составила 10 бит/пиксель, что позволяет получить более детализированное изображение. Для эксперимента были выбраны два алгоритма: метод водораздела и метод заполнения водой. Цель эксперимента – выделить на снимках водные объекты (реки, озера и т. д.). Метрикой оценки методов сегментации будет служить точность соответствия границ объектов на исходных и обработанных снимках. На основе этой метрики будет оцениваться пригодность используемых в эксперименте алгоритмов сегментации для решения поставленной задачи.

Метод водоразделов. Данный метод основан на представлении изображения в виде топографической поверхности, где высоты относительно определённого уровня представлены значениями яркости пикселей [1]. Эта аналогия разбивает изображение на «вершины» и «равнины». Равнины – это области изображения, где разница в яркости пикселей не превышает заданный порог. Вершины наоборот возникают в местах, где абсолютное значение градиента яркости максимально. Вначале определяются координаты глобального минимума яркости на изображении и начинается постепенное «заполнение водой», которое подразумевает последовательное добавление слоёв пикселей к сегментам. В точках соприкосновения этих сегментов (вершинах) строятся разбиения, называемые водоразделами, которые препятствуют объединению отдельных участков в один. Эксперимент начинается с приблизительной оценки водных объектов. Для этого можно использовать бинаризацию, представляющую собой процесс установки определенного порога яркости (в примере порог яркости был равен 105) с последующим присвоением каждому пикселю либо значения 0 (черный) при превышении порога, либо значения 255 (белый) при его снижении. Следующий шаг – удаление любых мелких шумов на изображении с помощью морфологического раскрытия (эрозия, за которой следует дилатация) и удаление любых мелких отверстий в объекте с помощью морфологического закрытия (дилатация, за которой следует эрозия). Остальные области не определены как водные объекты или

фон. Для разрешения этих неоднозначных зон используется алгоритм водораздела.

Метод заполнения водой. Метод заполнения водой — это алгоритм, позволяющий определить и/или изменить смежные значения на изображении на основе их близости к исходной точке [2]. Этот метод используется для выделения областей однородного цвета. Для реализации алгоритма выбирается начальный пиксель и задаётся интервал изменения цвета соседних пикселей относительно исходного. Процесс объединяет пиксели в один сегмент, заполняя их одним цветом, при условии, что они попадают в заданный диапазон. Результатом является сегмент, залитый определённым цветом. Данный алгоритм полезен для заливки областей с небольшим различием цвета и однородным фоном.

При решении задач в контексте обработки спутниковых снимков данный алгоритм может быть полезен, например, при определении границ озёр и поиске точек соприкосновения нескольких водоёмов на карте, поскольку при наличии пересечений они будут окрашены в один цвет. Для корректной работы заливки необходимо передавать координаты пикселей, расположенных в сегментах, границы которых требуется найти. Поэтому для тестирования данного алгоритма в качестве входных данных передавались координаты точек только некоторых областей.

II. Анализ результатов

Метод заполнения водой удовлетворительно решает задачу поиска границ водоёмов. Алгоритм относительно точно выделяет мелкие детали, такие как участки суши, расположенные в середине выделяемых сегментов. При оценке результатов работы можно заметить некоторый шум на выделенных кластерах в виде неокрашенных пикселей в середине кластера. Этот шум устраняется небольшой корректировкой начальных параметров алгоритма. В данном случае цветовой диапазон пикселей, присоединяемых к сегменту, был изменён с [-10, +10] на [-20, +20]. Основным недостатком данного метода является необходимость реализации поиска координат пикселей, которые будут располагаться на необходимых кластерах. Решением является либо вмешательство человека в качестве посредника между процессами сбора данных и их обработки, либо введение дополнительного алгоритма сегментации, который будет выделять необходимые сегменты. Другой недостаток заключается в том, что если в процессе работы данный метод сталкивается с контрастными препятствиями, то кластер, представляющий собой один объект на исходном изображении, будет выбран неверно.

Метод водоразделов при определённых условиях может создавать избыточную сегментацию: места, которые должны быть единым кластером, разбиваются на множество более мелких, что малозаметно в итоговых результатах из-за плотности этих «подкластеров». Однако, в отличие от метода заливки, алгоритм водоразделов не требует промежуточного этапа в виде поиска необходимых координат на изображении, и при этом данный алгоритм, при правильном выборе начальных параметров, не испытывает больших потерь в точности выделения объектов по сравнению с заливкой. Как уже упоминалось ранее, сравнение алгоритмов проводилось по метрике точности границ объектов на исходных изображениях и изображениях после обработки. Это подразумевает, что объекты на исходном изображении принимаются за эталонные, а результаты работы алгоритмов оцениваются по проценту совпадения эталонных объектов с выделенными сегментами. Средний уровень соответствия стандарту при использовании алгоритма заливки составляет 85%, однако в ходе эксперимента потребовалась дополнительная помощь для поиска объектов на изображении и передачи их координат алгоритму. Алгоритм водораздела, напротив, показал соответствие 80—82% без необходимости дополнительной помощи.

На основе полученной информации был сделан вывод, что при сравнении по выбранной в начале эксперимента метрике предпочтительным является метод водораздела. Преимущество этого метода заключается в отсутствии необходимости дополнительной предварительной обработки изображений при сохранении того же уровня точности.

III. Заключение

Проведя исследование алгоритмов в контексте задачи обработки спутниковых снимков, был сделан вывод о преимуществе метода водоразделов в общем случае. Также необходимо обратить внимание, что космическая фотосъёмка – это процесс с большим числом нестабильных условий, к которым постоянно необходимо адаптироваться и минимизировать их последствия. Поэтому оптимальным выбором при решении задач сегментации изображений является комбинация нескольких алгоритмов, каждый из которых будет выполнять те этапы сегментации, на которых он справляется лучше всего. Таким образом, решение об использовании каждого алгоритма должно приниматься исходя из специфики задачи и условий получения каждого изображения.

IV. Список литературы

1. Roerdink, J. The watershed transform: definitions, algorithms and parallelization strategies / J. Roerdink, A. Meijster // Fundamenta Informaticae. – 2000. – Vol. 41 – P. 187–228.
2. Nosal, E.-M. Flood-fill algorithms used for passive acoustic detection and tracking / E.-M Nosal // New Trends for Environmental Monitoring Using Passive Systems. — 2008. -– P. 1–5.

БЕЗОПАСНЫЕ ОБЛАЧНЫЕ ХРАНИЛИЩА БИОМЕДИЦИНСКИХ ДАННЫХ

Хлебанова К. О.
Факультет компьютерного проектирования,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: ksenia.hlebanova@gmail.com

В последние годы объем биомедицинских данных, генерируемых клиническими системами, геномными исследованиями и носимыми устройствами, стремительно возрастает. Это требует разработки архитектур облачных хранилищ, обеспечивающих высокую производительность, масштабируемость и защиту персональных медицинских данных. В работе рассматриваются современные подходы к организации хранения биомедицинских данных в облаке, включая распределённые системы, микросервисные архитектуры и механизмы обеспечения конфиденциальности и целостности информации.

Введение

Биомедицинские данные характеризуются высокой гетерогенностью (большие бинарные объекты – медицинские изображения, большие файлы геномики; структурированные EHR/метаданные; потоковые сенсорные данные), строгими требованиями к конфиденциальности (PHI, GDPR/HIPAA), потребностью в аналитике в реальном времени и архивном хранении [1-2]. Проектируемая архитектура должна обеспечивать:

- масштабируемость по объёму и запросам;
- различную производительность для горячих и холодных данных;
- контроль доступа, шифрование и доказуемый аудит;
- поддержку стандартов обмена (FHIR, DICOM, HL7) и возможность приватных вычислений (secure enclave, TEEs, SMC/HE).

Далее рассматриваются два подхода, реалистичных для облачной среды.

I. Централизованный облачный Data Lake

Централизованный облачный *Data Lake* представляет собой архитектуру, где объектное хранилище (S3-подобного типа) служит единым источником данных – DICOM, геномных (FASTQ/BAM) и электронных медицинских записей (EHR). Метаданные и индексы хранятся в NoSQL или реляционной базе, связанной с FHIR-ресурсами. Поток данных проходит через слой преобразования (ETL/ELT), нормализующий их в аналитические форматы Parquet/ORC.

Аналитический слой (BigQuery, Athena, Spark) обеспечивает масштабируемую пакетную аналитику. Безопасность реализуется с помощью шифрования (at-rest, in-transit), IAM, KMS и аудита. Для оптимизации затрат используется многоуровневое хранение (hot–cold) с автоматическим tiering. DICOM-объекты индексируются через метаданные и миниатюры, геномные данные преобразуются в табличный формат, EHR – в FHIR-структуры [2]. Время отклика системы можно аппроксимировать выражением:

$$L = L_{\text{net}} + L_{\text{meta}} + L_{\text{io}}, \quad L_{\text{io}} \approx \frac{S}{B},$$

где L_{net} – сетевые задержки при передаче данных, L_{meta} – время поиска и обработки метаданных, L_{io} – задержка при чтении объектов из хранилища, зависящая от их размера S и пропускной способности канала B.

Надёжность системы определяется вероятностью доступности данных при k-кратной репликации:

$$A = 1 - (1 - a)^k,$$

где a – индивидуальная надёжность отдельного узла хранения, а A – совокупная доступность объекта. Увеличение числа копий данных повышает отказоустойчивость, но увеличивает издержки хранения.

Стоимость хранения оценивается как:

$$Cost = V_{\text{hot}} c_{\text{hot}} + V_{\text{cold}} c_{\text{cold}} + T_{\text{egress}} c_{\text{egress}},$$

где V_{hot} и V_{cold} – объёмы данных в горячем и холодном уровнях хранения, c_{hot} и c_{cold} – их удельные стоимости, T_{egress} – объём исходящего трафика, а c_{egress} – стоимость передачи. Данная модель позволяет балансировать между производительностью и затратами при больших объёмах медицинских данных.

Преимущества: централизованное управление, готовность к аналитике, эффективное хранение крупных файлов. Недостатки: единая точка отказа, риск утечки PHI, возможные сетевые задержки при высокой нагрузке. Подход эффективен для крупных исследовательских и национальных платформ с централизованным контролем доступа.

II. Федеративная архитектура с безопасными вычислениями

Федеративная (*Data Mesh / Federated*) архитектура строится на распределённой модели, где каждое доменное хранилище (клиника или

лаборатория) сохраняет собственные данные локально или в изолированном облачном пространстве. Глобальный каталог метаданных содержит только описания и минимальные индексы, а взаимодействие между доменами осуществляется через контракты данных и единые API-интерфейсы (FHIR). Центральная оркестрационная платформа маршрутизирует запросы, выполняет авторизацию и обеспечивает виртуализацию данных, не требуя их физического перемещения.

Вычисления над конфиденциальными данными выполняются в доверенных средах (*secure enclaves*) или с применением методов защищённых вычислений – многопартийных протоколов (MPC) и гомоморфного шифрования (HE). Для кросс-доменных исследований применяются федеративное обучение и приватизация данных по модели дифференциальной приватности.

Время выполнения распределённого запроса описывается выражением:

$$T = \max_{i \in [1..n]} (T_{local,i}) + T_{agg} + T_{net},$$

где $T_{local,i}$ – время обработки на домене i, T_{agg} – агрегация результатов, T_{net} – сетевые накладные расходы. Стоимость распределённого запроса определяется как:

$$Cost_{\text{query}} = \sum_i cost_{\text{compute},i} + cost_{\text{orchestration}},$$

где $cost_{\text{compute},i}$ – вычислительная стоимость обработки на каждом домене, а $cost_{\text{orchestration}}$ – накладные расходы на координацию, маршрутизацию и объединение результатов между участниками федерации. Применение ε-дифференциальной приватности позволяет ограничить вероятность раскрытия персональных данных при агрегации.

Основными преимуществами подхода являются высокая степень защиты данных, соответствие юридическим требованиям (данные не покидают юрисдикцию) и распределённый контроль. Недостатками выступают сложность согласования форматов и политик доступа, повышенные задержки при агрегации и более высокие операционные затраты. Такой подход предпочтителен при организации межведомственных или международных биомедицинских систем, где конфиденциальность превалирует над скоростью обработки.

III. Сравнительный анализ и вывод

Централизованный *Data Lake* обеспечивает высокую производительность и удобство аналитической обработки за счёт унифицированного хранилища и зрелых облачных инструментов. Однако он создаёт единую точку отказа и повышенные риски компрометации конфиденциальных медицинских данных. Федеративная архитектура, напротив, распределяет ответственность между доменами, снижает вероятность утечек и лучше соответствует нормативным требованиям (HIPAA, GDPR), но требует сложной оркестрации и имеет большие сетевые задержки.

Сравнение по ключевым параметрам представлено в табл. III.

Таблица 1 – Сравнение архитектурных подходов

Критерий	**Централизованный Data Lake**	**Федеративная архитектура**
Производительность	Высокая, оптимизированные OLAP-запросы	Зависит от числа доменов и сети
Безопасность и приватность	Централизованное управление, высокий риск утечек	Распределённый контроль, минимизация PHI-передачи
Масштабируемость	Простое горизонтальное масштабирование	Ограничено политиками и совместимостью
Соответствие регуляторным требованиям	Ограничено при кросс-юрисдикционном обмене	Полное соответствие локальному законодательству
Операционные затраты	Ниже при большом объёме данных	Выше из-за оркестрации и синхронизации

Оптимальным направлением развития является гибридный подход, объединяющий сильные стороны обоих решений: использование централизованного *Data Lake* для обезличенных или агрегированных данных и федеративных механизмов для чувствительных данных. Такая гибридная архитектура обеспечивает баланс между безопасностью, масштабируемостью и аналитической эффективностью, что делает её наиболее перспективной для будущих облачных платформ хранения и анализа биомедицинской информации.

IV. Список литературы

1. Tahir, A.; Rehman, A.; Iqbal, S. et al. A Systematic Review on Cloud Storage Mechanisms in Healthcare / A. Tahir, A. Rehman, S. Iqbal // Sensors. – 2020. – Vol. 20, № 18. – P. 5392.
2. Hashem, I. A. T.; Yaqoob, I.; Anuar, N. B. et al. The rise of "big data" on cloud computing: Review and open research issues / I. A. T. Hashem, I. Yaqoob, N. B. Anuar // Information Systems. – 2015. – Vol. 47. – P. 98–115.
3. Manco, C.; et al. HEALER: A Data Lake Architecture for Healthcare / C. Manco, et al. // CEUR Workshop Proceedings, Vol. 3379. – 2023. – P. 602
4. Simhadri, S. Y.; Deshpande, L. K. Evaluating Data Mesh and Traditional Data Architectures: Implications for Governance, Scalability, Cost, and Enterprise Adoption / S. Y. Simhadri, L. K. Deshpande // Frontiers in Health Informatics. – 2024. – Vol. 13, № 8. – P. 5095–5106.

УГРОЗА БОТНЕТОВ ДЛЯ IOT-УСТРОЙСТВ

Чаган Н. Ф., Белоусова Е. С.
Кафедра защиты информации,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: chagan_nf@pac.by, belousova@bsuir.by

В материалах доклада рассматривается развитие интернета вещей с представлением подробной статистики по количеству IoT-устройств и прогнозом их увеличения в ближайшие несколько лет. Особо уделяется внимание проблеме кибербезопасности интернета вещей, а именно механизму формирования ботнетов, таких как AISURU, в сетях с IoT-устройствами. На основе оценки последствий кибератак, реализованных посредством ботнета AISURU, описывается проблема необходимости создания подходов и средств по обеспечению безопасности интернета вещей.

Введение

В настоящее время создание моделей будущего, где общество сформировало общественный строй, члены которого живут комфортной жизнью с максимальной автоматизацией и интеграцией информационных систем, уже не представляется чем-то нереальным. В условиях стремительного развития искусственного интеллекта, интернета вещей IoT (Internet of Things) и автоматизированных систем уже представляется возможным построение модели, основанной на объединении физических и виртуальных объектов в единую инфраструктуру – киберфизическую систему, обеспечивающую оптимизацию ресурсов и повышение качества жизни.

Основная часть

Интернет вещей уже активно проникает в различные сферы жизнедеятельности человека. На текущий момент наблюдается экспоненциальный рост умных устройств интернета вещей. По данным сервиса Statista [1], в настоящее время в мире насчитывается порядка 19,8 млрд. подключенных IoT-устройств. Таким образом, количество устройств, которые подключены к сети Интернет, уже превышает более чем вдвое население Земли. С течением времени эта разница будет только увеличиваться в геометрической прогрессии. По прогнозу (рис. 1) к 2030 году таких устройств будет насчитываться более 31,2 млрд [1], что обусловлено стремлением производителей к совершенствованию производимых ими IoT-устройств с целью получения конкурентного преимущества.

Однако вместе с перспективами и возможностями развития IoT неизбежно возникают новые социальные, этические и экономические проблемы. Также развивается зависимость человека от интеллектуальных систем, которая ставит под угрозу конфиденциальность и безопасность персональных данных.

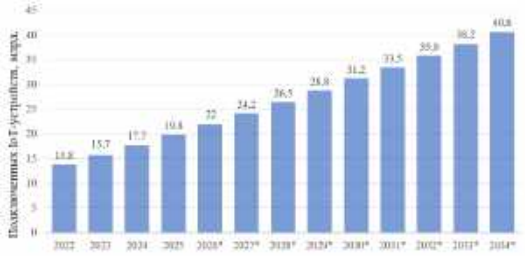


Рис. 1 – Количество устройств, подключений к Интернету вещей по всему миру в период с 2022 по 2024 гг. с прогнозами на период с 2025 по 2034 гг.

Каждый год нарушители получают доступ к миллионам устройств по всему миру, превращая их в ботнеты, или так называемые «зомби-сети». Таким образом нарушители осуществляют свои противоправные действия в отношении организаций, осуществляют кражу персональных данных, нарушают работу важных информационных систем либо манипулируют процессами. Любое устройство, подключенное к сети Интернет, может стать элементом ботнета без ведома пользователя. Происходит это скрытно, в процессе функционирования взломанного устройства, которое при этом выполняет указания киберпреступников.

Создание ботнета является лишь шагом к достижению конечной цели нарушителя. Сформировав мощную инфраструктуру, появляются способы не только проводить различные кибератаки, такие как DDoS, спам, кража данных и т.д., но и заниматься процессом масштабирования ботнета с целью последующей продажи или аренды другим нарушителям.

Примером зафиксированной DDoS-атаки является ботнет AISURU. По оценкам исследователей QiAnXin XLab [2], в данном ботнете оказались порядка 300 тыс. маршрутизаторов по всему миру. В основном использовались уязвимости N-day в IP-камерах и маршрутизаторах D-Link, Linksys, Zyxel, SDK Realtek, а именно уязвимости CVE-2017-5259 [3] в устройствах Cambium, CVE-2023-28771 [4] в оборудовании Zyxel, CVE-2023-50381 [5] в SDK Realtek Jungle. К ботне-

ту AISURU оказались подключенными видеорегистраторы nvms9000, ранее относившиеся к другому ботнету RapperBot. Для повышения пропускной способности трафика операторы задействовали GRE-туннели, которые были установлены на четырёх командных узлах.

Стоит учитывать, что весной 2025 года ботнет AISURU проводил волну кибератак [2] мощностью 5,8 Тбит/с, в сентябре мощность увеличилась почти вдвое и достигала 11,5 Тбит/с, в октябре произошло очередное более чем двукратное увеличение мощности атаки – 29,69 Тбит/с, что на данный момент является новым рекордом.

Потенциальные жертвы для формирования ботнетов выбираются случайным образом. В настоящее время отсутствуют свидетельства того, что целью нарушителя является какая-либо конкретная отрасль или страна. География кибератак затрагивает организации таких стран, как Китай, США, Великобритания, Германия. Несмотря на это, можно смело сказать, что защита крупнейших игровых и облачных платформ в настоящее время находится под угрозой кибератак данного или аналогичного ботнета, так как атака, проведенная 6 октября 2025 года, которая затронула игровые сервисы Steam (рис. 2), Xbox (рис. 3), Riot Games, PlayStation Network и другие популярные онлайн-сервисы согласно информации онлайн системы мониторинга сбоев онлайн-сервисов Downdetector [6].

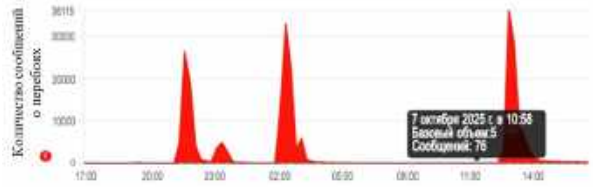


Рис. 2 – Количество сообщений о перебоях в работе сервиса Steam в сравнении с базовым объемом

Рис. 3 – Количество сообщений о перебоях в работе сервиса Xbox в сравнении с базовым объемом

В результате кибератаки по всему миру пользователи испытывали трудности с входом в их игровые аккаунты, подключением к игровым магазинам и серверам. Геймеры жаловались на невозможность подключения к таким платформам, как Counter-Strike, Dota 2, Valorant, League of Legends и др.

Интересным является подход к маскировке ботнета. Вредоносное программное обеспечение проверяет наличие анализаторских инструментов или виртуальных сред и прекращает свое функционирование при их обнаружении. Также используются TCP Carpet Bomb-атаки [7], имитирующие легитимный трафик. С марта 2025 года нарушители внедрили два крупных обновления, направленных на усиление шифрования и изменение сетевого протокола. Одной из характерных черт AISURU является использование модифицированного алгоритма RC4 со статическим ключом PJbiNbbeasddDfsc.

Несмотря на то, что AISURU является наглядным примером злонамеренного использования IoT-устройств, он далеко не единственный в своем роде. В даркнете продолжают формироваться и другие ботнеты. Эти сети действуют незаметно, накапливая ресурсы и охватывая все больше уязвимых устройств. Неизвестно, когда именно, при каких условиях и с какой целью они будут активированы, что делает угрозу еще более тревожной и непредсказуемой.

Заключение

Ботнеты не являются принципиально новой угрозой, но в то же время выходят за рамки обычного инструмента для целенаправленных атак на IoT-устройства и представляют собой многофункциональную платформу. Для эффективного противодействия данной нарастающей угрозе требуются конкретные оперативные действия в сфере кибербезопасности по созданию подходов и средств, позволяющих обеспечить защиту информации, в том числе совместные действия международного сообщества и правоохранительных органов.

1. Компания Statista [Электронный ресурс]. – Режим доступа: `https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/`. – Дата доступа: 15.10.2025
2. Компания XLab [Электронный ресурс]. – Режим доступа: `https://blog.xlab.qianxin.com/super-large-scale-botnet-aisuru-en/`. – Дата доступа: 15.10.2025
3. The National Institute of Standards and Technology (NIST). – [Электронный ресурс]. – Режим доступа: `https://nvd.nist.gov/vuln/detail/CVE-2017-5259`. – Дата доступа: 15.10.2025
4. The National Institute of Standards and Technology (NIST). – [Электронный ресурс]. – Режим доступа: `https://nvd.nist.gov/vuln/detail/CVE-2023-28771`. – Дата доступа: 15.10.2025
5. The National Institute of Standards and Technology (NIST). – [Электронный ресурс]. – Режим доступа: `https://nvd.nist.gov/vuln/detail/CVE-2023-50381`. – Дата доступа: 15.10.2025
6. Компания Downdetector [Электронный ресурс]. – Режим доступа: `https://downdetector.com`. – Дата доступа: 15.10.2025
7. Российский портал SecurityLab.ru [Электронный ресурс]. – Режим доступа: `https://www.securitylab.ru/glossary/carpet_bombing/`. – Дата доступа: 15.10.2025

РЕШЕНИЕ СЛОЖНЫХ ВЫЧИСЛИТЕЛЬНЫХ ЗАДАЧ НА БАЗЕ КЛАСТЕРНОГО КОМПЬЮТЕРА

Черемисинова Л. Д.
Объединённый институт проблем информатики Национальной академии наук Беларуси
Минск, Республика Беларусь
E-mail: cld@newman.bas-net.by

Описывается технология распараллеливания алгоритмов решения логико-комбинаторных задач, возникающих при автоматизации логического проектирования СБИС, с целью их реализации на многопроцессорных вычислительных системах кластерного типа.

Введение

Логико-комбинаторные задачи характеризуются тем, что объемы перерабатываемой информации при поиске их решений относительно невелики, а сами процессы переработки достаточно сложны, так как требуют, как правило, перебора и анализа значительного числа вариантов промежуточных решений. В силу этого становится проблематично решать такие задачи большой размерности за приемлемое время даже с использованием самых быстродействующих персональных компьютеров. Для решения таких задач все чаще используют многопроцессорные вычислительные системы, наиболее популярным типом которых являются кластеры.

В настоящей работе предлагается технология выполнения распараллеливания алгоритмов решения логико-комбинаторных на примере выполнения верификации результатов проектирования в рамках САПР логического проектирования заказных КМОП СБИС на основе удаленного решения на кластере задачи нахождения корней логических уравнений.

I. Логические уравнения и поиск их корней

Логическими уравнениями называются выражения вида $F(X) = 1$ и $F(X) = 0$, где $F(X)$ – формула булевой алгебры. Одним из важнейших классов логических уравнений являются уравнения, в которых формула $F(X)$ задана в виде КНФ K или ДНФ D: $K = 1$, $D = 0$ (задачи двойственны).

Широкий класс практически значимых задач, связанных с управлением и обработкой информации в дискретных системах, а также задач синтеза и верификации в микроэлектронике и других, сводится к проблеме булевой выполнимости КНФ или проверке ДНФ на тавтологию. Эти ключевые операции часто фигурируют во внутренних циклах алгоритмов, решающих эти задачи. Проблема выполнимости КНФ К состоит в проверке существования корня уравнения $K = 1$, т. е. такого набора значений переменных, который обращает в 1 КНФ (как и каждый ее дизъюнкт) и называется выполняющим. Задачу выполнимости КНФ называют в литературе SAT-задачей.

В матричной интерпретации задача сводится к задаче анализа на выполнимость троичной матрицы $\boldsymbol{T}$, строки которой задают дизъюнкты исходной КНФ. Проверить троичную матрицу $\boldsymbol{T}$ на выполнимость означает найти троичный вектор $\boldsymbol{t}$, который обращает в 1 каждый дизъюнкт. Или убедиться в том, что такого вектора не существует (когда матрица $\boldsymbol{T}$ не выполнима, т. е. $K = 0$). Анализ матрицы $\boldsymbol{T}$ производится методом обхода дерева поиска выполняющего вектора $\boldsymbol{t}$. Текущая ситуация, представляемая в узле дерева, характеризуется значением вектора $\boldsymbol{t}$, отдельные компоненты которого определены (т.е. имеют значение 0 или 1), а другие (отмеченные «–») подлежат доопределению в процессе движения по дереву. Троичный вектор $\boldsymbol{t}$ однозначно определяет минор $\boldsymbol{T}$ ($\boldsymbol{t}$) матрицы $\boldsymbol{T}$, расположенный на пересечении ее строк, соответствующих невыполнимым относительно вектора $\boldsymbol{t}$ дизъюнктам, и столбцов, отмеченных его неопределенными компонентами.

II. Распараллеливание задачи поиска корней логических уравнений

Распараллеливание алгоритмов решения SAT-задач может производиться на основе динамического или статического распределения нагрузки процессоров. Для систем кластерного типа более подходит метод статического распределения нагрузки процессоров, который не требует интенсивного взаимодействия узлов. Он предполагает декомпозицию исходной SAT-задачи на семейство непересекающихся подзадач, каждая из которых обрабатывается SAT-решателем на отдельном вычислительном узле вычислительной системы.

В основе реализованного в настоящей работе метода лежит идея гибридного распределения нагрузки между подчиненными процессами, когда сначала вся нагрузка делится, а затем освободившиеся процессоры подключаются для решения подзадач, решаемых загруженными процессорами.

Метод распараллеливания с предварительным распределением нагрузки основан на расщеплении анализируемых булевых формул, ис-

пользуя разложение Шеннона по k переменным. Задача разбивается на не более чем $l = 2^k$ подзадач (не более, так как некоторые из подзадач могут иметь тривиальное решение, не требующее обсчета на процессоре).

Полученные частные задачи решаются параллельно l процессорами решающими свою задачу над своими исходными данными до конца, а корневой процессор собирает результаты частичных решений и формирует из них итоговое решение общей задачи. Если хотя бы для одной частной задачи найдется решение, то по нему однозначно находится набор значений переменных, выполняющих исходную КНФ. Хотя время решения задачи существенно зависит от выбора переменных для разложения, используется тривиальная процедура разложения, так как она выполняется целиком на корневом процессоре в режиме последовательного выполнения.

III. Программа решения логических уравнений на кластерном компьютере

Метод решения логических уравнений реализован MPI-программой «Вычислитель» для кластерного компьютера [1]. Исходными данными для программы «Вычислитель» служит описание КНФ в текстовом виде или в формате DIMACS, который представляет собой стандартный интерфейс для SAT-решателей. Результатом выполнения программы является текст, содержащий режим и длительность решения, а также выполняющий вектор, если КНФ выполнима.

В мультипроцессорных системах с индивидуальной памятью текст MPI-программы перед исполнением копируется в память каждого из обрабатывающих процессов. Сама параллельная MPI-программа делится на общие, корневые и подчиненные участки, исполняемые всеми, корневым и подчиненными процессами. Для организации обмена сообщениями предусматриваются специальные приемо-передающие буферы. Для рассматриваемой задачи через них передаются векторы, задающие миноры и выполняющие наборы, сигналы окончания работы. При реализации программы «Вычислитель» был выделен минимально необходимый набор MPI-операций: для запуска и завершения работы, определения выделенной конфигурации кластера, выделения и распознавания блоков программы для процессов, работы с приемо-передающими буферами (опроса, приема и посылки), операций обмена сообщениями через буфера.

Распределение нагрузки в MPI-программе «Вычислитель» производится корневым процессом динамически через магазин миноров по мере их формирования при обходе дерева поиска. Минор матрицы $\boldsymbol{T}$ задается вектором $\boldsymbol{t}$. Когда какой-либо подчиненный процесс при достраивании вектора $\boldsymbol{t}$ в процессе обхода своего поддерева поиска: 1) находит искомый вектор $\boldsymbol{t}$(выполняющий все строки обрабатываемого минора), тогда задача решена и все остальные процессы прекращают работу, 2) убеждается в невыполнимости минора, тогда он освобождается и получает очередной минор из магазина или минор, полученный (по запросу корневого процесса) на очередном шаге обхода дерева поиска некоторым занятым процессом (делит с ним нагрузку). Поиск выполняющего вектора $\boldsymbol{t}$ продолжается до тех пока, пока он ни будет найден (КНФ выполнима) или пока все подчиненные процессы не закончат работу (КНФ не выполнима).

IV. Интеграция программы решения логических уравнений в маршрут проектирования

MPI-программа «Вычислитель» в целом является частью трехслойной системы, состоящей из программы для суперкомпьютера, сервиса связи и агента пользователя. Протокол работы грид-сервиса состоит из следующих шагов:1) транспортировки файла с анализируемой КНФ в файловую систему на управляющей машине кластера; 2) запуска программы «Вычислитель»; 3) обратной транспортировки файла с результатами решения на клиентский компьютер.

Методика запуска на кластерном компьютере программ решения комбинаторных задач, возникающих в процессе проектирования интегральных схем, отработана на примере разработанного в лаборатории логического проектирования ОИПИ НАН Беларуси программного комплекса ЭЛС «Энергосберегающий логический синтез» [2]. Программа верификации из ЭЛС была модифицирована следующим образом: после получения КНФ в процессе выполнении операции верификации в среде комплекса ЭЛС файл с анализируемой КНФ, заданной в формате Dimacs, транспортируется по сети на управляющую машину кластера; запускается программа «Вычислитель» на суперкомпьютере; файл с результатами решения возвращается в систему ЭЛС. Если файл содержит найденный вектор (что говорит о том, что КНФ выполнима), делается вывод о неэквивалентности сравниваемых описаний, иначе описания эквивалентны.

V. Список литературы

1. Черемисинов, Д. И. Использование параллельных вычислений при автоматизированном проектировании СБИС / Д. И. Черемисинов, Л. Д. Черемисинова // Проблемы разработки перспективных микро- и наноэлектронных систем – 2016. Сборник трудов / под общ. ред. академика РАН А. Л. Стемпковского. М. : ИППМ РАН, 2016. Часть I. – С. 32-39.
2. Бибило, П. Н. Автоматизация логического синтеза КМОП схем с пониженным энергопо-треблением / П. Н.Бибило, Л. Д.Черемисинова, С. Н. Кардаш и др. // Программная инженерия. – 2013. – № 8. – С. 35-41.

STATISTICAL ANALYSIS OF PIXEL DISTRIBUTION FOR IMAGE DISTORTION DETECTION

Wang Jing, Yu.German
Department of Information Technologies in Automated Systems,
Belarussian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: qiaonu38@gmail.com, jgerman@bsuir.by

This study investigates the use of statistical methods to detect image distortions by analyzing pixel distributions. Using Kolmogorov–Smirnov (KS) tests and moment-based comparisons across varying sample sizes, we compare normal and distorted images to determine the minimal sample size required for reliable discrimination. Our findings demonstrate that statistical pixel analysis provides a computationally efficient alternative to traditional image quality metrics, particularly suitable for real-time applications where processing resources are limited.

Introduction

Image quality assessment is essential in digital imaging and computer vision applications ranging from medical imaging to autonomous systems. Traditional full-image analysis methods often require substantial computational resources, making them impractical for real-time applications. This research evaluates whether statistical pixel distribution analysis can distinguish between normal and distorted images efficiently, with a focus on sample size optimization. By analyzing partial pixel data rather than complete images, we aim to develop a lightweight yet effective distortion detection methodology that balances accuracy with computational efficiency.

The significance of this work lies in its potential applications across various domains including remote sensing, medical diagnostics, and industrial quality control. Current approaches often rely on computationally intensive methods that analyze entire images, limiting their applicability in resource-constrained environments. Our approach addresses this limitation by exploring the minimal data requirements for reliable distortion detection. This research builds upon established statistical theory while adapting it to the specific challenges of digital image analysis.

Methodology

We developed a Python-based analyzer to sample pixels from grayscale images and compute statistical features—mean, variance, skewness, kurtosis, and moments—across sample sizes of 10 to 500 pixels. The KS test was used to evaluate distribution similarity, with 10 trials per sample size for reliability. Our experimental setup included 10 different distortion types including Gaussian noise, motion blur, and compression artifacts, each compared against reference images under controlled conditions. The analysis framework employed random sampling to ensure unbiased representation of image regions.

The experimental design incorporated a systematic evaluation of statistical power across different sample sizes. For each combination of sample size and distortion type, we conducted multiple independent trials to assess consistency and reliability. The images were pre-processed to ensure consistent lighting conditions and normalized intensity ranges, eliminating potential confounding factors in our analysis. Additionally, we implemented a bootstrapping technique to estimate confidence intervals for our statistical measures, providing a more robust assessment of the method's reliability across different image types and distortion levels.

Experimental Setup and Analysis

The analysis followed a structured methodology with specific parameters: Sample Sizes Tested: 10, 50, 100, 200, 500 pixels; Number of Trials: 10 for each sample size; Statistical Significance Level: $\alpha = 0.05$; Images Analyzed: 10 distorted images compared against a reference normal image. Each analysis was repeated 10 times for statistical reliability, providing a robust dataset for evaluating the relationship between sample size and detection accuracy.

The statistical validation methods included accuracy analysis and optimal sample size determination algorithms. We implemented additional validation through the analyze_sample_size_accuracy method, which calculated accuracy for each sample size by comparing predictions with ground truth assumptions. The determine_best_sample_size method identified the optimal sample size based on ρ-value comparisons, ensuring statistically sound recommendations for practical applications.

Key Findings

Sample Size Impact: Small samples (e. g., 10-50 pixels) yield high variability and unreliable KS results due to insufficient representation of overall distribution characteristics. Samples of 100–500 pixels improve discrimination power and stability, with diminishing returns observed beyond 200 pixels. The relationship between sample size and detection accuracy followed a logarithmic pattern, with rapid improvement up to 100 pixels and gradual refinement thereafter. Interestingly, extremely large sample sizes (above 500 pixels) showed minimal additional

benefit while significantly increasing computational overhead.

KS Test Effectiveness: The test consistently identified significantly different distributions for heavily distorted images, especially with larger samples. For subtle distortions, the KS test required sample sizes above 100 pixels to achieve statistical significance ($p < 0.05$). The test demonstrated particular sensitivity to distribution shape changes, outperforming traditional mean-based comparisons for non-Gaussian distortions. We observed that the KS statistic showed strong correlation with human perceptual ratings of image quality, suggesting its potential as a proxy for subjective quality assessment.

Statistical Moments: Distorted images showed altered skewness and kurtosis, indicating distribution shape changes. Moment analysis revealed that variance differences provided the most consistent discrimination across distortion types, while higher-order moments were particularly sensitive to specific artifacts like salt-and-pepper noise. The combination of multiple statistical features created a robust feature space for distortion classification. Notably, the third and fourth moments (skewness and kurtosis) proved especially valuable for identifying asymmetric distortions and outlier-prone distributions respectively.

Optimal Sample Size

We recommend 100–200 pixels for balancing accuracy and computational cost. This range reliably differentiates image types without excessive processing, achieving 85-92% discrimination accuracy across various distortion types. The 200-pixel sample size consistently provided the best trade-off, offering robust performance while maintaining computational efficiency suitable for embedded systems and real-time processing applications.

Further analysis revealed that the optimal sample size varies slightly depending on image content and distortion characteristics. For high-frequency images with complex textures, samples towards the upper end of this range (150-200 pixels) provided more reliable results. Conversely, for smooth images with uniform regions, samples as small as 100 pixels achieved comparable performance. We also found that adaptive sampling strategies, which allocate more samples to regions with higher local variance, could further optimize the trade-off between accuracy and efficiency.

Conclusion

Statistical pixel analysis, particularly using KS tests and moment comparisons, is effective for detecting image distortions. Sample size plays a critical role, with medium-sized samples offering the best trade-off between discrimination accuracy and computational requirements. The methodology demonstrates particular promise for applications in resource-constrained environments such as mobile devices and IoT systems. Future work may include adaptive sampling strategies that dynamically adjust sample size based on image complexity and machine learning integration for improved distortion classification.

The practical implications of this research extend to real-time quality monitoring in manufacturing, automated image curation in social media platforms, and resource-efficient medical image analysis. By establishing the minimal data requirements for reliable distortion detection, our work provides a foundation for developing more efficient computer vision systems that can operate within strict computational constraints. The statistical approach outlined here offers a principled alternative to heuristic methods, with well-understood theoretical properties and predictable performance characteristics across diverse application scenarios.

References

1. Kolmogorov, A. N. (1933). Sulla determinazione empirica di una legge di distribuzione. *Giornale dell'Istituto Italiano degli Attuari*.
2. Smirnov, N. V. (1948). Table for estimating the goodness of fit of empirical distributions. *Annals of Mathematical Statistics*.
3. Press, W. H., et al. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.
4. Gonzalez, R. C., & Woods, R. E. (2018). *Digital Image Processing*. Pearson.
5. Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*.
6. Wang, Z., & Bovik, A. C. (2006). Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*.

I-OPTIMAL DESIGN FOR AFFINE MULTIPLE REGRESSION

Mukha V. S.
Department of Information Technologies of Automated Systems,
Belarusian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: mukha@bsuir.by

This article is devoted to the development of the experiment design theory. The new so called I-optimality criterion is substantiated and I-optimal design is developed which do not inferior to the known designs in the sense of known-optimality and has no restrictions on the number of the observations.

Introduction

The optimal designs of experiment use widely for identification of the subjected control objects in many technical and technological systems and processes [1]. This implies the further development of the theory in this area. We will consider the regression experiment for the affine multiple regression.

It is now traditionally accepted to represent the mathematical model of the observations for the affine multiple regression function in the following form [2]:

$$y_{o,i} = \theta_* + x_i^T\Theta + \epsilon_i,\ \ i = 1, 2, \ldots, n,\ \ m+1 \le n,$$

or

$$y_{o,i} = H_i^T\Theta_1 + \epsilon_i,\ \ i = 1, 2, \ldots, n,\ \ m+1 \le n, \quad (1)$$

where $x_i^T = (x_{i,1}, x_{i,2}, \ldots, x_{i,m}) \in E^m$, $i = 1, 2, \ldots, n,$, are input vector variables, so called factors, $\Theta^T = (\theta_1, \theta_2, \ldots, \theta_m)$, $\Theta_1^T = [\theta_*, \Theta^T]$ is the vector-rows of the unknown coefficients, $H_i^T = [1, x_i^T] \in E^{m+1}$ is the vectors of the basis functions, ϵ_i are uncorrelated random errors with zero mean value and variation σ^2 , $y_{o,i}$ is observed values of the output variable, n is the size of the sample. The factors $x_{i,j}$ are subjected to the constraints in the form $-1 \le x_{i,j} \le 1$. Such factors are called normalized. The expression (1) can be written also in the vector-matrix form as

$$Y = F\Theta_1 + E,$$

where

$$F = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,m} \\ 1 & x_{2,1} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,m} \end{pmatrix}, Y = \begin{pmatrix} y_{o,1} \\ y_{o,2} \\ \vdots \\ y_{o,n} \end{pmatrix},$$

$$E = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}, \Theta_1 = \begin{pmatrix} \theta_* \\ \theta_1 \\ \vdots \\ \theta_m \end{pmatrix},$$

The vector $\hat{\Theta}_1$ of the LSE-estimations of the vector Θ_1 is defined by the expression

$$\hat{\Theta}_1 = (F^TF)^{-1}F^TY, \quad (2)$$

and the covariance matrix of the estimation (2) is defined by the following expression [2]:

$$V_{\hat{\Theta}_1} = \sigma^2(F^TF)^{-1}.$$

The set of the points $x_i^T = (x_{i,1}, x_{i,2}, ..., x_{i,m}) \in E^m$, $i = 1, 2, ..., n$, is called a design of the experiment and is denoted $\xi_m(n)$. The design $\xi_m(n)$ is represented in the form of the following $(n \times m)$-matrix:

$$\xi_m(n) = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,m} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,m} \end{pmatrix}. \quad (3)$$

The number of the rows of the matrix $\xi_m(n)$ (3) not less than the number of the columns. The matrix $A = F^TF$ is called the information matrix of the design $\xi_m(n)$ [2]. The inverse matrix $A^{-1} = (F^TF)^{-1}$ is called the covariance matrix of the design $\xi_m(n)$. The task of the optimal experiment designing consists of the finding the design $\xi_m(n)$ satisfying the given constraints and optimal in some sense.

There are a number of optimality criteria for the designs [2]. A wide set of the optimal designs there exists, but they have the constraints on the numbers of the experiments (points). It is desirable that the researcher has the complete freedom in choosing the number of the experiments.

We introduce in this work the new optimality criterion which allows us to develop the designs for arbitrary number of observations.

I. The model of the observations and the I-optimal design

We will consider the observations model for the affine multiple regression function in opposition to the model (1) in the following form:

$$y_{o,i} = \theta_0 + (x_i - \overline{x})^T\Theta + \epsilon_i,\ \ i = 1, 2, ..., n,\ \ m+1 \le n, \quad (4)$$

where $x_i^T = (x_{i,1}, x_{i,2}, ..., x_{i,m}) \in E^m$, $y_{o,i}, \theta_0, \epsilon_i \in R$, $\Theta^T = (\theta_1, \theta_2, ..., \theta_m) \in E^m$,

$$\overline{x}^T = \frac{1}{n}\sum_{i=1}^{n} x_i^T = (\overline{x}_1, \overline{x}_2, ..., \overline{x}_m),$$

θ_0, Θ are unknown coefficients of the regression function, ϵ_1 is the uncorrelated random error with zero

mean value and variation σ^2 , x_i is the point of the experiment, n is the size of the sample (the number of the experiments, the number of the design points, the number of runs), $y_{o,i}$ is the observed value of the output variable. The LSE-estimations $\hat{\theta}_0, \hat{\Theta}$ of the coefficients θ_0, Θ are defined by the expressions

$$\hat{\theta}_0 = (\sum_{i=1}^{n} y_{o,i}^T)/n = \overline{y}, \;\; \hat{\Theta} = \mu_{xx}^{-1}\mu_{xy}, \qquad (5)$$

where

$$\mu_{xx} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(x_i - \overline{x})^T,$$

$$\mu_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})y_{o,i}.$$

If we denote $\dot{x}_i = x_i - \overline{x}$ the centered vector x_i , $\dot{y}_i = y_{o,i} - \overline{y}$ the centered observation $y_{o,i}$,

$$\dot{\xi}_m(n) =$$

$$= \begin{pmatrix} x_{1,1} - \overline{x}_1 & x_{1,2} - \overline{x}_2 & \cdots & x_{1,m} - \overline{x}_m \\ x_{2,1} - \overline{x}_1 & x_{2,2} - \overline{x}_2 & \cdots & x_{2,m} - \overline{x}_m \\ \vdots & \vdots & \vdots & \vdots \\ x_{n,1} - \overline{x}_1 & x_{n,2} - \overline{x}_2 & \cdots & x_{n,m} - \overline{x}_m \end{pmatrix} \qquad (6)$$

the centered design $\dot{\xi}_m(n)$, then

$$\mu_{xx} = \frac{1}{n}\dot{\xi}_m^T(n)\dot{\xi}_m(n), \;\; \mu_{xy} = \frac{1}{n}\dot{\xi}_m^T(n)\dot{y}, \qquad (7)$$

and the expression (5) for the estimation $\hat{\Theta}$ takes the following form:

$$\hat{\Theta} = (\dot{\xi}_m^T(n)\dot{\xi}_m(n))^{-1}\dot{\xi}_m^T(n)\dot{y}$$

by analogy with the expression (2).

We will call the matrix (5) (or (7)) the information matrix of the centered design (6) (or of the design $\xi_m(n)$ (3)). The variation $V_{\hat{\theta}_o}$ of the estimation θ_0 and the covariance matrix $V_{\hat{\Theta}}$ of the estimation $\hat{\Theta}$ are defined by the following expressions [3]:

$$V_{\hat{\theta}_o} = \frac{\sigma^2}{n}, \;\; V_{\hat{\Theta}} = \frac{\sigma^2}{n}\mu_{xx}^{-1}.$$

The estimations $\hat{\theta}_0$ and $\hat{\Theta}$ are uncorrelated [3], in opposite to the coefficients of the traditional model (1), what is the advantage of the model (4) .

We will call the design $\xi_m(n)$ (**??**) I-optimal if it provides the maximal value of the trace of the information matrix μ_{xx}. It is can to show that the trace of the information matrix μ_{xx} is represented in the following form:

$$tr(\mu_{xx}) = \frac{1}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}||x_i - x_j||^2. \qquad (8)$$

The problem of building of I-optimal design is formulated as follow: find the design $\xi_m(n)$ (3), providing a maximum for the function $tr(\mu_{xx})$ (8), or for the function $I = n^2 tr(\mu_{xx})$:

$$I = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n}||x_i - x_j||^2. \qquad (9)$$

provided the points of the design are pairwise distinct, do not lie on the same linear variety in E^m, satisfy the constrains

$$a \le x_i \le b, \;\; i = 1, 2, ..., n,$$

$a = (a_1, a_2, ..., a_m) \in E^m$, $b = (b_1, b_2, ..., b_m) \in E^m$, and the number of the points $n \ge m + 1$.

As a result, we propose the following formalized algorithm for synthesis of the -optimal design.

Let k be the remainder of dividing n by 4. If $n < 4$ $(m < n - 1 = 3)$, then the saturated nonsingular design $S_m(m + 1)$ (10) below augmented with $n - m - 1$ points in accordance with criterion I-optimality (9) is used:

$$S_m(m+1) = \begin{pmatrix} b_1 & b_2 & \cdots & b_m \\ b_1 & a_2 & \cdots & a_m \\ \vdots & \vdots & \vdots & \vdots \\ a_1 & a_2 & \cdots & b_m \end{pmatrix} = (s_{i,j}), \qquad (10)$$

$$s_{1,j} = b_j, \;\; i = 1, \;\; j = 1, 2, ..., m,$$

$$s_{i,j} = \begin{pmatrix} b_j, & j = i - 1 \\ a_j, & j \ne i - 1 \end{pmatrix}, i = 2, 3, ..., m + 1.$$

If $n \ge 4$, then provided $n - k \ge m + 1$ and $n - k = o \bmod 4$ the Plackett–Burman design $PB_m(n - k)$ is used [4], otherwise the design $S_m(m+1)$ (10) augmented with $n-m-1$ points in accordance with criterion I -optimality (9) is used.

II. References

1. Rachel T. Johnson, Douglas C. Montgomery, Bradley A. Jones. An Expository Paper on Optimal Design. Quality Engineering, 23:287–301, 2011. DOI: 10.1080/08982112.2011.576203.
2. Ermakov S. M., Zhigljavsky A. A. Mathenatical Theory of the Optimal Design. Moscow, Science, 1987, 320 p. (in Russian).
3. Mukha V. S. Multidimensional-matrix Linear Regression Analysis: Distributions and Properties of the Parameters Estimations. Proceedings of the National Academy of Sciences of Belarus. Physics and Mathematics series. 2014, no. 2, pp. 71-81 (in Russian).
4. Plackett R. L., Burman J.P. The Design of Optimum Multifactor Experiments. Biometrika, 1946, 33, № 4, pp. 305-325. DOI: 10.2307/2332195.

STATISTICAL EVALUATION OF NETWORK TRAFFIC VARIATIONS USING HYPOTHESIS TESTING METHODS

Orazdurdyyeva G. O., Bekiyeva M. B., Bekiyev A. R.
Department of Computer Sciences and Information Technologies,
Oguz han Engineering and Technology University of Turkmenistan
Department of Applied Mathematics and Informatics Oguz han Engineering and Technology University of Turkmenistan
Faculty of Engineering and Economics,
Belarusian State University of Informatics and Radioelectronics
Ashgabat, Turkmenistan; Minsk,Belarus
E-mail: gulshatorazdurdyyewa3@gmailcom, maral.bekiyeva@etut.edu.tm, successbmb@gmail.com

This paper presents a statistical approach for detecting and analyzing variations in network traffic using hypothesis testing methods such as the z-test, t-test, and chi-square test. The study aims to determine whether changes in traffic behavior are statistically significant and could indicate potential cyberattacks or anomalies. The proposed approach provides an interpretable, mathematically grounded framework for network monitoring without relying on complex machine learning algorithms. Experimental results demonstrate that statistical hypothesis testing can effectively differentiate normal traffic from abnormal or attack traffic, thereby contributing to improved cybersecurity analysis.

Introduction

The exponential increase in online services, cloud computing, and digital transactions has made network security one of the most critical challenges of the modern era. Network traffic often exhibits variations in volume, frequency, and packet distribution due to user activity or external interference. However, some of these variations may signify malicious activities such as Denial of Service (DoS) attacks, port scanning, or data exfiltration.

Traditional cybersecurity systems rely heavily on machine learning or signature-based detection, which may not always detect new or evolving threats. Therefore, there is a growing need for statistical methods capable of identifying anomalies in network behavior based on measurable, quantifiable differences.

This paper focuses on the use of statistical hypothesis testing methods—specifically, the z-test, t-test, and chi-square test—to evaluate whether observed changes in network traffic are statistically significant. Unlike heuristic or AI-based methods, this approach is grounded in mathematical rigor and can be easily implemented using standard statistical tools. The main objective is to provide a reliable and explainable framework for anomaly detection that enhances the overall cybersecurity infrastructure.[1-3]

I. Network Traffic Analysis

Network traffic represents the continuous flow of data packets transmitted across network devices such as routers, switches, and servers. Each packet contains information including its size, transmission rate, source and destination addresses, and protocol type. Under normal conditions, network traffic exhibits consistent and predictable statistical characteristics – for example, stable average packet sizes and steady transmission frequencies.

However, during malicious activities such as denial-of-service (DoS) attacks, port scans, or intrusion attempts, these characteristics deviate from their expected ranges. Detecting these variations through statistical evaluation allows for the identification of anomalies at an early stage, thereby enhancing cybersecurity defense mechanisms.[1,3]

II. Statistical Hypothesis Testing

Statistical hypothesis testing provides a formal framework to determine whether an observed pattern in data significantly differs from an expected norm. In the context of network security, it is used to evaluate whether current traffic behavior deviates from the baseline of normal activity. The process involves two competing hypotheses:

- Null hypothesis (H_0): There is no statistically significant difference between normal and current network traffic patterns.;
- Alternative hypothesis (H_1): There exists a statistically significant difference, suggesting potential anomalous or malicious activity.

A test statistic is calculated based on the traffic data, and its corresponding p-value is compared to a chosen significance level (commonly $\alpha = 0.05$). If the p-value is smaller than α, the null hypothesis is rejected, indicating that the difference between traffic samples is not due to random variation but rather a meaningful change – possibly related to a cyberattack.[5]

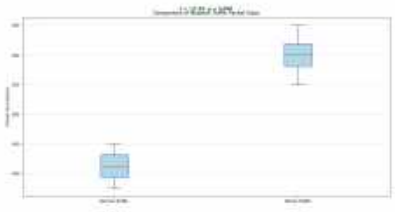


Figure 1 – Comparison on Network Traffic Packet Size

This example demonstrates how a t-test can identify statistically significant differences in average packet size between normal and attack traffic.

III. Tests Used

Different statistical tests are employed depending on the nature and size of the data:

Table 1 – Summarizes the statistical tests

Test Type	Sample Size Condition	Data Type	Purpose in Traffic Analysis
Z-test	Large sample ($n > 30$), known variance	Continuous	Detects deviation in mean packet size or transmission rate
T-test	Small sample ($n \geq 30$), unknown variance	Continuous	Compares means of normal and suspected traffic
Chi-square Test	Any sample size	Categorical	Compares frequency of packet types or IP categories

summarizes the statistical tests applied for evaluating variations in network traffic[3]

IV. Methodology

1. Data Collection. Network traffic data was collected from publicly available datasets, including NSL-KDD and UNSW-NB15, which provide both normal and attack traffic samples. These datasets contain a variety of attack types such as Denial of Service (DoS), Probe, and User to Root (U2R), allowing for

2. Feature Selection. The following parameters were selected as indicators of network behavior:

Table 2 – indicators of network

Feature	Description	Purpose
Packet Size	Number of bytes per packet	Detect anomalies in payload volume
Packet Rate	Number of packets per unit time	Identify traffic spikes
Protocol Type	TCP, UDP, ICMP, etc.	Observe changes in protocol distribution
Connection Duration	Time of each session	Capture unusual session behavior

3. Data Preprocessing

1. Normalization – Traffic parameters were scaled to ensure comparability and reduce bias from extreme values.

2. Grouping – Data was divided into two sets: Normal traffic Potential attack traffic

3. Statistical Testing. Different statistical tests were applied to identify anomalies in network traffic: Z-test: Compares average packet sizes between normal and abnormal traffic. Suitable for large sample sizes with known variance.T-test: Examines differences in packet transmission rates between groups with unknown variance or smaller sample sizes.Chi-square test: Evaluates deviations in categorical data, such as protocol type or source IP frequency.[3]

4. Decision Process. The results of the statistical tests are interpreted using ρ-values: ρ-value < 0.05: The null hypothesis is rejected; traffic variation is statistically significant, indicating potential cyber threats and ρ-value ≥ 0.05: The null hypothesis is not rejected; observed variations are likely due to random fluctuation.[4]

V. Conclusion

This study demonstrates that statistical hypothesis testing methods can be effectively applied to evaluate variations in network traffic and detect anomalies that may indicate cyberattacks. The proposed framework is mathematically rigorous, transparent, and computationally efficient. Unlike machine learning models, it does not require extensive training data or complex tuning, making it suitable for lightweight intrusion detection systems. Future research may extend this approach by integrating Bayesian inference or combining statistical testing with AI-based classifiers to achieve even higher detection accuracy.

1. Lippmann, R., Haines, J., Fried, D., Korba, J., & Das, K. The 1999 DARPA off-line intrusion detection evaluation / R. Lippmann, J. Haines, D. Fried, J. Korba, K. Das // Computer Networks. – 2000. – Vol. 34, № 4. – P. 579–595. – DOI: https://doi.org/10.1016/S1389-1286(00)00075-X.
2. Tavallaee, M., Bagheri, E., Lu, W., & Ghorbani, A. A. A detailed analysis of the KDD CUP 99 data set / M. Tavallaee, E. Bagheri, W. Lu, A. A. Ghorbani // Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications. – 2009. – P. 1–6. – IEEE.
3. Moustafa, N., & Slay, J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set) / N. Moustafa, J. Slay // 2015 Military Communications and Information Systems Conference (MilCIS). – 2015. – P. 1–6. – IEEE.
4. Chandola, V., Banerjee, A., & Kumar, V. Anomaly detection: A survey / V. Chandola, A. Banerjee, V. Kumar // ACM Computing Surveys. – 2009. – Vol. 41, № 3. – P. 1–58. – DOI: https://doi.org/10.1145/1541880.1541882.
5. Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. Network anomaly detection: Methods, systems and tools / M. H. Bhuyan, D. K. Bhattacharyya, J. K. Kalita // IEEE Communications Surveys & Tutorials. – 2014. – Vol. 16, № 1. – P. 303–336. – DOI: https://doi.org/10.1109/SURV.2013.052213.00146.

AN IMPROVED Q-LEARNING ALGORITHM WITH OPTIMIZED INITIALIZATION AND ANNEALED BOLTZMANN EXPLORATION

Tang Yi, Yu.O.German
Department of Information Technologies in Automated Systems,
Belarussian State University of Informatics and Radioelectronics
Minsk, Republic of Belarus
E-mail: tangyijcb@163.com, jgerman@bsuir.by

This paper proposes a hybrid enhancement method for Q-learning that combines direction-sensitive Q-table initialization with annealing-based Boltzmann exploration. Initialization leverages geometric priors to bias actions toward the target without leaking obstacle information; the annealing-based Boltzmann method achieves a smooth transition from extensive exploration to exploitation. By leveraging the symmetry of isometric states and an adaptive exploration strategy, the improved Q-learning algorithm achieves faster convergence in discrete action environments.

Introduction

In intelligent transportation systems (ITS), navigation and path planning are foundational capabilities that underpin applications such as autonomous driving, unmanned delivery, and warehouse logistics. Typical scenarios include efficient traversal within urban road networks, congestion and obstacle avoidance in warehouse grid environments, and policy evaluation in simulated traffic networks. Given the discrete layout of environments, constrained traversability, and complex dynamics, achieving efficient and robust path planning in unknown or partially known settings carries substantial research and engineering significance.

From an application perspective, an effective navigation strategy should satisfy three criteria: reachability, efficiency, and robustness. Methodologically, traditional rule-based or heuristic planners (e.g., fixed-cost search and hand-crafted policies) often require extensive parameter tuning and complete map information in large-scale, dynamically changing scenarios; they generalize poorly to novel layouts or constraints and incur high maintenance costs[1].

Against this backdrop, reinforcement learning (RL), with its capacity to learn adaptively through interaction with the environment, has emerged as a powerful approach to path planning in grid-based settings. Among RL methods, Q-learning method has been widely adopted for grid navigation with discrete actions due to its simplicity, computational lightness, and strong interpretability.

I. Weaknesses of existing methods

Despite substantial progress, Q-learning and its variants continue to face efficiency and robustness bottlenecks in large-scale, obstacle-dense, and sparse-reward grid environments. First, the lack of reliable navigational priors at initialization leads to directionless early exploration when Q-tables are uniform or random. Although distance- or heuristic-based initializations can accelerate learning to some extent, they tend to over-bias the agent toward blocked directions under high obstacle density, and their benefits are highly map-dependent. Second, exploration mechanisms are largely heuristic and insensitive to learning progress. Specifically, fixed ϵ-greedy induces high variance and repeated probing, while Boltzmann/entropy-regularized policies or adaptive ϵ decay can improve stability but introduce additional hyperparameters and tuning burdens, and still exhibit oscillations, cul-de-sac trapping, and premature greediness in heavily cluttered settings[2]. In terms of scalability, approaches based on local-search decomposition and hybrids with evolutionary/swarm intelligence enhance global exploration but increase computational complexity and parameter sensitivity, limiting real-time deployment.

Overall, existing improvements partially alleviate exploration and reward sparsity issues, yet remain highly sensitive to priors and hyperparameters. As obstacle density and map scale grow, robustness and generalization remain inadequate.

II. Problem statement

We model path planning in a 2D grid environment as a discrete Markov Decision Process (MDP) $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$. On a grid of size $H \times W$, the state $s = (i, j) \in \{0, \ldots, H-1\} \times \{0, \ldots, W-1\}$ denotes the current position of agent, with a designated start s_{start} and goal s_{goal}. The action set $\mathcal{A} = \{\text{up}, \text{down}, \text{left}, \text{right}\}$ corresponds to four-connected movements.

The reward function is designed to jointly capture task completion, collision penalties, and time costs[3]. with the objective of maximizing the expected discounted return which can be formed below in Equation (1):

$$\max_{\pi} \ \mathbb{E}_{\pi}\left[\sum_{t=0}^{T_{\max}} \gamma^t \, r_t \;\middle|\; s_0 = s_{\text{start}}\right] \tag{1}$$

where $\gamma \in [0, 1)$ is the discount factor and r_t is the instantaneous reward induced by the current transition. For small to medium discrete state spaces,

we adopt tabular Q-learning as a baseline solver to estimate the action-value function in a model-free manner and iteratively improve the policy. The update rule is shown in Equation (2):

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha\Big(r_t + \gamma \max_{a' \in \mathcal{A}} Q(s_{t+1}, a') - Q(s_t, a_t)\Big) \tag{2}$$

where $\alpha \in (0, 1]$ is the learning rate. Exploration–exploitation trade-offs are handled via ϵ-greedy.

III. Method description

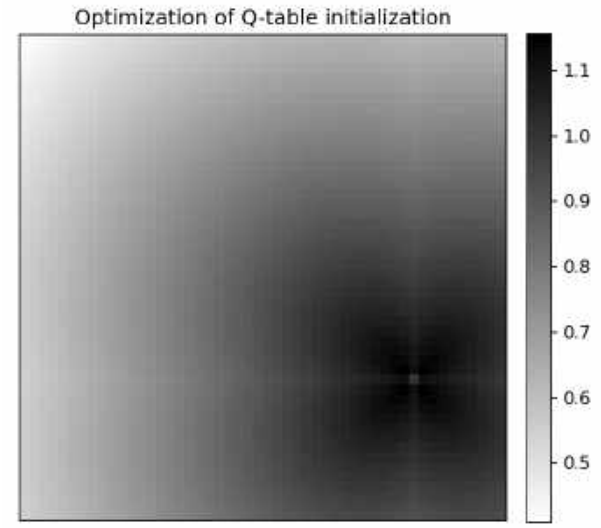


Figure 1 – Optimized initialization of Q table

Optimized Q Table Initialization: To mitigate the drawback of standard initialization that ignores navigational priors, we provide a warm start to reduce early blind exploration. Unlike zero or random initialization, we compute $Q_0(i, j, a)$ using goal direction and distance, where (i, j) denotes the state position, a denotes the action index, and Q_0 is the initial Q-value. The initialization is defined as in Equation(3):

$$Q_0(i, j, a) = \beta_0 + \beta_1 \left(1 - \hat{d}(i, j; g_i, g_j)\right) + \beta_2 \kappa(i, j, a) \tag{3}$$

where $\beta_0 = 0$ is a constant bias term, $\beta_1 = 0.8$ is the distance weight, and $\beta_2 = 0.5$ is the direction weight; $\hat{d}(i, j; g_i, g_j)$ is the normalized Manhattan distance (with (g_i, g_j) denoting the goal position), computed using $\hat{d}(i, j; g_i, g_j) = \frac{|i-g_i|+|j-g_j|}{(H-1)+(W-1)+\epsilon_0}$, ($\epsilon_0 = 1\text{e}{-8}$ is a small constant to avoid division by zero). $\kappa(i, j, a)$ is the action alignment score, computed via the dot product between the unit goal vector field $\vec{u}_g(i, j)$ and the unit action vector $\vec{u}_a$. The calculation is shown below in Equation (4):

$$\kappa(i, j, a) = \max\left(0, \vec{u}_g(i, j) \cdot \vec{u}_a\right) \tag{4}$$

This method biases actions that are closer to the goal and aligned in direction, significantly accelerating convergence. Figure 1 shows the visualization of optimized initialization of the table.

Annealed Boltzmann Exploration To address the limitations of standard ϵ-greedy with high noise and insufficient late-stage exploration. we adopt an annealed Boltzmann scheme whose temperature decays exponentially with episode e[4]. The calculation for temperature T_e is in Equation(5):

$$T_e = \max(T_{\min}, T_0 \cdot \rho^e) \tag{5}$$

where $T_0 = 1.5$ is the initial temperature, $T_{\min} = 0.02$ is the lower bound, and $\rho = 0.998$ is the decay rate. The action probability $P(a \mid s)$ is computed as:

$$P(a \mid s) = \frac{\exp\left(Q'(s, a)/T\right)}{\sum_{a' \in A} \exp\left(Q'(s, a')/T\right)} \tag{6}$$

To improve efficiency, we add a UCB bonus(Upper Confidence Bound bonus): $Q'(s, a) = Q(s, a) + c\sqrt{\ln(t + 1)/(N(s, a) + \epsilon)}$, where $Q'(s, a)$ is the adjusted Q-value, $c = 0.1$ is the exploration constant, t is the total time step count, and $N(s, a)$ is the state–action visit count. This mechanism encourages actions with higher uncertainty and promotes detour discovery in cluttered environments, contrasting with the inefficient exploration reported for standard Q-learning in computationally complex dynamic warehouses. Visit counts are tracked via state and action arrays, and the global step t increments per action.

IV. conclusion

We propose an improved Q-learning method for grid-based navigation, centered on the synergy between direction-sensitive Q-table initialization and annealed Boltzmann exploration. The former provides a mild yet explicit directional prior, while the latter implements a smooth, adaptive softening mechanism to transition from exploration to exploitation. Without revealing obstacle information and while retaining a model-free paradigm, the method is expected to improve sample efficiency and convergence stability.

1. Zhang, Y., Zhao, W., Wang, J., & Yuan, Y. (2024). Recent progress, challenges and future prospects of applied deep reinforcement learning: A practical perspective in path planning. Neurocomputing, 608, 128423.
2. Pei, M., An, H., Liu, B., & Wang, C. (2021). An improved dyna-q algorithm for mobile robot path planning in unknown dynamic environment. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52(7), 4415-4425.
3. Ben-Akka, M., Tanougast, C., & Diou, C. (2025). Novel design of reward and epsilon-greedy decay strategy tailored for Q-learning in optimizing local mobile robot path planning. Knowledge-Based Systems, 113836.
4. Zhou, Q., Lian, Y., Wu, J., Zhu, M., Wang, H., & Cao, J. (2024). An optimized Q-Learning algorithm for mobile robot local path planning. Knowledge-Based Systems, 286, 111400.

Секция «Технологии цифровой экономики»

Section «Digital Economy Technologies»

ПРИМЕНЕНИЕ БЛОКЧЕЙН-ТЕХНОЛОГИЙ ДЛЯ ОБЕСПЕЧЕНИЯ ПРОЗРАЧНОСТИ ЦЕПОЧЕК ПОСТАВОК

Бекиева М. Б., Ораздурдыева Г. О., Бекиев А. Р., Дурдыев С. М.
Кафедра прикладной математики и информатики,
Кафедра Компьютерных наук и информационных систем,
Инженерно-технологический университет Туркменистана имени Огуз хана
Инженерно-экономический факультет,
Белорусский государственный университет информатики и радиоэлектороники
Экономический факультет, Российский университет дружбы народов
Ашхабад, Туркменистан; Минск, Республика Беларусь; Москва, Российская Федерация
E-mail: successbmb@gmail.com, maral.bekiyeva@etut.edu.tm

В статье рассматривается использование блокчейн-технологий для повышения прозрачности, доверия и эффективности в управлении цепочками поставок. Представлены основные принципы построения распределённых реестров и их преимущества по сравнению с традиционными централизованными системами. Приведена формальная модель оценки прозрачности цепочки поставок и сравнительная таблица ключевых параметров двух подходов. Показано, что применение блокчейна позволяет повысить достоверность данных, сократить количество посредников и ускорить процесс аудита, обеспечивая при этом высокий уровень безопасности и верификации информации. Сделан вывод о перспективности интеграции блокчейна с технологиями Интернета вещей и искусственного интеллекта в целях дальнейшей цифровизации и автоматизации логистических процессов.

Введение

Цепочки поставок представляют собой многоуровневые системы взаимодействия производителей, поставщиков, логистических компаний и розничных продавцов. Эффективность их функционирования во многом зависит от прозрачности процессов и достоверности информации. Однако традиционные централизованные системы управления поставками подвержены ряду рисков – искажениям данных, подделкам документов, низкой скорости аудита и необходимости доверия посредникам [1].

Блокчейн-технология, основанная на принципах децентрализации, неизменяемости и прозрачности, предоставляет новые возможности для решения этих проблем. Она позволяет фиксировать каждое событие в распределённом реестре, что обеспечивает достоверность данных и снижает вероятность мошенничества. Дополнительно блокчейн повышает эффективность процессов верификации и позволяет интегрировать различные звенья логистической сети в единое цифровое пространство. Таким образом, технология становится фундаментом для цифровой трансформации цепочек поставок и внедрения концепции «умной логистики».

I. Проблематика традиционных цепочек поставок

Современные цепочки поставок страдают от недостаточной прозрачности. Информация о происхождении и движении продукции часто скрыта от участников, а верификация данных требует значительных затрат времени. Централизованные системы хранения данных уязвимы для несанкционированного доступа и подделки документов. Это приводит к потере доверия между участниками и повышению транзакционных издержек [2].

Применение блокчейна позволяет устранить необходимость в посредниках, создавая единый источник достоверной информации, доступный всем участникам сети. Каждое изменение состояния товара фиксируется в распределённом реестре и подтверждается консенсусом, что обеспечивает неизменяемость записей. Подобная система минимизирует человеческий фактор, улучшает контроль качества и способствует устойчивому развитию логистики.

II. Архитектура блокчейн-системы для логистики

Блокчейн-система для цепочек поставок может включать следующие элементы:

- Распределённый реестр транзакций – хранит информацию обо всех этапах перемещения товара;
- Смарт-контракты – автоматически проверяют условия поставки и исполняют соглашения между сторонами;
- Интерфейс участников – предоставляет безопасный доступ к данным и инструментам мониторинга.

Типовая последовательность взаимодействий имеет вид:

Производитель → Блокчейн-узел → Логистический оператор → Блокчейн-узел → Розничный продавец → Покупатель.

Каждое событие в процессе поставки подтверждается несколькими узлами, что исключает возможность изменения информации без общего согласия [3].

III. Формальная модель прозрачности цепочки поставок

Для количественной оценки прозрачности системы введём показатель P:

Для количественной оценки прозрачности системы введём показатель:

$$P = \frac{\sum_{i=1}^{n} T_i \cdot W_i}{N}$$

где T_i – количество проверяемых транзакций на i-м этапе, W_i – весовой коэффициент значимости этапа, N – общее количество звеньев цепочки.

Рост показателя P свидетельствует о повышении прозрачности системы. Применение блокчейна позволяет увеличить количество проверяемых транзакций T_i, а также снизить вероятность искажения данных.

IV. Сравнение традиционных и блокчейн подходов

Для демонстрации преимуществ децентрализованного подхода приведём сравнительную таблицу.

Таблица 1 – Сравнительный анализ характеристик традиционных и блокчейн-систем поставок

Критерий	Традиционная система	Блокчейн-система
Хранение данных	Централизованное	Распределённое между участниками
Проверка достоверности	Ручная, требует посредников	Автоматическая с помощью консенсуса
Изменяемость данных	Возможна	Практически исключена
Скорость аудита	Низкая	Высокая
Прозрачность	Ограниченная	Контролируемая и полная
Требование доверия к посреднику	Высокое	Отсутствует

Из таблицы видно, что блокчейн-модель обеспечивает более высокий уровень прозрачности и снижает операционные риски, связанные с человеческим фактором и централизацией данных [4].

Исследования показали, что внедрение блокчейн-технологий повышает показатель прозрачности P на 35–50% [5]. Это особенно актуально для отраслей с высокими требованиями к сертификации и отслеживанию происхождения товаров – например, пищевой промышленности и фармацевтики.

Блокчейн способствует сокращению времени на проведение аудита, минимизации споров и повышению доверия между сторонами. Однако внедрение технологии требует стандартизации протоколов обмена данными и адаптации существующих информационных систем [6]. Кроме того, успех интеграции зависит от правовых норм и уровня цифровой зрелости компаний, участвующих в цепочке поставок.

V. Заключение

Блокчейн-технологии обеспечивают новый уровень прозрачности и надёжности в управлении цепочками поставок. Их применение позволяет исключить посредников, предотвратить подделку информации и ускорить контроль над движением товаров. Несмотря на существующие проблемы масштабируемости и затрат на внедрение, децентрализованные системы становятся ключевым элементом цифровой экономики.

В перспективе интеграция блокчейна с интернетом вещей и искусственным интеллектом создаст полностью автоматизированные и самопроверяемые цепочки поставок, что станет важным шагом в развитии индустрии 4.0.

VI. Список литературы

1. Saberi, S., Kouhizadeh, M., Sarkis, J. Blockchain technology and its relationships to sustainable supply chain management / S. Saberi, M. Kouhizadeh, J. Sarkis // International Journal of Production Research. – 2020. – Vol. 58, № 7. – P. 2063–2081.
2. Wang, Y., Singgih, M., Wang, J., Rit, M. Making sense of blockchain technology: How will it transform supply chains? / Y. Wang, M. Singgih, J. Wang, M. Rit // International Journal of Production Economics. – 2021. – Vol. 231. – P. 107–122.
3. Francisco, K., Swanson, D. The Supply Chain Has No Clothes: Technology Adoption of Blockchain for Supply Chain Transparency / K. Francisco, D. Swanson // Logistics. – 2020. – Vol. 4, № 1. – P. 7–17.
4. Kouhizadeh, M., Saberi, S., Sarkis, J. Blockchain technology and the sustainable supply chain: Theoretically exploring adoption barriers / M. Kouhizadeh, S. Saberi, J. Sarkis // International Journal of Production Economics. – 2021. – Vol. 231. – P. 108–119.
5. Bai, C., Dallasega, P., Orzes, G., Sarkis, J. Industry 4.0 technologies assessment: A sustainability perspective / C. Bai, P. Dallasega, G. Orzes, J. Sarkis // International Journal of Production Economics. – 2022. – Vol. 243. – P. 111–128.
6. Bekiyeva, M. B., Orazdurdyyeva, G. O. Application of Semantic Analysis and GMM Models for Anomaly Detection in Network Traffic / M. B. Bekiyeva, G. O. Orazdurdyyeva // Open Semantic Technologies for Intelligent Systems Research Papers Collection. – Minsk, 2025. – P. 45–51.

ИСПОЛЬЗОВАНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В DIGITAL-ПРОДВИЖЕНИИ

Верняховская В. В., Усова В. А.
Кафедра экономики, кафедра экономической информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: verniahovskaya@bsuir.by, uvikamail@gmail.com

В статье анализируется трансформация digital-продвижения под влиянием искусственного интеллекта. Рассматриваются ключевые направления применения ИИ: глубокая персонализация коммуникаций, предиктивная аналитика, автоматизация создания контента и интеллектуальный таргетинг. Особое внимание уделяется преимуществам внедрения – рост конверсии, снижение затрат и повышение скорости принятия решений. Также затрагиваются проблемы внедрения, включая этические аспекты и зависимость от качества данных. Определены перспективы интеграции ИИ в маркетинговые стратегии.

Введение

Современная цифровая экономика характеризуется стремительным развитием технологий, изменением потребительского поведения и усилением конкуренции в онлайн-пространстве. В этих условиях искусственный интеллект становится одним из наиболее перспективных инструментов digital-маркетинга, радикально меняя подходы к анализу данных, таргетингу, созданию контента и взаимодействию с потребителем.

Если раньше интернет-маркетинг основывался преимущественно на статистических методах и ручном управлении рекламными кампаниями, то сегодня ключевым конкурентным преимуществом становится способность компании использовать интеллектуальные системы для автоматизации и персонализации продвижения. Это позволяет не только предсказывать тенденции и поведение клиентов с высокой точностью, но и в реальном времени адаптировать маркетинговые стратегии под их меняющиеся потребности. Таким образом, ИИ трансформирует маркетинг из просто инструмента продвижения в стратегическую систему управления клиентским опытом.

I. Роль искусственного интеллекта в цифровом маркетинге

Искусственный интеллект в digital-маркетинге представляет собой совокупность технологий, основанных на машинном обучении, нейронных сетях и анализе больших данных (Big Data), направленных на оптимизацию маркетинговых процессов.

Главная задача ИИ в этой сфере – автоматизация рутинных операций и повышение точности маркетинговых решений. Алгоритмы способны самостоятельно анализировать огромные объёмы информации о поведении пользователей, определять закономерности, прогнозировать спрос и формировать персонализированные предложения [1].

Таким образом, ИИ становится не просто вспомогательным инструментом, а стратегическим элементом системы маркетинговых коммуникаций, обеспечивающим повышение эффективности digital-продвижения.

II. Основные направления применения ИИ в digital-продвижении

Одним из ключевых преимуществ ИИ является возможность глубокой персонализации пользовательского опыта. На основе анализа поведения клиента (просмотры, клики, покупки, время на сайте) алгоритмы создают индивидуальные рекомендации и формируют динамические рекламные сообщения.

Предиктивная аналитика позволяет прогнозировать вероятность конверсии, оттока клиентов и интерес к конкретным товарам. Например, платформы Google Ads и Meta Ads (Facebook, Instagram) активно используют машинное обучение для автоматической оптимизации ставок и выбора целевой аудитории, что существенно повышает эффективность рекламных кампаний.

ИИ-системы всё чаще применяются для создания и редактирования текстового, визуального и видео-контента. Инструменты вроде ChatGPT, Jasper, Copy.ai или Midjourney позволяют генерировать рекламные тексты, заголовки, сценарии и даже изображения, адаптированные под целевую аудиторию.

Благодаря анализу откликов аудитории ИИ способен определять, какие темы и форматы контента вызывают наибольший интерес, и предлагать варианты оптимизации. Это делает коммуникации бренда более точными и релевантными, снижая издержки на контент-продакшн.

Традиционные методы сегментации (по возрасту, полу, географии) постепенно уступают место интеллектуальному таргетингу, основанному на поведении и интересах пользователей. ИИ анализирует не только демографические характеристики, но и цифровые следы: поисковые запросы, время активности, покупки, реакции на контент.

На основе этих данных создаются динамические сегменты, которые обновляются в реальном времени, что позволяет показывать рекламу тем, кто с наибольшей вероятностью совершит целевое действие.

Использование чат-ботов на базе искусственного интеллекта стало одним из самых популярных направлений digital-продвижения. Они обеспечивают круглосуточную коммуникацию с пользователями, отвечают на вопросы, подбирают товары, помогают оформить заказ.

Интеллектуальные ассистенты (например, чат-боты на базе GPT-технологий) способны вести диалог в естественной форме, запоминать предпочтения клиента и повышать уровень удовлетворённости обслуживанием. Для бизнеса это не только улучшение клиентского опыта, но и значительное сокращение затрат на персонал.

ИИ активно применяется в системах автоматического медиапланирования. Алгоритмы анализируют эффективность разных каналов продвижения, рассчитывают оптимальный бюджет, перераспределяют ресурсы и формируют отчёты в режиме реального времени. Например, платформа Google Marketing Platform использует машинное обучение для прогнозирования рентабельности инвестиций (ROI) и повышения конверсии за счёт автоматической корректировки стратегии размещения рекламы [2].

III. Влияние искусственного интеллекта на Digital-маркетинг: преимущества и проблемы внедрения

Применение ИИ в digital-продвижении обеспечивает следующие преимущества:

- точность таргетинга и снижение издержек;
- рост конверсии и лояльности клиентов;
- скорость анализа данных;
- автоматизация рутинных задач;
- прогнозирование трендов.

Таким образом, искусственный интеллект превращает digital-маркетинг из набора отдельных инструментов в интегрированную, самонастраивающуюся систему коммуникаций [3].

Это проявляется в создании единого контура управления, где CRM, аналитические платформы и рекламные каналы обмениваются данными без участия человека, постоянно повышая релевантность коммуникаций.

Несмотря на значительные преимущества, внедрение ИИ сопровождается рядом проблем:

- этические и правовые аспекты;
- отсутствие прозрачности алгоритмов;
- зависимость от качества данных;
- необходимость квалифицированных кадров.

Эти проблемы замедляют темпы внедрения ИИ в маркетинговую практику, особенно в странах с ограниченными цифровыми ресурсами и недостаточным уровнем подготовки специалистов [4].

Кроме того, существует фундаментальное ограничение: ИИ эффективен в оптимизации известных процессов, но его способность к генерации по-настоящему креативных и инновационных маркетинговых стратегий без участия человека остается предметом споров.

Таким образом, развитие искусственного интеллекта в digital-продвижении представляет собой сложный и противоречивый процесс.

IV. Перспективы развития

Будущее digital-маркетинга неразрывно связано с дальнейшей интеграцией искусственного интеллекта во все этапы коммуникации с клиентом.

Наиболее перспективными направлениями можно считать:

- развитие гиперперсонализации;
- использование нейромаркетинга;
- интеграцию ИИ с технологиями AR/VR;
- формирование адаптивных экосистем брендов.

В перспективе искусственный интеллект станет основой не только для анализа и автоматизации, но и для стратегического управления маркетингом, превращая цифровое продвижение в самообучающуюся систему [5].

Заключение

Использование искусственного интеллекта в digital-продвижении – это не просто технологическая тенденция, а логический этап эволюции маркетинга в эпоху данных. ИИ обеспечивает бизнесу возможность точнее понимать аудиторию, быстрее реагировать на изменения рынка и выстраивать долгосрочные отношения с клиентами.

Компании, внедряющие интеллектуальные решения уже сегодня, получают стратегическое преимущество, поскольку именно искусственный интеллект становится главным драйвером эффективности, креативности и инновационности цифровых коммуникаций.

V. Список литературы

1. Иванов, А. В. Цифровая трансформация маркетинга / А. В. Иванов // М.: Инфра-М – 2021. – 256 с.
2. Инструменты и технологии digital-маркетинга [Электронный ресурс] / APNI.ru. – Режим доступа: `https://apni.ru/article/7286-instrumenti-i-tekhnologii-digital-marketinga`. – Дата доступа: 18.10.2025.
3. Цифровой маркетинг: обзор каналов и инструментов [Электронный ресурс] /Cyberleninka.ru. – Режим доступа: `https://cyberleninka.ru/article/n/tsifrovoy-marketing-obzor-kanalov-i-instrumentov`. – Дата доступа: 18.10.2025.
4. Котлер, Ф. Маркетинг менеджмент / Ф. Котлер, К. Л. Келлер – СПб.: Питер, 2018. – 800 с
5. Ламбен, Ж. Ж. Стратегический маркетинг. Европейская перспектива / Ж. Ж. Ламбен // М.: ИД «Вильямс» – 2020. – 736 с.

АНАЛИЗ ПОТРЕБИТЕЛЬСКОГО ПОВЕДЕНИЯ НА ОСНОВЕ МЕТОДОВ ЦИФРОВОГО МАРКЕТИНГА И МАШИННОГО ОБУЧЕНИЯ

Дайнович О. Л., Ситников А. В., Мигалевич С. А.
Центр информатизации и инновационных разработок,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: {o.dajnovich, a.sitnikov, migalevich}@bsuir.by

В статье рассматриваются современные методы анализа потребительского поведения в цифровой среде с использованием инструментов машинного обучения и цифрового маркетинга. Приведена классификация аналитических подходов, а также методы интеллектуальной обработки данных. Особое внимание уделено применению предиктивной и поведенческой аналитики для прогнозирования действий клиентов и персонализации маркетинговых стратегий. Отмечается актуальность интеграции технологий машинного обучения в маркетинговые процессы для повышения эффективности взаимодействия с аудиторией и оптимизации бизнес-решений в условиях цифровой экономики.

Введение

Современная цифровая экономика характеризуется ростом объёмов информации и развитием технологий, позволяющих собирать и анализировать данные о поведении пользователей в сети. Потребитель взаимодействует с брендами через социальные сети, поисковые системы, электронную почту, мобильные приложения и онлайн-платформы. В результате компании получают доступ к большим массивам данных, которые могут быть использованы для глубокого понимания предпочтений и мотивации клиентов.

Использование инструментов машинного обучения для анализа этих данных позволяет выявлять скрытые закономерности, прогнозировать поведение клиентов и формировать индивидуальные стратегии продвижения товаров и услуг [1].

I. Аналитические методы цифрового маркетинга

Аналитические методы направлены на извлечение и интерпретацию данных о поведении пользователей в цифровой среде.

К их числу относятся:

- Веб-аналитика – измерение активности пользователей на сайтах и платформах. Основные показатели (Key Performance Indicators): трафик, источники переходов, время на странице, глубина просмотра, путь пользователя (User Journey). Веб-анаитика применяется для выявления наиболее интересных страниц, точек оттока, анализа конверсий.
- UBA (User Behaviour Analytics) – поведенческая аналитика действий пользователя: клики, просмотры, добавления в корзину, реакции на уведомления. UBA позволяет построить карты тепла (heatmaps), воронки поведения и сегменты пользователей по активности.
- Когортный анализ (Cohort Analysis) – метод анализа метрик продукта на основе поведения групп людей (когорт), объединенных по какому-либо признаку во времени. Обычно в одну когорту объединяют людей, которые начали пользоваться продуктом в один и тот же месяц (или неделю).
- RFM-анализ. Данный позволяет разделить клиентов на сегменты, основываясь на данных об их покупках. Он учитывает три признака: recency (давность покупки), frequency (частота покупок) и monetary (сумма трат клиента).
- Анализ воронки продаж (Conversion Funnel Analysis) – анализ маркетингового инструмента для визуализации шагов клиента от предложения товара или услуги до покупки (осведомленность, интерес, действие, покупка) [2].

II. Методы машинного обучения и интеллектуальной аналитики

В контексте цифрового маркетинга методы машинного обучения позволяют анализировать большие массивы данных о поведении пользователей, прогнозировать их действия, выявлять скрытые сегменты клиентов и персонализировать взаимодействие с аудиторией.

Основная идея заключается в том, чтобы использовать поведенческие и транзакционные данные (клики, просмотры, покупки, реакции на рекламу) для построения моделей, которые предсказывают будущее поведение клиента или автоматически группируют пользователей по схожим признакам. К методам, используемым для построения этих моделей, относятся:

- Кластеризация. Цель – выявить группы потребителей с похожими характеристиками без заранее заданных меток. Основными алгоритмами для кластеризации являются K-Means (делит пользователей на заданное число кластеров по близости призна-

ков), hierarchical Clustering (строит древовидную структуру связей между пользователями)и DBSCAN (выделяет плотные группы и отбрасывает шум (удобен для больших и неструктурированных данных)).

- Классификация. Цель – определить, к какому классу относится пользователь (например, «совершит покупку» или «уйдёт»). Основные алгоритмы классификации: Decision Tree (дерево решений), Random Forest (случайный лес), Gradient Boosting (XGBoost, LightGBM, CatBoost), Support Vector Machines (SVM).
- Регрессионный анализ. Цель – предсказать числовое значение, например сумму следующей покупки, длительность жизненного цикла клиента или вероятность клика по рекламе. Для регрессионного анализа используются линейные, нелинейные либо регуляризованные модели.
- Анализ тональности и текстов (Sentiment Analysis, NLP) – анализ, применяемый к отзывам, комментариям, постам. Позволяет определить эмоциональную окраску (положительная, нейтральная, отрицательная). В качестве инструментов используются модели spaCy, NLTK, BERT, ChatGPT API [3].
- Рекомендательные системы – алгоритмы коллаборативной фильтрации, контентной фильтрации, применяемые для персональных рекомендаций товаров, статей, услуг.

Применение методов машинного обучения и интеллектуальной аналитики в цифровом маркетинге открывает новые возможности для глубокого понимания потребительского поведения. Использование алгоритмов классификации, кластеризации и прогнозирования позволяет не только описывать текущее состояние аудитории, но и предсказывать её будущие действия, создавая условия для персонализированных стратегий взаимодействия и повышения эффективности маркетинговых решений.

III. Методы предиктивной и поведенческой аналитики

Современные компании, работающие в цифровой среде, стремятся не только понимать текущее поведение потребителей, но и предсказывать их будущие действия. Для этого используются методы предиктивной (прогностической) и поведенческой аналитики, основанные на сборе, моделировании и интерпретации данных о взаимодействиях пользователей с брендом.

Эти подходы позволяют организациям переходить от реактивного управления маркетингом к проактивному – предвосхищать поведение клиентов и автоматически адаптировать маркетинговые стратегии под их потребности.

Предиктивная аналитика (Predictive Analytics) – это область аналитики, использующая статистические модели и методы машинного обучения для прогнозирования вероятности будущих событий на основе исторических данных. В цифровом маркетинге она применяется для прогнозирования поведения покупателей (покупка, отток, возврат), оценки вероятности конверсии, прогнозирования спроса и динамики продаж, определения оптимального времени для контакта с клиентом, а также автоматизации персонализированных предложений. К методам предикивной аналитики относятся регрессионные модели, прогнозирование спроса с помощью временных рядов, модели оттока клиентов, а также прогноз жизненной ценности клиента [4].

Поведенческая аналитика (Behavioral Analytics) – это анализ действий пользователей с целью понимания их мотивов, привычек и закономерностей взаимодействия с цифровыми платформами. В отличие от предиктивной аналитики, она фокусируется на описании и объяснении текущего поведения, а не на прогнозе. Для проведения поведенческой аналитики используются: анализ пользовательских путей, когортный анализ, RFM-анализ, анализ микровзаимодействий и событий [5].

Заключение

Методы предиктивной и поведенческой аналитики являются неотъемлемой частью современной системы цифрового маркетинга. Благодаря интеграции с инструментами машинного обучения и большими данными, эти методы формируют основу интеллектуальных маркетинговых систем, способных принимать решения в реальном времени и обеспечивать максимальную эффективность взаимодействия с потребителем.

IV. Список литературы

1. Marketing Analytics: A Machine Learning Approach / A. Mansurali, P. Mary Jeyanthi – 2023.
2. Time analysis of online consumer behavior by decision trees, GUHA association rules, and formal concept analysis, Journal of Marketing Analytics. – 09.01.2024.
3. Machine Learning for Marketing / Hiroshi Mamitsuka – 2019.
4. Predictive Analytics and Generative AI for Data-Driven Marketing Strategies / Hemachandran K., Debdutta Choudhury, Raul Villamarin Rodriguez, Jorge A. Wise, Revathi T. – 2025.
5. Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics and Big Data / Omer Artun, Dominique Levin – 2015.

ПРОГРАММНОЕ СРЕДСТВО АНАЛИЗА НОВОСТНОГО СЕНТИМЕНТА НА КРИПТОВАЛЮТНОМ РЫНКЕ

Галяк И. П.
Кафедра проектирования информационно-компьютерных систем,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: galiak.ivan@gmail.com

Настоящая работа посвящена разработке программного средства для автоматизированного анализа новостного фона на криптовалютном рынке. В условиях высокой волатильности и информационной перегрузки предложенное решение использует открытые новостные API для агрегации данных и большую языковую модель (LLM) для оценки потенциального влияния новостей. Реализован алгоритм приоритизации информации на основе достоверности и важности источника. Результаты анализа, включая ключевые метрики и краткие сводки, представляются пользователю через веб-интерфейс. Это позволяет сократить время на оценку новостного фона и предоставить инструмент для поддержки принятия инвестиционных решений.

Введение

Криптовалютный рынок характеризуется экстремальной волатильностью и высокой степенью зависимости от новостного фона. Ввиду этого, для инвесторов одной из ключевых задач становится оперативная обработка и оценка обширного потока неструктурированной текстовой информации. Ручной мониторинг многочисленных новостных источников является трудоемким процессом, подверженным значительным когнитивным искажениям.

Целью данной работы является реализация методов обработки новостной информации, позволяющих эффективно снизить информационную перегрузку за счет применения технологий обработки естественного языка (NLP) посредством больших языковых моделей (LLM) [1, 2].

I. Проблематика

Криптовалютный рынок характеризуется экстремальной волатильностью, круглосуточным функционированием и критической зависимостью от непрерывного, зачастую гетерогенного новостного потока. Это создает информационную перегрузку, в которой рыночные сигналы (регуляторные изменения или крупные инциденты безопасности) с трудом различимы среди доминирующего объема «шума», состоящего из спекуляций, слухов и малозначимых событий. В таких условиях ручная обработка и контекстуальная оценка потенциального влияния новостей на активы становится не только чрезвычайно трудоемкой и требующей глубоких экспертных знаний, но и крайне необъективной. Человеческое восприятие подвержено значительным когнитивным искажениям и уязвимо к целенаправленным информационным манипуляциям типа FUD (Fear, Uncertainty, Doubt) и FOMO (Fear of Missing Out). В результате участники рынка принимают неоптимальные, эмоционально окрашенные решения, основанные не на объективном анализе, а на реактивном поведении, что ведет к снижению эффективности и увеличению рисков.

II. Реализация алгоритмов оценки

Программное средство анализа новостного сентимента «Axon» построено на модульной архитектуре и реализует многоступенчатый алгоритм обработки и оценки новостных данных, предоставляя пользователю инструмент анализа и поддержки принятия решений. Ядром системы является алгоритм семантической оценки, реализованный посредством взаимодействия с API большой языковой модели (в прототипе используется Groq API, модель llama-3.1-8b-instant). Для каждой входящей новости модель, предварительно настроенная системным промптом как финансовый аналитик и обязанная возвращать структурированный JSON-ответ, генерирует четыре ключевые метрики. К ним относятся sentiment_score (числовая оценка тональности в диапазоне [-1.0, 1.0]), summary (краткая аналитическая сводка из 2-3 предложений), mentioned_coins (список тикеров криптовалютных активов для тематической фильтрации) и impact_score (целочисленная оценка потенциального влияния новости на рынок от 1 до 5).

Семантический анализ, выполняемый LLM, дополняется алгоритмом приоритезации, основная задача которого – ранжировать новости не только по содержанию, но и по потенциальной силе влияния источника на рынок, имитируя логику опытного аналитика. Система проверяет полное текстовое содержание каждой новости (заголовок, основной текст и, что критично, имя источника) на наличие заранее определенных ключевых слов-маркеров, классифицируя новость по одному из четырех уровней. Приоритет 1 (Критический, «Market-Moving») присваивается при обнаружении упоминаний центральных банков стран G7 и ключевых финансовых регуляторов, таких как «Federal Reserve», «SEC», «European Central Bank» или «IMF», спо-

собных вызывать максимальные изменения. Приоритет 2 (Высокий, «Trend-Setting») получают новости, связанные с действиями и заявлениями крупнейших игроков традиционного финансового мира или наиболее авторитетных фигур криптоиндустрии, идентифицируемых по маркерам вроде «BlackRock», «JPMorgan», «Elon Musk» или «Vitalik Buterin». Приоритет 3 (Средний, «Information Flow») присваивается новостям, исходящим от ведущих мировых или отраслевых информационных агентств, чьи имена (например, «Reuters», «Bloomberg», «CoinDesk») непосредственно присутствуют в поле «источник». Все прочие новости, не прошедшие фильтрацию по вышестоящим уровням, автоматически получают Приоритет 4 (Стандартный, «General Noise»), охватывая менее влиятельные издания и блоги.

Реализованный веб интерфейс предоставляет пользователю расширенные функциональные возможности для эффективной работы с информацией. Новостная лента автоматически сортируется по убыванию приоритета и дате публикации, а также может быть отфильтрована по упомянутым ранее параметрам. Система предлагает многоуровневое изложение новости: на карточке по умолчанию отображается краткая AI-сводка, а детальное изложение объемом около 300 слов генерируется LLM по запросу пользователя (ленивая загрузка), что оптимизирует использование ресурсов API и ускоряет первоначальную загрузку страницы. Все проанализированные новости сохраняются в локальной базе данных SQLite, обеспечивая персистентность данных и реализуя эффективный механизм кэширования для мгновенной отдачи результатов при повторных запросах. Архитектура предусматривает возможности кастомизации пользовательского промпта, позволяя адаптировать AI-анализ под индивидуальные инвестиционные стратегии, например, с акцентом на конкретные активы или технологические аспекты.

III. Реализация

Программное средство «Axon» реализовано как веб-приложение с разделением на серверную и клиентскую части, где выбор стека технологий определялся требованиями к производительности и скорости разработки. Серверная часть построена на Python 3 с использованием микро-фреймворка Flask [3] для создания RESTful API (/api/analyzed_news, /api/detailed_summary). Взаимодействие с внешними API (новостными и LLM) осуществляется через библиотеку Requests. Интеграция с Groq API [4] (llama-3.1-8b-instant) через официальную Python-библиотеку groq обеспечивает высокую скорость AI-анализа (<2 сек/статья). Персистентность данных и кэширование реализованы на SQLite3 посредством встроенного модуля sqlite3. Клиентская часть разработана на «ванильном» HTML5, CSS3 (с Flexbox/Grid) и JavaScript ES6 (с fetch API), что гарантирует быструю загрузку и отзывчивость. Тестирование показало: время ответа бэкенда без кэша ∼10–15 сек на 10 новостей, но при кэшировании снижается до <50 мс, подтверждая критическую эффективность механизма кэширования для пользовательского опыта.

Заключение

В рамках данной работы спроектировано и реализовано программное средство «Axon», предназначенное для автоматизации анализа новостного потока криптовалютного рынка. Система решает задачи снижения информационной перегрузки и устранения субъективности при оценке инфоповодов за счет интеграции больших языковых моделей (LLM) и авторского алгоритма контекстной приоритизации источников. Веб-интерфейс предоставляет пользователям оперативный доступ к агрегированным и структурированным данным, включая оценку тональности, автоматические сводки и прогнозируемый уровень влияния событий. Применение системы позволяет сократить временные затраты на первичный мониторинг рынка, минимизировать влияние когнитивных искажений и повысить обоснованность принимаемых инвестиционных решений.

1. Pang, B. Opinion mining and sentiment analysis / B. Pang, L. Lee // Foundations and Trends in Information Retrieval. – 2008. – Vol. 2, № 1–2. – P. 1–135.
2. Vaswani, A. Attention Is All You Need / A. Vaswani [et al.] // Advances in Neural Information Processing Systems 30 (NIPS 2017). – 2017. – P. 5998–6008.
3. Grinberg, M. Flask Web Development: Developing Web Applications with Python / M. Grinberg. – O'Reilly Media, 2018. – 256 p.
4. Groq API Documentation [Electronic resource] / Groq Inc. – Mode of access: `https://console.groq.com/docs`. – Date of access: 07.10.2025.

ГИБРИДНАЯ АДАПТИВНАЯ МОДЕЛЬ УПРАВЛЕНИЯ ИТ-КОМПАНИЕЙ В УСЛОВИЯХ РОСТА

Илюкевич А. А.
Директор ОАО "Центр банковских технологий"
Кафедра менеджмента, Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: iluk@tut.by

Статья посвящена изучению проблем управления быстрорастущими ИТ-компаниями в условиях их организационной трансформации и цифровизации. Выявлены ключевые управленческие вызовы, связанные с масштабированием бизнеса, и обоснована необходимость перехода от традиционных иерархических структур к адаптивным гибридным моделям, которые сочетают принципы процессного и проектного управления для обеспечения гибкости и устойчивости роста.

Введение

Рынок информационных технологий (ИТ) демонстрирует устойчивый рост в условиях быстрой цифровизации. Однако интенсивное масштабирование бизнеса, выражающееся в увеличении штата сотрудников и диверсификации продуктового портфеля, закономерно приводит к изменению организационной структуры и усложнению внутренних коммуникаций.

Проблема заключается в том, что традиционные управленческие модели, основанные на жесткой иерархии и функциональном подходе (наследие рациональной бюрократии М. Вебера), теряют свою эффективность при превышении определенного порога роста, что может привести к так называемому организационному кризису роста. [1] В фазе быстрого масштабирования возникает острый конфликт между необходимостью сохранения гибкости, характерной для проектного подхода, и потребностью в стандартизации и формализации, характерной для функционального подхода.

Исследования показывают, что типичными недостатками организационных структур ИТ-компаний являются разобщенность и слабая скоординированность деятельности структурных подразделений, негативно сказываясь на результативности предоставляемых услуг. Это противоречие ставит под угрозу оперативность принятия решений и требует переосмысления базы накопленных знаний в области управления для формирования новой научной поддержки бизнеса в вопросах организационной трансформации.[2-4]

I. Ключевые вызовы роста и структурные ограничения

Рост организации в долгосрочной перспективе неизбежно приводит к увеличению числа иерархических уровней управления, требуя перегруппировки сил и адаптации механизмов менеджмента. В условиях ИТ-отрасли, характеризующейся высокой динамичностью внешней среды, традиционные подходы порождают следующие проблемы:[4]

- Классические линейно-функциональные схемы, основанные на функциональной департаментизации, приводят к сверхспециализации, создавая "непроницаемые организационные перегородки которые ослабляют горизонтальные связи и порождают функционализм — организационную болезнь, при которой размывается общая цель организации.
- Увеличение иерархических уровней и широкий масштаб управляемости могут стать причиной возникновения «эффекта бутылочного горла» на уровне высшего руководства, которое оказывается перегружено оперативной работой, что приводит к потере контроля над стратегическими задачами.
- В проектном подходе, доминирующем в ИТ-сфере, даже исторически распространенная каскадная модель (Waterfall) остро проявляет недостаток в виде слабой скоординированности деятельности подразделений при реализации проекта.

Разрешение этих кризисов требует перехода к адаптивным (органическим) структурам. Органический тип характеризуется децентрализацией, гибкой структурой власти и эффективен в условиях нерутинной технологии. В высокотехнологичной сфере наиболее эффективна адхократия, где власть основана не на иерархическом положении, а на экспертных знаниях и компетентности. [5]

II. Комплексная адаптивная модель управления (КАМУ)

Для обеспечения устойчивого роста ИТ-компании в условиях цифровизации критически необходим гибридный подход к управлению, который стратегически сочетает проектную гибкость и органическую структуру для инноваций с процессной формализацией для обеспечения контроля и управляемости при масштабировании.

Разработанная Комплексная Адаптивная Модель Управления (КАМУ) является таким гибридным синтезом, структурированным через три взаимосвязанные подсистемы:

структурно-функциональную, информационно-поведенческую и подсистему саморазвития.

Структурно-функциональная подсистема:

- Гибридная департаментизация: Происходит отказ от чистой функциональной схемы в пользу матричных или адхократических элементов. КАМУ сочетает функциональную специализацию (для обеспечения экспертизы) с проектным подходом (для фокусировки на конечном результате).
- Децентрализация и Проектный менеджмент: Вводится институт Проект-менеджера для координации команды и контроля сроков. Формируются полуавтономные команды под конкретный проект/цель, что усиливает горизонтальные и неформальные связи.[5]

Информационно-поведенческая подсистема:

Цифровая трансформация выступает ключевым связующим звеном, устраняющим кризис координации. Центральный элемент – переход к принятию решений на основе данных (Data-Driven Decision Making) и системное сокращение рутинных элементов в деятельности персонала.[7] Внедрение Систем управления задачами и проектами (СУЗП) и Корпоративных систем управления знаниями (КСУЗ) является важным условием для обеспечения формальных внутрифирменных коммуникаций и прозрачности, а также устраняет информационную разобщенность. Модель требует предпринимательской культуры и партисипативного управления (привлечение работников к установлению целей). Мотивация акцентируется на нематериальных стимулах (профессиональный рост, самоактуализация).[7-10]

Подсистема саморазвития:

Эта подсистема отвечает за гибкость и адаптивность к изменениям. Она ориентирована на стратегическое предвидение. Цифровой двойник организации (DTO) ичспользуется как ключевой инструмент формального моделирования и прогнозирования сценариев роста и изменений, позволяя руководству избегать реальных ошибок. Для радикальных преобразований используется реинжиниринг бизнес-процессов. Ротация кадров применяется как метод обучения и развития, обеспечивая движение системы управления по горизонтали.[11]

III. Заключение

Комплексная Адаптивная Модель Управления (КАМУ) является необходимым ответом на вызовы, которые ставит перед ИТ-компаниями этап интенсивного роста. Она обеспечивает баланс между гибкостью органической структуры (для инноваций) и цифровой стандартизацией процессов (для управляемости).

Практическая значимость модели заключается в предоставлении количественных метрик для верификации ее эффективности:[3]

- Снижение текучести кадров (индикатор стабильности информационно-поведенческой подсистемы).
- Сокращение Time-to-Market и рост производительности команд.
- Уменьшение «эффекта бутылочного горла».

Внедрение КАМУ требует системной трансформации, охватывающей структурный реинжиниринг, внедрение ITSM-подхода и использование DTO как инструмента непрерывного самосовершенствования.[7-8] КАМУ обеспечивает организации способность сохранять системную целостность и адаптивность в условиях высокодинамичной внешней среды, подтверждая, что гибридный подход является единственным путем к преодолению кризисов роста в ИТ-отрасли.

Дальнейшее развитие темы представляется целесообразным сосредоточить на эмпирической верификации предложенной модели на основе корреляционно-регрессионного анализа фактических данных ИТ-компаний, успешно совершивших переход к адаптивному управлению.

1. Виханский, О. С. Менеджмент: учебник / О. С. Виханский, А. И. Наумов. – 4-е изд., перераб. и доп. – М.: Экономистъ, 2006. – 670 с.
2. Мескон, М. Х. Основы менеджмента / М. Х. Мескон, М. Альберт, Ф. Хедоури; пер. с англ. – М.: Дело, 2000. – 704 с.
3. Калязина, Е. Г. Трансформация подходов к управлению в условиях цифровизации / Е. Г. Калязина, Е. Ю. Плешакова // Теория и практика общественного развития. – 2024. – № 6. – С. 85–93.
4. Головина, О. Д. Организация деятельности IT-компаний в области управления проектами / О. Д. Головина, О. А. Воробьева // Вестник Удмуртского университета. Серия Экономика и право. – 2023. – Т. 33, № 2. – С. 223–228.
5. Беляцкая, Т. Н. Предпринимательская деятельность и управление проектами в ИТ-сфере: учеб. пособие. – Минск: РИПО, 2023. – 256 с.
6. Иванов, Д. А. Современные подходы к управлению ИТ. Процессная парадигма / Д. А. Иванов // CIO. – 2011. – № 1. – С. 24–30.
7. Романова, Ю. Д. Информационные технологии в менеджменте (управлении): учебник и практикум / под ред. Ю. Д. Романовой. – М.: Издательство Юрайт, 2018. – 478 с.
8. Петров, И. С. Современные информационные системы поддержки управленческих решений в IT-отрасли / И. С. Петров // Экономика и управление. – 2022. – № 2. – С. 45–53.
9. Козлов, А. В. ITSM-подход в управлении ИТ-инфраструктурой сети МПК / А. В. Козлов // Вестник транспорта. – 2020. – № 3. – С. 112–120.
10. Сидорова, О. Л. Сервисный подход в управлении ИТ-подразделениями / О. Л. Сидорова // IT Expert. – 2018. – № 4. – С. 15–22.
11. Беляцкий, Н. П. Основы лидерства: учебник. – Минск: БГЭУ, 2010. – 256 с.

ОБЪЯСНИМЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И МЕТОД SHAP КАК ОСНОВА ПРОЗРАЧНОГО ПРОГНОЗИРОВАНИЯ В АГРОПРОМЫШЛЕННОМ КОМПЛЕКСЕ РЕСПУБЛИКИ БЕЛАРУСЬ

Клименко И. В.
Кафедра экономической информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: znachnamail@gmail.com

В статье рассматривается применение методов объяснимого искусственного интеллекта (XAI) в контексте цифровой трансформации агропромышленного комплекса Республики Беларусь. Раскрываются возможности интерпретации сложных моделей машинного обучения с использованием методов PDP, ICE, LIME, Permutation Importance и особенно SHAP, позволяющего количественно оценивать вклад факторов в прогноз. На примере прогнозирования урожайности пшеницы и анализа молочных удоев показано, как объяснимость повышает доверие к цифровым решениям и способствует управляемости производственных процессов. Сформулированы рекомендации по институционализации XAI в национальных информационных системах АПК.

Введение

Цифровая трансформация АПК Беларуси рассматривается как ключ к росту эффективности, конкурентоспособности и устойчивости как отрасли, так и национальной экономики. В данный момент внедряются элементы точного земледелия, мониторинг животных, облачные отраслевые ИС, однако прогресс неравномерен из-за кадровых и инфраструктурных ограничений, а также технологического отставания и низкой скорости внедрения инноваций [1]. Нормативную рамку внедрения цифровизации формируют Декрет № 8 и Госпрограмма «Цифровое развитие Беларуси», задающие стратегические ориентиры обновления сектора [2].

Главным барьером внедрения цифровых решений выступает недоверие управленцев и работников к «чёрным ящикам»: в прогнозах непрозрачных моделей людям сложно обнаруживать внятную логику, что снижает готовность к их применению [3]. В этих условиях раскрывается необходимость внедрения объяснимого ИИ (XAI): он раскрывает логику сложных моделей без заметной потери точности и превращает их из закрытого механизма в инструмент совместного анализа, дополняющий профессиональную экспертизу [1; 3].

Основная часть

Методы объяснимого искусственного интеллекта различаются по степени детализации анализа и нередко используются совместно, поскольку дают разные углы понимания поведения модели. Так, PDP (Partial Dependence Plot) отображает, как в среднем изменяется прогноз при варьировании одного из факторов, например уровня влажности или кислотности почвы, при прочих равных условиях. ICE (Individual Conditional Expectation) показывает подобную зависимость, но уже для каждого отдельного наблюдения – то есть для конкретного поля, хозяйства или региона, что позволяет учитывать индивидуальные особенности. LIME (Local Interpretable Model-Agnostic Explanations) строит простую приближенную модель вокруг конкретного решения, помогая понять, почему алгоритм сделал именно этот выбор. Permutation Importance оценивает важность признаков, случайно перемешивая их значения и проверяя, насколько сильно при этом падает точность модели [3].

Особое место среди методов объяснимости занимает SHAP (Shapley Additive Explanations), основанный на теории кооперативных игр. Он разлагает прогноз модели на вклады отдельных факторов, выраженные в тех же единицах, что и сам результат. Это позволяет не только определить важные признаки, но и количественно оценить их влияние: например, насколько температура или уровень осадков изменили ожидаемую урожайность. Для управленцев такой подход особенно ценен, поскольку делает возможным осмысленный количественный анализ решений.

Метод сопровождается наглядными визуализациями – beeswarm, summary bar, dependence и waterfall графиками, показывающими, как каждый фактор влияет на прогноз. Такие инструменты уже применяются в аграрных исследованиях: от выбора оптимальных культур до прогнозирования неисправностей техники и оценки производительности труда [4-8]

Использование SHAP требует осторожности и понимания его ограничений. При высокой корреляции признаков (например, температуры и влажности) метод может ошибочно распределять вклад между ними, так как классические варианты, такие как Kernel SHAP, подставляют нереалистичные комбинации данных. Для повышения достоверности применяются Conditional SHAP и

группировка связанных признаков [9]. Кроме того, SHAP объясняет корреляции внутри модели, но не причинно-следственные связи, что требует подтверждения выводов экспериментами или каузальным анализом [10-11].

Практическое применение методов объяснимого искусственного интеллекта наглядно демонстрирует анализ SHAP, используемый для интерпретации прогностических моделей в сельском хозяйстве. В исследовании по прогнозированию урожайности яровой пшеницы в Алтайском крае (2001–2024 гг.) модель XGBoost, обученная на 18 метео- и агротехнических факторах, показала высокую точность (R^2 = 0,95; MAE = 0,13 т/га) [6]. SHAP выявил ключевые детерминанты: количество осадков, паровой предшественник и дозы удобрений. Оптимальное увлажнение (596 мм) повышало прогноз на +1,19 т/га, дефицит влаги (317 мм) снижал на – 0,77 т/га. Это позволило обосновать решения по севообороту и нормам удобрений.

Схожие результаты получены в проекте «цифрового двойника» молочной фермы, где ансамбль моделей (SVR, AdaBoost, ANN) применялся для прогноза удоя и оценки экологической эффективности [7]. SHAP показал, что месяц лактации снижает удои, тогда как состав и энергетическая насыщенность рациона её повышают. Избыток азота увеличивает концентрацию мочевины в молоке, а рост удоя и массы животного, наоборот, снижает её уровень, что позволяет оперативно корректировать кормление и повышать эффективность производства.

Выводы

Таким образом, объяснимый искусственный интеллект меняет подход к цифровому анализу и управлению в аграрной сфере. Он делает работу моделей прозрачной и воспроизводимой, превращая алгоритмы из «чёрных ящиков» в понятные инструменты принятия решений.

Для агропромышленного комплекса Беларуси метод SHAP служит связующим звеном между точностью и управляемостью аналитических систем, позволяя интерпретировать прогнозы через вклад факторов и повышая доверие к цифровым решениям. Практические примеры, такие как прогноз урожайности, анализ молочной продуктивности и техническая диагностика, подтверждают, что объяснимость делает ИИ-системы более надёжными и полезными.

Внедрение XAI требует институциональной поддержки: разработки стандартов объяснимости, включения таких методов в национальные информационные системы и подготовки специалистов. Расширение образовательных модулей по XAI в рамках программы «Цифровое развитие Беларуси» обеспечит формирование кадровой базы для осмысленной и прозрачной цифровой трансформации агропромышленного комплекса страны.

1. Проблемы и перспективы цифровизации аграрной экономики в Республике Беларусь / К. В. Павлов, И. В. Зенькова, И. А. Поздянкова [и др.] // Вестник УГНТУ. Наука, образование, экономика. Серия экономика. – № 2 (44). – 2023. – С. 7–13.
2. Сырокваш, Н. А. О развитии цифровизации АПК Беларуси / Н. А. Сырокваш, О. Л. Сапун // Актуальные проблемы инновационного развития агропромышленного комплекса Беларуси: сб. науч. тр. по материалам XIV Междунар. науч.-практ. конф., Горки, 30–31 мая 2022 г. : в 2 ч. / [редкол. : И. В. Шафранская (отв. ред.) и др.]. – Горки. – 2023. – Ч. 2. – С. 54–58.
3. Шевская, Н. В. Объяснимый искусственный интеллект и методы интерпретации результатов [Электронный ресурс]. – Режим доступа: https://moitvivt.ru/ru/journal/pdf?id=1005. – Дата доступа: 12.09.2025.
4. Akkem, Y. Role of Explainable AI in Crop Recommendation Technique / Y. Akkem, S. K. Biswas, A. Varanasi // I. J. Intelligent Systems and Applications. – 2025. – Vol. 1. – P. 31–52.
5. Suresh, S. E., Venkateswara Reddy K. Explainable AI for Predictive Maintenance in Smart Agricultural Facilities / S. E. Suresh // International Journal of Research Publication and Reviews. – 2025. – Vol. 6, Issue 5. – P. 11806–11808.
6. Интерпретируемая модель машинного обучения для прогнозирования урожайности яровой пшеницы / В. К. Каличкин, В. И. Усенко, А. А. Гаркуша [и др.] // Российская сельскохозяйственная наука. – 2025. – № 2. – С. 59-66
7. Engineering-Oriented Explainable Machine Learning and Digital Twin Framework for Sustainable Dairy Production and Environmental Impact Optimisation / R. Xing, B. Li, S. Dora [et al.] // Algorithms. – 2025. – 18(10):670. – Mode of access: https://www.mdpi.com/1999-4893/18/10/670. – Date of access: 20.10.2025.
8. A Proactive Predictive Model for Machine Failure Forecasting / O. O. Ajayi, A. M. Kurien, K. Djouani [et al.] // Machines. – 2025. – 13(8):663. – Mode of access: https://www.mdpi.com/2075-1702/13/8/663. – Date of access: 20.10.2025.
9. Aas, K. Explaining individual predictions when features are dependent: more accurate approximations to Shapley values / K. Aas, M. Jullum, A. Løland // Artificial Intelligence. – 2021. – Vol. 298. Article 103502. – Mode of access: https://arxiv.org/pdf/1903.10464. – Date of access: 20.10.2025.
10. Takefuji, Y. Reevaluating feature importance in machine learning: concerns regarding SHAP interpretations in the context of the EU artificial intelligence act / Y. Takefuji // Water Research. – 2025. – Vol. 280. Article 123514. – Mode of access: https://doi.org/10.1016/j.watres.2023.123514. – Date of access: 20.10.2025.
11. Be careful when interpreting predictive models in search of causal insights (blog article) / E. Dillon, J. LaRiviere, S. M. Lundberg [et al.] // Tech Community Microsoft (SHAP documentation). – Mode of access: https://shap.readthedocs.io/en/latest/example_notebooks/overviews/Be%20careful%20when%20interpreting%20predictive%20models%20in%20search%20of%20causal%20insights.html. – Date of access: 20.10.2025.

ВЛИЯНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА ЦИФРОВОЙ МАРКЕТИНГ

Кропотин Д. Д.
Кафедра экономической информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: krapotsin.daniil@gmail.com

В данной статье рассматривается влияние искусственного интеллекта на цифровой маркетинг и его развитие. Определены основные направления внедрения искусственного интеллекта в маркетинговую деятельность организаций, включая анализ данных, персонализацию контента и другие. Приведены примеры использования искусственного интеллекта в цифровом маркетинге и управлении взаимоотношениями с клиентами, выделяются преимущества и риски применения искусственного интеллекта в маркетинге.

Введение

Развитие цифровых технологий существенно изменило подход к маркетингу. В условиях роста объемов данных, многоканальности коммуникаций и повышения требований потребителей значительно снижается эффективность традиционных инструментов маркетинга и существующих подходов к решению задач, поставленных перед маркетологами. Инновационные подходы в маркетинге включают использование технологий искусственного интеллекта и машинного обучения для анализа потребительских данных [1]. Технологии искусственного интеллекта обеспечивают интеллектуальный анализ данных, автоматизацию и упрощение процессов, необходимых для достижения высоких результатов и повышения удовлетворенности клиентов. Актуальность исследования обусловлена тем, что внедрение технологий искусственного интеллекта позволяет компаниям формировать индивидуальные предложения для клиентов, повышать точность прогнозов и оптимизировать расходы на продвижение на основе анализа целевой аудитории, потребностей клиентов, посадочных страниц. Данные технологии также упрощают сбор семантического ядра, написание рекламных объявлений, составление контент-плана и многое другое.

I. Роль искусственного интеллекта в цифровом маркетинге

Искусственный интеллект охватывает совокупность технологий, включающих машинное обучение, обработку естественного языка, интеллектуальную аналитику. В маркетинге данные инструменты используются для автоматизации аналитических и коммуникационных процессов. Наиболее распространенные направления применения искусственного интеллекта в маркетинге:

- анализ и интерпретация больших массивов данных (Big Data);
- прогнозирование поведения и потребностей клиентов;
- персонализация контента и рекламы;
- динамическое ценообразование;
- управление клиентским опытом через чат-боты и виртуальных ассистентов;
- оптимизация медиапланирования.

Использование искусственного интеллекта позволяет перейти от массового к адресному маркетингу, формируя индивидуальные сценарии взаимодействия с клиентом. Алгоритмы машинного обучения выявляют скрытые закономерности, что делает возможным точное прогнозирование реакций аудитории на рекламные сообщения.

II. Практическое применение технологий искусственного интеллекта

Применение искусственного интеллекта в цифровом маркетинге охватывает все стадии маркетингового цикла – от анализа целевой аудитории до оценки эффективности кампаний.

1. Аналитика и прогнозирование. С помощью алгоритмов машинного обучения осуществляется сегментация клиентов, определение их ценности для компании, выявление потенциальных оттоков и формирование индивидуальных предложений. Примером может служить использование RFM-анализа и усовершенствованных аналитических моделей, применяемых в компании, в сочетании с моделями кластеризации для прогнозирования поведения клиентов.

2. Персонализация и автоматизация коммуникаций. Технологии обработки естественного языка (NLP) позволяют анализировать тональность отзывов, создавать текстовые описания товаров и вести диалог с клиентами. Чат-боты, в которые внедрен искусственный интеллект, обеспечивают круглосуточное взаимодействие. Искусственный интеллект может вести десятки диалогов одновременно и обрабатывает запросы быстрее операторов [2].

3. Контент-маркетинг и генерация креативов. Генеративные модели (в частности, GPT и diffusion-модели) применяются для автоматического создания рекламных текстов, визуальных материалов и сценариев видеорекламы. Это позволяет значительно ускорить процесс разработки

креативных материалов и адаптировать их под различные целевые аудитории.

4. Оптимизация рекламы и ценообразования. Маркетинговые платформы используют алгоритмы искусственного интеллекта для автоматических стратегий назначения ставок. На основе анализа откликов пользователей на рекламу корректируются параметры аукционов, ставки и размещение объявлений. Системы динамического ценообразования, основанные на прогнозных моделях спроса, позволяют устанавливать наиболее выгодные цены с учетом рыночных факторов и поведения конкурентов. Примеры инструментов, использующих технологии искусственного интеллекта в цифровом маркетинге, представлены в таблице 1.

Таблица 1 – Примеры применения искусственного интеллекта в цифровом маркетинге

Инструмент / Платформа	Технология ИИ	Назначение / Применение
Google Ads Smart Bidding	Машинное обучение	Автоматическая оптимизация ставок на основе вероятности конверсии
Meta Advantage+	Предиктивная аналитика	Оптимизация креативов и аудитории в рекламных кампаниях
ChatGPT / Jasper AI	Генеративные языковые модели (NLP)	Создание текстового контента, описаний товаров, рекламных сообщений
DALL·E, Midjourney	Диффузионные нейросети	Генерация изображений и визуальных креативов
HubSpot AI Tools	Аналитика	Автоматизация CRM, прогнозирование потребностей

Применение искусственного интеллекта в маркетинге обеспечивает следующие преимущества:

1. повышение точности прогнозирования продаж и спроса;
2. рост конверсии за счет персонализированных предложений;
3. сокращение затрат на анализ, планирование;
4. повышение уровня удовлетворенности клиентов;
5. возможность непрерывного обучения и адаптации маркетинговых стратегий.

Однако, несмотря на преимущества, наблюдаются и риски. К числу основных из них можно отнести:

- качество алгоритмов принятия решений;
- расхождения с более детальным прогнозированием, выполненным специалистами;
- угрозы утечки и несанкционированного использования персональной информации клиентов;
- высокая стоимость внедрения решений и подготовки специалистов в области технологий искусственного интеллекта.

Кроме того, чрезмерная автоматизация коммуникаций может привести к потере «человеческого» аспекта взаимодействия, снижая эмоциональную вовлеченность потребителя. Помимо этого, технологии искусственного интеллекта имеют множество недоработок, которые могут привести к критическому состоянию не только процесса, но и организации в целом.

III. Заключение

Влияние искусственного интеллекта на цифровой маркетинг носит системный характер. Искусственный интеллект не только повышает эффективность маркетинговых коммуникаций, но и трансформирует саму концепцию взаимодействия с клиентом, делая ее более интеллектуальной, динамичной и адаптивной.

Компании, активно внедряющие ИИ-инструменты, получают конкурентные преимущества за счет точного анализа данных, предиктивного моделирования и персонализации предложений. В перспективе развитие технологий машинного обучения, нейросетевых моделей и генеративных систем приведет к появлению новых форм цифрового маркетинга, основанных на автономных интеллектуальных платформах. Дальнейшие исследования могут быть направлены на разработку гибридных моделей, сочетающих машинное обучение и эвристический анализ, а также на внедрение искусственного интеллекта в мультиканальную аналитику [3].

При этом ключевой задачей остается поиск баланса между эффективностью автоматизации и ошибками, возникающими в результате использования искусственного интеллекта, что требует выявления оптимального подхода для того, чтобы принимать как можно более эффективные маркетинговые решения.

IV. Список литературы

1. Тихонов, А. В. Инновационные подходы к маркетингу и управлению бизнес-процессами в эпоху цифровых технологий / А. В. Тихонов // Профессиональный Вестник: Экономика и управление. – 2024. – № 1. – С. 20-23.
2. Горчаков, И. Ю. Направления реализации возможностей искусственного интеллекта в управлении бизнесом / И. Ю. Горчаков, Ю. Е. Галкина // Журнал У. Экономика. Управление. Финансы. – 2025. – № 2(40). – С. 370-379.
3. Шардаков, Е. А. Применение искусственного интеллекта для анализа поведения пользователей на вебсайтах / Е. А. Шардаков // Вестник науки. – 2025. – Т. 2, № 4(85). – С. 777-781.

ВИЗУАЛЬНЫЕ КОММУНИКАЦИИ В РЕКЛАМЕ: КУЛЬТУРНЫЕ ОСОБЕННОСТИ И ЭФФЕКТИВНОСТЬ ВОСПРИЯТИЯ

Марченкова Е. Е.
Кафедра экономики, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: marchenkova@bsuir.by

В статье исследуются теория и практика визуальных коммуникаций в рекламной сфере: основные элементы визуального воздействия, их психологическое восприятие, статистика эффективности, современные технологии. Особое внимание уделяется специфике визуального восприятия рекламы в Беларуси и России.

Введение

Современная реклама всё чаще основывается на визуальных формах коммуникации, которые обеспечивают быстрое восприятие информации и эмоциональный отклик аудитории. В условиях роста информационной насыщенности и конкуренции за внимание потребителя визуальные средства становятся важнейшим инструментом формирования имиджа бренда. Исследование особенностей визуального восприятия в разных культурных контекстах позволяет адаптировать рекламные стратегии к специфике целевых аудиторий.

I. Культурные особенности визуального восприятия

Люди из западных индивидуалистических культур (например, США, Великобритания) склонны фокусироваться на центральных или ключевых объектах сцены: они быстро замечают “главные” элементы. В то время как в культурах с коллективистским уклоном (например, многие восточноазиатские) внимание распределяется более широко: они более чувствительны к взаимоотношениям между объектами и фоном. Например, достоверно известно, что представители разных культур по-своему воспринимают "позитивные"эмоции в рекламе:выражение радости, улыбка, степень возбуждения воспринимаются с разной степенью интенсивности. Например, американцы лучше различают “высокую активность” (HA – high arousal) и “среднюю/низкую” активность в эмоциональных выражениях, тогда как японцы менее чувствительны к этой градации.

Практически это означает, что визуальные образы с выраженными эмоциональными лицами и жестами могут сильно отличаться в том, как они воспринимаются в разных культурах (более или менее “сильно”, более или менее “искренне”). Кроме того, люди из разных культур по-разному ассоциируют цвета с настроением и ценностями. Например, в Европе и Китае разные ассоциации цветов: белый в Европе – символ чистоты, а в Китае – траура. Жёлтый в Европе ассоциируется с предательством, в то время как в Китае – с императорской властью и богатством. Красный – символ тревоги и агрессии в Европе, а в Китае это цвет удачи и праздника. Зелёный в Европе символизирует природу и рост, а в Китае может означать измену.

В исследовании “Cultural Differences & Visual Attention” отмечается, что участники из коллективистских культур демонстрируют больше фиксаций, дольше исследуют детали сцены, менее “ускоренный” поиск, чем участники из более индивидуалистичных культур [1]. Это может означать, что реклама, ориентированная на аудиторию с “широким” визуальным вниманием, выиграет, если содержит “фоновые” детали, визуальные связи, контекст, а не просто сильный центральный объект.

Эти особенности объясняют необходимость адаптации визуальных коммуникаций под культурные коды. Исследование J. Engelmann (2021) показало, что визуальные образы, соответствующие культурным ожиданиям аудитории, вызывают эмоциональный отклик на 40–45 % выше, чем универсальные нейтральные изображения [2].

II. Визуальные коммуникации на российском рынке

Для российской рекламы характерны яркость, контрастность и эмоциональная выразительность. Я. Грибер (2022) отмечает, что русскоязычная аудитория демонстрирует высокую чувствительность к насыщенным цветам и динамичным визуальным элементам [3].

Согласно данным CNews (2025), каждый второй россиянин при выборе товара ориентируется на качество изображения и оформление упаковки. Аналитики фиксируют рост эффективности визуальной рекламы: кликабельность графических объявлений в интернете выше текстовых в среднем на 37 % [4]. Исследования в области нейромаркетинга подтверждают, что визуальные стимулы активируют зоны мозга, отвечающие за удовольствие и память.

В последние годы российская реклама активно обращается к культурным архетипам – традициям, фольклору, мотивам семьи и природы. Российский рынок рекламы отличается стремлением

к сочетанию выразительности и функциональности визуального ряда. Среди доминирующих направлений выделяются минимализм, использование национальных мотивов и эмоционально окрашенные сюжеты, апеллирующие к семейным ценностям и традициям.

В России многие крупные компании активно используют визуальный сторителлинг для продвижения своих продуктов и сервисов. Например, компания «Яндекс» применяет анимацию и яркие визуальные истории для продвижения карт, такси и маркетплейса, используя простую иконографию и динамичные цвета, что значительно повышает вовлечённость пользователей. Банк «Сбер» строит свои визуальные кампании вокруг фирменного зелёного цвета, лаконичных графических элементов и анимации, что усиливает ощущение технологичности и надёжности. Розничная сеть «М.Видео» использует яркие баннеры, акционные цвета и динамичные GIF-анимации для цифровых рекламных площадок, а сеть «Пятёрочка» применяет контрастные цвета, лаконичные шрифты и крупные цифры акций, что привлекает внимание и повышает конверсию покупателей.

III. Визуальные предпочтения и рекламные тенденции в Беларуси

Белорусский рынок рекламы демонстрирует схожие тенденции, но с выраженным акцентом на умеренность и природную гармонию. Ключевыми аспектами визуального восприятия белорусских потребителей являются цветовая гамма, типографика и используемые изображения. Природные и приглушённые оттенки, такие как зелёный, голубой и бежевый, вызывают ассоциации с надёжностью, стабильностью и экологичностью, что делает визуальное оформление более доверительным для аудитории. Чёткие шрифты и аккуратная верстка текста воспринимаются как показатель профессионализма и усиливают ощущение качества продукта. Кроме того, использование фотографий с живыми эмоциями и природными мотивами способствует более сильному эмоциональному вовлечению потребителей и формирует позитивное впечатление о бренде.

Социальные сети играют важную роль в восприятии брендов. Короткие визуальные сюжеты и лаконичные визуальные сообщения обеспечивают высокий уровень вовлечения, особенно у молодых пользователей.

Многие белорусские компании успешно используют визуальный сторителлинг для создания уникального имиджа. Например, компания «Коммунарка» строит визуальные рекламные кампании вокруг традиционных ценностей, подчёркивая натуральность продуктов и качество изготовления. Такой подход усиливает доверие к бренду и делает его более узнаваемым. Бренд «Савушкин продукт» использует визуальные элементы, подчёркивающие свежесть и экологичность своей молочной продукции, формируя у покупателей ощущение натуральности и безопасности. Аналогичным образом компания «Санта Бремор» сочетает современный дизайн упаковки с национальными мотивами, что создаёт эмоциональную связь с аудиторией и повышает вовлечённость потребителей.

Для эффективного использования визуального сторителлинга в Беларуси рекламодатели должны учитывать культурные и эмоциональные особенности аудитории. Они должны акцентировать внимание на локальных мотивах и культурных кодах, включая орнаменты, природные символы и элементы национальной архитектуры. Следует придерживаться визуальной умеренности, используя лаконичные композиции, мягкую цветовую палитру и минимализм, что повышает доверие аудитории. Важно подчёркивать экологичность и натуральность продукции с помощью изображений природы, свежих продуктов и натуральных материалов. Рекламодатели должны создавать визуальные истории, включая короткие сюжетные ролики и карусели в социальных сетях, чтобы повышать эмоциональное вовлечение и запоминаемость бренда.

IV. Список литературы

1. Masuda N. Cultural effects on visual attention: Eye-tracking evidence [Electronic resource] / N. Masuda, T. Nisbett // Journal of Personality and Social Psychology. – APA, 2010. – Mode of access: https://pubmed.ncbi.nlm.nih.gov/20234851/ – Date of access: 17.10.2025.
2. Engelmann J. B. Cultural differences in perception and emotion regulation [Electronic resource] / J. B. Engelmann, P. D. Mauss // Annual Review of Psychology. – Stanford University Press, 2021. – Mode of access: https://pubmed.ncbi.nlm.nih.gov/22084635/ – Date of access: 17.10.2025.
3. Griber Y. A. Colour terms and object referents in the Russian language [Electronic resource] / Y. A. Griber // Language and Cognition. – Cambridge University Press, 2022. – Mode of access: https://pubmed.ncbi.nlm.nih.gov/ – Date of access: 17.10.2025.
4. Каждый второй россиянин выбирает товар по изображению [Электронный ресурс] / CNews. – Москва, 2025. – Режим доступа: https://www.cnews.ru/news/ – Дата доступа: 17.10.2025.

СОВЕРШЕНСТВОВАНИЕ МЕТОДОВ УПРАВЛЕНИЯ ЛИКВИДНОСТЬЮ БАНКА

Нагулевич Р. С.
Кафедра экономической информатики,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: nagulevicrita@gmail.com

Введение

Основой для устойчивого функционирования кредитно-финансового учреждения является его нормативный капитал. Банки обязаны поддерживать его в размере, достаточном для покрытия принимаемых им рисков [1]. Требование к управлению рисками при осуществлении деятельности банками Республики Беларусь содержится в инструкции «Об организации корпоративного управления» [2], в соответствии с которой банк выявляет риск ликвидности, определяет его источники возникновения, проводит «измерение (оценку), мониторинг (контроль) и ограничение (снижение)» уровня [2, с. 17]. Требование об обеспечении достаточности нормативного капитала вынуждает банки постоянно совершенствовать методы оценки и процедуры управления рисками.

I. Результаты и их обсуждение

Для оценки (измерения) риска ликвидности банки используется метод, основанный на анализе GAP-разрывов (GAP-анализ). Метод предполагает расчет несбалансированности сумм требований (активов) и обязательств (пассивов) банка по срокам их погашения (востребования). Разность между активами и пассивами в разрезе сроков и формирует чистую позицию, которую называют GAP-разрывом. Пример расчета GAP-разрывов приведен в таблице (рис. 1).

Расчет суммы риска ликвидности сводится к определению будущей потребности банка в ликвидных активах на каждом из интервалов срочности и расходов для закрытия отрицательных накопленных позиций ликвидности по всем интервалам срочности, где отрицательные GAP-разрывы не компенсированы положительными разрывами в предыдущих периодах.

Название показателя/ Сроки погашения	До востребования	1-30 дней	…	331-365 дней	свыше 1 года	Всего
ИТОГО АКТИВОВ	5 872	55 955	578	5 948	23 134	91 488
ИТОГО ПАССИВОВ	14 258	45 878	690	3 656	282	64 765
ПОКАЗАТЕЛИ РИСКА ЛИКВИДНОСТИ						
GAP-разрыв	- 8 386	10 077	- 112	2 292	22 852	
Итого активов нарастающим итогом	5 872	61 827	62 405	68 354	91 488	
Итого пассивов нарастающим итогом	14 258	60 136	60 826	64 483	64 765	
Накопленный GAP-разрыв	- 8 386	1 691	1 579	3 871	26 723	
Середина интервала (дней)	15	60	135	270		
СТОИМОСТЬ ПРИВЛЕЧЕНИЯ РЕСУРСОВ , %	15,0					
Расходы по привлечению	- 1 205	211	148	145		
Сумма потерь по риску ликвидности	- 701	млн.руб				
Условно-постоянный остаток	3 100					
Сумма потерь по риску ликвидности с учетом УПО	- 255	млн.руб				

Рис. 1 – Оценка риска ликвидности по методу GAP-разрывов

Как видно из приведенного анализа GAP-разрывов (рис. 1) риск ликвидности составил 701,0 млн.руб.

Алгоритм GAP-расчета имеет недостатки: точность анализа зависит от широты временных интервалов, в пределах которых группируются активы и пассивы (чем шире интервал, тем менее точен результат расчета). Принимаемое допущение о неизменности объемно-временной структуры баланса банка на горизонте оценки (до 1 года) делает метод статичным. Это означает проведение политики активно-пассивных операций, направленной на достижение «нулевого GAP», без учета ежедневного изменения структуры.

С учетом отмеченных выше недостатков метода и существенности риска для банка предлага-

ется модифицировать алгоритм расчета и ввести в базовую статичную модель элементы модели динамической.

Учет стабильной части средств.

Дополним GAP-модель в части пассивов (обязательств) расчетной суммой в виде условно-постоянного остатка средств (далее–УПО) в разрезе видов валют:

$$УПО = k * V * P * T,$$

где:

k – квантиль нормального распределения (для 95% вероятности: 1,645);

V – волатильность (стандартное отклонение), определяемое по ежедневному изменению остатков средств, участвующих в алгоритме расчета УПО;

P – сумма (остаток) средств, участвующих в алгоритме расчета УПО, на дату проведения анализа;

T – корень квадратный из числа рабочих дней в месяце, следующем за датой расчета.

В качестве суммы остатков (P) на дату расчета используются: средства на текущих расчетных счетах субъектов хозяйствования; средства физических лиц на счетах «до востребования»; отзывные срочные средства физических лиц и субъектов хозяйствования на депозитах (вкладах).

Как показывает расчет, учет *УПО* (рис. 1) существенно снизил величину возможного убытка: GAP-разрыв компенсируется средствами, которые с вероятностью в 95% составят величину не менее рассчитанной (3800,0 млн руб.). Величина риска с учетом в обязательствах банка стабильной части средств (*УПО*) составила величину в 255,0 млн руб.

Расчет резервов ликвидности.

Для учета статистических характеристик денежных потоков в процессе управления ликвидностью банка (в т. ч. и риском ликвидности) предлагается с использованием модели для GAP-анализа провести расчет двух видов резервов ликвидности: требуемого (необходимого) и фактического.

Необходимый (требуемый) резерв (Vmp) рассчитывается с учетом ежедневного изменения остатков средств, сложившегося за последние 10 лет по следующим счетам: депозитным и текущим – субъектов хозяйствования; отзывным, безотзывным, до востребования – физических лиц; средств, привлеченных посредством ценных бумаг – клиентов и банков; корреспондентским и депозитным – других банков.

Расчет производится с применением VaR – методики с доверительной вероятностью в 95% (квантиль нормального распределения 1,645). Полученная сумма означает, что отток средств со счетов клиентов не превысит рассчитанную величину, и приравнивается к требуемому (необходимому) резерву ликвидности.

В целях контроля достаточности активов банка для компенсации возможного оттока с пассивных счетов на сроке до 1 месяца банком ежедневно проводится расчет фактического (имеющегося) объема (резерва) ликвидных средств (V). В него включаются активы, которые наиболее легко реализуются (востребуются) и могут послужить источником для обеспечения обязательств банка перед клиентами. Это остатки со сроками до востребования и до 30-ти дней на счетах актива : касса; средства на корреспондентском счете и в фонде обязательных резервов сверх суммы фиксированной части резервных требований; средства на корреспондентских счетах в других банках; кредиты юридических и физических лиц, вложения в ценные бумаги; прочие активы.

Далее величины двух резервов (требуемого и фактического) сопоставляются по абсолютной величине и, в случае недостаточности фактического объема активов ($V < Vmp$), риск-менеджмент инициирует принятие оперативных управленческих решений по наращиванию фактической подушки ликвидности (Vф). Например такие, как: приобретение государственных краткосрочных обязательств (наращивание активов), краткосрочное кредитование, размещение средств в депозиты других банков и др.

II. Выводы

По итогам проведенного исследования следует отметить, что:

- управление риском ликвидности, основанное на анализе действующей структуры активов и пассивов, должно сопровождаться принятием мер по регулированию объемно-временных показателей ресурсов банка;
- дополнение статичной модели GAP-анализа элементами модели динамической, построенной на основании исторических данных с использованием элементов теории вероятностей, повышает адекватность (точность) оценки риска ликвидности;
- практическое использование GAP-модели в ежедневной деятельности банка при управлении ликвидностью и риском ликвидности реализуется посредством контроля показателей, характеризующих фактическое и спрогнозированное по историческим данным состояние разрыва на коротком сроке.

Список литературы

1. Инструкция об организации внутренней процедуры оценки достаточности капитала. Постановление Правления Национального банка Республики Беларусь от 22.01.2018 г. № 25 [Электронный ресурс]. – Режим доступа: https://www.nbrb.by/legislation/documents/PP_25_571.pdf. – Дата доступа: 06.10.2025.
2. Инструкция об организации корпоративного управления [Электронный ресурс]. – Режим доступа: Постановление Правления Национального банка Республики Беларусь от 21.02.2024 г. № 62. Источник: https://www.nbrb.by/legislation/documents/pp_-62-2024.pdf. – Дата доступа: 06.10.2025.

ИСПОЛЬЗОВАНИЕ ВОЗМОЖНОСТЕЙ СОВРЕМЕННЫХ МЕССЕНДЖЕРОВ В ПРОГРАММНЫХ ПРОДУКТАХ НА ПРИМЕРЕ TELEGRAM MINI APPS

Нестеренков С. Н., Асепков Д. И.
Кафедра программного обеспечения информационных технологий,
Белорусский государственный университет информатики и радиоэлектроники
Минск, Республика Беларусь
E-mail: s.nesterenkov@bsuir.by, danilamail05@gmail.com

Рассматривается эволюция мессенджеров от простых инструментов для общения до многофункциональных платформ. Особое внимание уделяется технологии Telegram Mini Apps (TMA) как яркому примеру этой тенденции. Анализируются ключевые преимущества, архитектура и практическое применение TMA для современного бизнеса, а также их стратегическое значение в сравнении с традиционными мобильными приложениями.

Введение

Современные мессенджеры эволюционировали от инструментов для общения до многофункциональных платформ. Эта трансформация началась с появления API и чат-ботов, однако их текстовый интерфейс был неудобен для сложных задач. Для понимания значимости текущих изменений необходимо проанализировать модель WeChat, которая стала успешным доказательством концепции экосистемы «приложение в приложении». Успех WeChat продемонстрировал, что мессенджер может стать центральной операционной системой для цифровой жизни пользователя, интегрируя социальные функции, электронную коммерцию, платежи и сторонние сервисы (Mini Programs) в единый, бесшовный интерфейс.

Ключевыми факторами успеха WeChat стали глубокая интеграция собственной платежной системы (WeChat Pay), захваченный внутренний рынок и преимущество первопроходца, что создало мощные сетевые эффекты. Это задает эталон, с которым можно сравнивать глобальную и более децентрализованную стратегию Telegram.

При этом важно понимать, что Telegram не просто копирует WeChat, а адаптирует модель суперприложения для глобальной, ориентированной на конфиденциальность аудитории. Модель WeChat тесно связана с регуляторной средой Китая, требуя верификации по реальному имени и работая под строгим государственным надзором. Такую модель крайне сложно экспортировать на западные рынки с их строгими законами о защите данных. Бренд Telegram, напротив, построен на конфиденциальности, шифровании и анонимности пользователей. Выстраивая свою экосистему на этих принципах, Telegram стратегически позиционирует себя как «дружелюбную к Западу» платформу-суперприложение.

I. Ключевые аспекты применения

1. *Экономическая эффективность и скорость разработки.* Ключевое преимущество TMA – значительное снижение затрат и времени на разработку. В отличие от нативных приложений, требующих отдельных версий для iOS и Android, TMA строятся на стандартных веб-технологиях и используют единую кодовую базу. Разработчики полностью освобождаются от комиссий магазинов приложений (15–30%) и могут интегрировать любые платежные системы. Этот финансовый выигрыш дополняется отсутствием длительного процесса модерации в App Store и Google Play. Обновления развертываются на сервере мгновенно и сразу становятся доступны всем пользователям, что значительно ускоряет циклы итераций и вывода новых функций на рынок.
2. *Бесшовный пользовательский опыт и высокая конверсия.* TMA устраняют главные барьеры для пользователя: поиск в магазине, скачивание, установку и регистрацию. Приложение запускается мгновенно из чата, а авторизация происходит в один клик через виджет Telegram. Весь путь клиента, от знакомства с продуктом до покупки, проходит в едином окне мессенджера. Такой «бесшовный» процесс минимизирует отвлекающие факторы и снижает вероятность того, что пользователь покинет воронку продаж, что напрямую ведет к росту конверсии. Кроме того, TMA предлагает настоящую кроссплатформенную универсальность: одно веб-приложение без дополнительной адаптации работает на всех устройствах, поддерживающих Telegram – iOS, Android и десктопных версиях.
3. *Прямая коммуникация и удержание клиентов.* Каждое Mini App неразрывно связано с Telegram-ботом. После первого взаимодействия бот остается в списке чатов пользователя, создавая постоянный и бесплатный канал для прямой коммуникации. Это позволяет бизнесу эффективно возвращать клиентов, отправляя персонализированные уведомления и специальные предложения без

затрат на SMS или email-маркетинг. При первом запуске бизнес получает Telegram ID пользователя, что является ценным источником собственных данных (first-party data) для дальнейшего маркетинга.

II. Архитектура и технологический стек

В своей основе TMA – это веб-приложение, отображаемое во встроенном в клиент Telegram компоненте WebView. Такая архитектура делает технологию доступной для огромного сообщества веб-разработчиков. Она состоит из трех ключевых элементов :

- *Клиентская часть (Frontend):* веб-приложение, созданное с помощью JavaScript или фреймворков (React, Vue).
- *Серверная часть (Backend):* обрабатывает бизнес-логику и данные на любом серверном языке.
- *Telegram Bot API:* выступает мостом между приложением и мессенджером, обеспечивая аутентификацию и обмен данными.

Для глубокой интеграции используется специальный SDK (telegram-web-app.js), который позволяет управлять нативными элементами интерфейса (кнопками, всплывающими окнами), адаптировать внешний вид под цветовую схему пользователя, вызывать тактильную обратную связь и запрашивать доступ к геолокации или камере.

III. Практическое применение

Технология TMA нашла широкое применение в различных сферах:

- *Электронная коммерция:* создание полноценных интернет-магазинов и сервисов доставки.
- *Игры и геймификация:* особую популярность приобрели игры в формате «tap-to-earn» (нажимай, чтобы заработать), такие как Blum, привлекающие миллионы пользователей за счет вирусного маркетинга.
- *Сервисы и утилиты:* платформы для онлайн-записи, системы бронирования, образовательные курсы и менеджеры задач.
- *Web3 и DeFi:* TMA становятся удобным шлюзом в мир децентрализованных финансов, предлагая встроенные криптокошельки и NFT-маркетплейсы, часто на базе блокчейна The Open Network (TON).
- *CRM/ERP системы:* Инструменты для малого бизнеса на основе TMA, позволяющие менеджерам по продажам вести учет клиентов, отслеживать статусы заказов или управлять задачами прямо из рабочего чата.

IV. Распространение и безопастность

Несмотря на преимущества, экосистема TMA имеет свои особенности. Основной проблемой является обнаруживаемость: в отличие от централизованных App Store и Google Play, у Telegram нет официального каталога, что затрудняет продвижение новых приложений без существующей аудитории. Продвижение зависит от ссылок в каналах, ботах и сторонних каталогах. Кроме того, безопасность полностью лежит на плечах разработчиков. Поскольку TMA – это веб-приложения, они уязвимы для стандартных веб-угроз, таких как XSS-атаки или утечки данных. Крайне важно применять лучшие практики безопасной разработки, включая шифрование данных и надежную аутентификацию.

Заключение

Telegram Mini Apps представляют собой фундаментальный сдвиг в подходе к разработке и дистрибуции цифровых продуктов. Они демократизируют доступ к «экономике приложений», предлагая бизнесу более дешевую и быструю альтернативу нативным приложениям. Устраняя барьеры для пользователей и создавая прямой канал для взаимодействия, TMA открывают новые горизонты для повышения конверсии и удержания клиентов. С ростом рынка «суперприложений», который, по прогнозам, к 2030 году достигнет 426 миллиардов долларов , роль мессенджеров как центральных цифровых платформ будет только усиливаться.

V. Списки литературы

1. Telegram Mini Apps: что это, как работает и зачем нужно вашему бизнесу [Электронный ресурс] // Блог SberCRM. – Режим доступа: `https://sbercrm.com/blog/telegram-mini-apps`. – Дата доступа: 18.10.2025.
2. Telegram Web App: полная инструкция по созданию мини-приложения в Телеграм [Электронный ресурс] // Timeweb.cloud. – Режим доступа: `https://timeweb.cloud/tutorials/react/telegram-web-app-kak-sozdat-mini-prilozhenie-v-telegram`. – Дата доступа: 18.10.2025.
3. Telegram vs WeChat: The Ultimate Comparison of Messaging Giants and Their Mini-App Ecosystems [Electronic resource] // Adsgram. – Mode of access: `https://adsgram.ai/telegram-vs-wechat-the-ultimate-comparison-of-messaging-giants-and-their-mini-app-ecosystems/`. – Date of access: 18.10.2025.
4. Security Concerns When Developing Telegram Mini-Games [Electronic resource] // Appicial. – Mode of access: `https://www.appicial.com/blog/security-concerns-when-developing-telegram-mini-games.html`. – Date of access: 18.10.2025.
5. Telegram Mini Apps [Electronic resource] // Telegram Core. – Mode of access: `https://core.telegram.org/bots/webapps`. – Date of access: 18.10.2025.
6. Как создать Telegram Web App: инструкция по разработке Mini App [Электронный ресурс] // Cloud.ru. – Режим доступа: `https://cloud.ru/blog/kak-sozdat-telegram-web-app`. – Дата доступа: 18.10.2025.
7. Гриз, Е. А. Блокчейн и эффективность цифровых платежных систем / Е. А. Гриз, С. Н. Нестеренков, Д. В. Кишкевич // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. X Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 13 марта 2024 года). В 2 ч. Ч. 1 / редкол. : В. А. Богуш [и др.]. – Минск : БГУИР, 2024. – С. 427-432.

УСТОЙЧИВОЕ СЕЛЬСКОЕ ХОЗЯЙСТВО: ПЛАН ДЕЙСТВИЙ ДЛЯ ОЦЕНКИ УРОВНЯ СЕЛЬСКИХ ХОЗЯЙСТВ БЕЛАРУСИ НА ОСНОВЕ КЛЮЧЕВЫХ ПОКАЗАТЕЛЕЙ ЭФФЕКТИВНОСТИ (KPI)

Русецкий А. Д.
Кафедра экономической информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: ruseckiy.alexander@gmail.com

Методологии ключевых показателей эффективности (KPI) становятся важным инструментом оценки устойчивости сельского хозяйства. Расширение цифровизации и доступность данных создают условия для внедрения систем мониторинга на уровне отдельных хозяйств и перехода от нормативных к результативным механизмам аграрной политики. Цель статьи – предложить план разработки и внедрения системы KPI для сельских хозяйств Беларуси. Объект исследования – процесс оценки устойчивости сельскохозяйственных предприятий. Предмет исследования – система ключевых показателей эффективности (KPI) как инструмент для перехода к политике, основанной на производительности. В статье обсуждаются институциональные и этические аспекты внедрения системы, а также её потенциал для повышения экологической и экономической результативности сельскохозяйственной политики.

Введение

Глобальная политика интенсификации сельского хозяйства, направленная на борьбу с голодом, привела к значительным экологическим последствиям: деградации почв, снижению биоразнообразия и существенному вкладу в антропогенные выбросы. Эта деградация угрожает здоровью людей и долгосрочному производству продовольствия.

В качестве решения предлагается создание научно обоснованных нормативных рамок, стимулирующих инновации. Исследование фокусируется на разработке системы оценки устойчивости сельхозпредприятий в Беларуси на основе ключевых показателей эффективности (KPI), поскольку системная оценка в стране отсутствует. Предлагаются восемь принципов построения такой системы и 5-этапная процессная рамка для её внедрения, интегрирующая эти принципы.

I. Использование KPI для реализации основанных на производительности сельскохозяйственных политик

Анализ агроэкологической политики выявляет преимущества подходов, ориентированных на результат, над традиционными практико-ориентированными мерами. Однако доминирование последних сохраняется из-за сложности создания систем верификации на основе экологического мониторинга.

В сельском хозяйстве KPI представляют собой расчетные величины, агрегирующие воздействие на окружающую среду (например, выбросы CO2-эквивалента), которые получаются с помощью моделей на основе данных хозяйства. Их преимущество — использование уже собранной агростатистики.

Ключевым ограничением для внедрения KPI остается недостаток открытых данных [1]. Решением может стать объединение данных из различных источников (ресурсы, урожайность, лабораторные и геопространственные данные), что требует разработки четких требований и схемы управления.

Для создания системы оценки устойчивости сельского хозяйства на основе KPI также необходимо обосновать целесообразность использования таких показателей для аграрной политики, ориентированной на результат [2]. Затем, на основе синтеза исследований существующих инструментов оценки, формулируются принципы построения системы KPI и описывается пошаговый процесс ее внедрения.

Следующий раздел представляет план создания такой системы, включающий принципы и процессную рамку для её развертывания.

II. Развертывание и внедрение системы KPI в сельском хозяйстве

Многообразие трактовок устойчивости и оценочных инструментов обусловило необходимость разработки принципов для создания интегрированной системы KPI в Беларуси. Эти принципы учитывают преимущества политик, ориентированных на результат.

На основе подходов к управлению устойчивостью, включая цикл PDCA, была разработана унифицированная 5-этапная структура, синтезирующая общие шаги различных методик оценки. Данная структура призвана помочь в создании и применении системы KPI для мониторинга и стимулирования устойчивости на уровне сельхозпредприятий (см. рис. 1).

Рис. 1 – Структура системы KPI для сельского хозяйства Беларуси

Пятиэтапная структура внедрения системы KPI включает:

1. Планирование: определение многоаспектной концепции устойчивости и формулировка конкретных целей на национальном уровне.
2. Проектирование связей: установление взаимосвязей между целями, данными хозяйств и геопространственной информацией для обеспечения релевантности KPI и формирования основы для системы поддержки принятия решений.
3. Разработка KPI: формирование расчетных показателей и их целевых значений с привлечением научного сообщества. Показатели адаптируются для различных секторов и регионов с использованием пространственно-дифференцированных методов.
4. Реализация: внедрение теоретических наработок в виде удобной онлайн-платформы для фермеров, обеспечивающей наглядную визуализацию прогресса и упрощающей процедуру получения субсидий.
5. Адаптация: регулярный анализ результатов и корректировка показателей, методов сбора данных или целей на основе обратной связи. Управление процессом осуществляется рабочей комиссией с участием заинтересованных сторон для актуализации системы.

Если анализ показывает расхождение между улучшением KPI и состоянием экосистем, целевые значения и расчетные модели стоит соответствующим скорректировать [3]. Этот цикл обратной связи помогает учитывать неопределенности научных моделей (см. рис. 2).

Рис. 2 – Цикличность работы и развития системы KPI

III. Необходимость адаптации политической и законодательной структуры

Успешное внедрение системы KPI в Беларуси, открывающее возможности для перехода к политике, основанной на производительности, зависит не только от технологических решений, но и от комплексной политической поддержки [4]. Необходима четкая структура управления, распределяющая роли между участниками, и разработка стандартов совместимости данных для интеграции информации из различных источников.

Критически важными являются сбалантированные финансовые стимулы, учитывающие этические аспекты связи субсидий с KPI, а также создание механизмов верификации данных для обеспечения достоверности системы. Целесообразно поэтапное внедрение через пилотные программы для тестирования и доработки методики в реальных условиях.

Заключение

Внедрение системы сопряжено с вызовами, включая потенциальное сопротивление существующих структур, необходимость обеспечения продовольственной безопасности и предстоящие структурные изменения в управлении. Тем не менее, переход к оценке устойчивости на основе KPI обладает значительным потенциалом для развития аграрного сектора. Перспективным направлением дальнейших исследований представляется разработка показателей на основе обмена данными между предприятиями различных регионов.

IV. Список литературы

1. Кравченко, Н. А. Сбалансированная система показателей как инструмент управления устойчивым развитием предприятия / Н. А. Кравченко. – СПб.: Питер, 2020. – 192 с.
2. Никонов, В. А. Устойчивое развитие сельских территорий: теория, методология, практика / В. А. Никонов, И. В. Савченко. – М.: РУСАЙНС, 2021. – 218 с.
3. Семенов, В. А. Методы оценки экологической эффективности сельскохозяйственных предприятий / В. А. Семенов // Вестник аграрной науки. – 2022. – № 5(86). – С. 112–119.
4. Гуринович, Л. В. Государственное регулирование и поддержка сельского хозяйства в Республике Беларусь / Л. В. Гуринович. – Минск: БГАТУ, 2021. – 168 с.

ГЕЙМИФИКАЦИЯ КАК ИНСТРУМЕНТ УПРАВЛЕНИЯ ПОЖИЗНЕННОЙ ЦЕННОСТЬЮ КЛИЕНТА

Саблина А. И.
Кафедра экономической информатики,
Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: oiveshka@gmail.com

В работе рассматривается геймификация как инструмент управления лояльностью и повышения пожизненной ценности клиента (Customer Lifetime Value, CLV). Анализируются основные механизмы её влияния – накопительные, достиженческие и случайные игровые элементы. Показано, что геймификация повышает вовлечённость, частоту взаимодействий и удержание клиентов, способствуя росту CLV и эмоциональной привязанности к бренду. Приводятся эмпирические данные о влиянии игровых механик на бизнес-показатели и обсуждаются подходы к расчёту и атрибуции CLV. Отмечаются риски пресыщения и манипулятивных практик, подчёркивается необходимость ответственного проектирования и персонализированного подхода на основе аналитики и искусственного интеллекта.

Введение

Современные компании сталкиваются с падением эффективности традиционных инструментов привлечения и удержания клиентов. Конкуренция и информационная перегрузка снижают отклик на классические программы лояльности и скидки, что требует поиска новых форм взаимодействия с аудиторией.

Одним из таких инструментов становится геймификация – использование игровых элементов (баллов, уровней, наград и рейтингов) в бизнес-процессах для повышения вовлечённости и лояльности клиентов [1]. Она перестала быть временным трендом и превратилась в стратегический механизм формирования потребительских привычек и увеличения пожизненной ценности клиента (CLV).

Особенно активно геймификация развивается в банковской сфере, где становится частью стандартов клиентского сервиса и укрепления долгосрочных отношений [2]. Актуальность темы заключается в необходимости определить, как именно игровые механики повышают вовлечённость и CLV, а также какие факторы ограничивают их эффективность.

I. Механизмы формирования лояльности

Геймификация использует игровые механики для стимулирования нужного поведения клиентов. Основные типы включают накопительные механики (баллы, бонусы, наклейки за покупки), которые превращают каждую транзакцию в шаг к цели и мотивируют возвращатьс; механики достижений (уровни, значки, рейтинги), удовлетворяющие потребность в признании; и случайные награды (лотереи, колесо фортуны), поддерживающие интерес эффектом неожиданности [1]. На практике компании часто комбинируют эти элементы, создавая гибридные программы вовлечения.

Психологическая сила геймификации объясняется активацией базовых мотивов человека – стремления к автономии, компетентности и прогрессу [2]. Игровые элементы позволяют пользователю видеть результат, получать признание и выбирать собственные цели, превращая рутинные действия в увлекательный процесс. Вознаграждения и «ощущение достижения» вызывают положительные эмоции и укрепляют привязанность к бренду, постепенно формируя долгосрочную лояльность.

II. Эмпирическая оценка эффективности геймификации

Исследования подтверждают, что геймификация существенно повышает активность и удержание клиентов. В эксперименте в одном из банков Латинской Америки отключение игрового модуля привело к снижению числа входов в приложение на 20%, платежных операций – на 18%, а доли своевременных выплат по кредитам – на 31% [2]. Особенно заметно падение проявилось у новых пользователей.

Обратный эффект наблюдается при активном применении геймификации: вовлечённость клиентов в среднем возрастает на 30%, а удержание – до 15% в первые месяцы после внедрения [2]. Таким образом время сессии в приложении выросло в 13 раз, доля еженедельных пользователей достигла 33%, а удержание депозитных клиентов увеличилось на 41% [2].

Похожие результаты демонстрирует программа «Плюс Сити» от Яндекса: за первый месяц игру установили 7 млн пользователей, ежедневно проводящих в ней до 30 минут, что повысило использование экосистемных сервисов и продлило подписки [3]. Эти данные подтверждают, что геймификация напрямую способствует росту пожизненной ценности клиента (CLV) за счёт увеличения частоты взаимодействий, удержания и лояльности.

III. Методы оценки и атрибуции пожизненной ценности клиента

Для оценки эффективности геймификации центральным показателем выступает *Customer Lifetime Value* (CLV) – пожизненная ценность клиента, показывающая, сколько прибыли компания получает за весь период взаимодействия. Один из распространённых способов её расчёта описывается формулой [4]:

$$CLV = AOV \times F \times GM \times \left(\frac{1}{Churn}\right).$$

где AOV – средний чек, F – частота покупок, GM – валовая маржа, $Churn$ – коэффициент оттока клиентов.

Из формулы видно, как геймификация влияет на рост CLV. Игровые механики повышают частоту покупок (F) и снижают отток ($Churn$) за счёт вовлечённости и удовлетворённости клиентов. Возможен рост среднего чека (AOV) через стимулы к крупным заказам и маржи (GM) – при продвижении более прибыльных продуктов. Дополнительно игровые элементы усиливают реферальный эффект, когда пользователи приглашают друзей, принося компании новых клиентов без затрат.

Для достоверной оценки важно корректно атрибутировать эффект. Простое сравнение метрик до и после внедрения геймификации может искажать результат. Более надёжны A/B-тесты, когортный и survival-анализ, позволяющие выявить чистый инкрементальный прирост CLV [5]. При этом рекомендуется фокусироваться не на игровых показателях (уровнях, значках), а на финансовых метриках – марже, частоте покупок, удержании. Интеграция CLV в модели атрибуции и машинного обучения помогает точнее оценить вклад геймификации в долгосрочную прибыльность.

IV. Ограничения и этика

Несмотря на преимущества, геймификация имеет и ограничения. Один из ключевых рисков – эффект пресыщения: при избытке игровых стимулов пользователи устают, и участие перестаёт приносить удовольствие [1]. Если программа требует постоянной активности без возможности перерыва, мотивация быстро падает. Устойчивый интерес поддерживают гибкие механики, например «пауза без вины», позволяющая сделать перерыв без потери прогресса [5]. Такой подход предотвращает выгорание и повышает доверие к бренду.

Этические аспекты также важны: игровые механики воздействуют на поведение, и компания несёт ответственность за пределы этого влияния. Недопустимы манипуляции, побуждающие к избыточным тратам; правила должны быть прозрачными, а вознаграждения – справедливыми. Кроме того, механики стоит адаптировать под разные типы пользователей, чтобы избежать перегрузки или демотивации.

Наконец, геймификация не заменяет качество самого продукта: она эффективна лишь там, где базовая услуга действительно ценна для клиента. Иначе эффект будет кратковременным – пользователь проявит интерес из любопытства, но не останется без ощутимой пользы.

Выводы

Геймификация зарекомендовала себя как эффективный инструмент управления лояльностью и повышения пожизненной ценности клиента (CLV). Она усиливает эмоциональную связь с брендом, повышает удержание, частоту покупок и средний чек, вовлекая пользователей через стремление к прогрессу, признанию и вознаграждению. Эмпирические данные показывают: внедрение игровых механик в разных отраслях приводит к росту повторных продаж и снижению оттока клиентов.

При этом геймификация требует продуманного проектирования и привязки к метрикам CLV. Эффективность следует измерять аналитически, оценивая как краткосрочные (вовлечённость), так и долгосрочные результаты (удержание, LTV). Успешные программы не перегружают пользователя и дают реальную ценность, сохраняя этический баланс: мотивируют, а не манипулируют.

С развитием аналитики и искусственного интеллекта геймификация станет более персонализированной, что позволит точнее управлять мотивацией клиентов и делать рост их ценности устойчивым.

1. Сухостав, Е. В. Повышение вовлеченности потребителей через элементы геймификации / Е. В. Сухостав // Экономика, предпринимательство и право. – 2025. – Т. 15, № 11. – DOI: 10.18334/epp.15.11.124130. – Режим доступа: `https://1economic.ru/lib/124130`. – Дата доступа: 19.10.2025.
2. Орлова, С. Клиент хочет играть [Электронный ресурс] / С. Орлова // Банковское обозрение. – 2025. – 5 мая. – Режим доступа: `https://bosfera.ru/bo/klient-hochet-igrat`. – Дата доступа: 20.10.2025.
3. Федоров, Е. Как Яндекс вовлекает миллионы пользователей в «Плюс Сити»? [Электронный ресурс] / Е. Федоров // VC.ru. – 2024. – 30 октября. – Режим доступа: `https://vc.ru/marketing/1624412-kak-yandeks-vovlekaet-milliony-polzovatelei-v-plyus-siti`. – Дата доступа: 20.10.2025.
4. Что такое Lifetime Value (LTV, CLTV, CLV) или пожизненная ценность клиента? [Электронный ресурс] // CXDP.ru. – 2024. – Обновлено 03.04.2024. – Режим доступа: `https://cxdp.ru/blog/wiki/ltv/`. – Дата доступа: 21.10.2025.
5. Мирошников, Д. Геймификация для лояльности: как увеличить продажи через игровые элементы [Электронный ресурс] / Д. Мирошников // Блог «Секреты бизнеса». – 2025. – 22 сентября. – Режим доступа: `https://secrets.tbank.ru/blogi-kompanij/geimifikaciya-dlya-loyalnosti/`. – Дата доступа: 19.10.2025.

ПРОГРАММНОЕ СРЕДСТВО УЧЁТА ДЕНЕЖНЫХ СРЕДСТВ ФИЗИЧЕСКИХ ЛИЦ ДЛЯ ПОВЫШЕНИЯ ФИНАНСОВОЙ ГРАМОТНОСТИ

Смоляков В. А.
Кафедра менеджмента, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: wladrimir@mail.ru

В работе анализируется проблема отсутствия системного подхода к учёту личных финансов среди населения Беларуси. Описаны результаты создания специального веб-приложения, предназначенного для автоматизации процессов отслеживания поступлений и расходов финансов. Подчёркивается значимость внедрения цифровых технологий учёта как инструмента формирования устойчивых финансовых привычек у граждан. Перспективное развитие проекта связывается с элементами машинного обучения и использованием банковских API.

Введение

Цифровые финансы представляют собой эволюцию традиционных механизмов финансового обслуживания посредством внедрения информационно-коммуникационных технологий. Это направление содержит не только электронные платежи, но и программные решения для личного финансового менеджмента, способствующие планированию и автоматизации денежных потоков. Дальнейшее развитие данной сферы является ключевым аспектом цифровой экономики, направленной на доступность финансовых инструментов для населения и повышения эффективности [1].

Необходимость исследования обусловлена существующим дисбалансом между расходами и доходами очень большой части граждан Республики Беларусь. Согласно результатам опросов, около 29 процентов жителей страны не ведут учёт собственных денежных средств, что повышает стрессовую нагрузку и затрудняет долгосрочное планирование. По данным опросов за 2025 год, именно среди тех, кто не контролирует свои расходы, чаще наблюдается превышение расходов над доходами. Порядка 13 процентов респондентов сталкивались с невозможностью покрытия затрат из-за отсутствия финансового контроля. Таким образом, внедрение программных систем для отслеживания доходов и расходов является актуальным инструментом решения имеющейся проблемы [2].

Цель исследования: установить причины и последствия дисбаланса расходов и доходов у населения Беларуси.

Задачи исследования: анализ поведения граждан в сфере контроля денежных средств; проектирование и реализация веб-приложения, обеспечивающего цифровизацию личного финансового менеджмента.

Основная часть

Полученные данные свидетельствуют о низком уровне финансовой дисциплины среди граждан. Постоянные методы учёта применяют не более 13 процентов респондентов. У безработных и пожилых людей наблюдается наиболее выраженная недостаточность финансовой организации, соответственно свыше 49 процентов и около 30 процентов не осуществляют учёт собственных денежных средств [3].

Исходя из этого, в Беларуси актуальна цифровая трансформация личных финансов, предполагающая разработку программного обеспечения повышения финансовой грамотности и автоматизации управления денежными потоками. Выделяются три основные формы ведения бюджета: специализированные программные средства, обеспечивающие комплексную автоматизацию финансового контроля; электронные таблицы (Google Sheets, Microsoft Excel), предоставляющие базовые способы анализа; традиционный бумажный контроль, отличающийся высокой трудоёмкостью. Объект исследования заключается в процессе создания веб-приложения для персонального финансового управления, предмет исследования является методология проектирования таких систем. Для корректной постановки исследовательской задачи необходимо предварительно рассмотреть предметную область, в рамках которой осуществляется работа. Это требует понимания сущности бухгалтерии и принципов функционирования системы бухгалтерского учёта [4].

Бухгалтерия представляет собой упорядоченную систему сбора, регистрации и обработки информации о финансовом состоянии организации, её имуществе и обязательствах. Учёт ведётся в денежном выражении и отражает все хозяйственные операции на основе принципов полноты, непрерывности и документальности. Объектами бухгалтерского учёта являются имущество организации, её обязательства, а также совершаемые в процессе деятельности хозяйственные действия.

Ключевая цель бухгалтерского учёта заключается в формировании достоверных и полных данных, используемых при составлении бухгалтерской отчётности и оценке имущественного со-

стояния экономического субъекта. В контексте данного исследования рассматривается аналогичная задача формирования информации о личных финансовых потоках, личный или семейный бюджет.

Система автоматизации бухгалтерского учёта предназначена для ведения различных видов учёта: как бухгалтерского, так и фискального, ориентированного на соблюдение требований государства в части расчёта и уплаты налогов.

В зависимости от объёма и сложности учётных задач система подразделяется на несколько категорий: персональные решения, системы для малого и среднего бизнеса, комплексы для крупных компаний и специализированные инструменты для банковских учреждений. В начале нашего века во всём мире наблюдается рост интереса к управлению личными финансами, что обусловлено усложнением структуры индивидуальных финансовых потоков, возникновением кредитных отношений, увеличением разнообразия доходов и расходов, а также развитием возможностей частных инвестиций. Для этих целей всё чаще используют индивидуальные бухгалтерские программы.

Эффективное финансовое планирование требует правильного понимания расходов, их основных и дополнительных категорий. Практика показывает, что большинство пользователей средств учёта не знакомы с теоретическими основами бухгалтерии, поэтому подход к ведению личной бухгалтерии должен быть интуитивно понятен и доступен широкой аудитории, не обладающей специальными знаниями. Именно на этом принципе строится концепция предлагаемого веб-приложения.

Реализуемое программное средство предусматривает создание системы, включающей такие функциональные возможности, как регистрация и авторизация пользователей, внесение и корректировка записей о расходах и доходах, а также удаление неактуальных данных.

Архитектура интерфейса пользователя организована по модульному принципу. Каждая часть системы отвечает за конкретный аспект взаимодействия с данными и представлена следующими структурными элементами: аналитический блок, включающий инструменты построения отчётности и экспорта данных в различные форматы; инструмент планирования бюджета, реализующий механизм установки и контроля лимитов; подсистема классификации и категоризации финансовых транзакций; модуль учёта расходов и доходов с возможностью внесения, корректировки и удаления операций; основная панель управления с графическим отображением ключевых финансовых параметров; система регистрации, аутентификации и авторизации пользователей.

В рамках работы реализована архитектура клиент-серверного типа на базе СУБД MySQL и фреймворка Laravel. Клиентская часть построена с использованием JavaScript, CSS, HTML и Bootstrap, обеспечивая адаптивный интерфейс. Программный функционал включает аутентификацию, категоризацию операций, инструменты бюджетирования и визуализацию данных в виде графиков и диаграмм. Интерфейс соответствует современным тенденциям снижения когнитивной нагрузки пользователя. Разработанное приложение выполняет не только функции фиксации операций, но и позволяет пользователю анализировать своё финансовое поведение, выявлять закономерности и формировать оптимальные стратегии планирования. Визуализированные данные способствуют более глубокому пониманию структуры расходов и доходов, повышая финансовую осведомлённость. Таким образом, предложенное решение служит инструментом организации личных финансов и осознанного управления ими [5].

Заключение

Итоги исследования подтверждают актуальность проблемы недостаточной финансовой дисциплины населения и демонстрируют эффективность её решения посредством цифровизации личных финансов. Созданное приложение является практическим средством перехода от хаотичных записей к системному учёту. Работа вносит вклад в развитие сферы цифровых финансов, предлагая прикладное решение, способствующее росту экономической грамотности.

Для разработчика срок окупаемости проекта один год эксплуатации. Пользователь не несёт капитальных затрат, оплачивает лишь подписку. Экономический эффект проявляется с первого месяца использования – при стоимости подписки 35 BYN в год средняя годовая экономия составляет до 1054 BYN.

Дальнейшее развитие проекта предполагает интеграцию с банковскими API для автоматического импорта транзакций, поддержку мультивалютного учета, применение алгоритмов машинного обучения для прогнозирования расходов, а также создание мобильной версии и адаптацию системы для образовательных целей в вузах и колледжах.

1. Козлова, А. В. Цифровизация личных финансов: современные программные решения для управления бюджетом / А. В. Козлова // Экономика и банки. – 2024. – Т. 30, № 2. – С. 115–123.
2. Ображей, О. Н. Отчёт о выполнении научно-исследовательской работы по теме: «Оценка и анализ финансовой грамотности населения Республики Беларусь» / О. Н. Ображей. – Минск: Издательский дом "Белорусская наука 2024. – 340 с.
3. Сидоренко, Д. И. Финансовая грамотность населения Беларуси в условиях цифровой трансформации / Д. И. Сидоренко // Финансы и кредит. – 2023. – № 8(896). – С. 45–53.
4. Анализ рынка приложений для учета финансов в России в 2020-2024 гг, прогноз на 2025-2029 гг / BusinesStat. – Москва, 2024. – 50 с.
5. Василенко, М. Г. Использование API банков в веб-приложениях финансового учёта / М. Г. Василенко // Информатика и цифровая экономика. – 2022. – № 4. – С. 79–87.

Содержание

Предисловие **7**

Секция «Системы управления» **7**

Борис Д. А., Гармашук В. Р.
Радиоуправляемый катер для обеспечения безопасности на воде **9**

Захарьев В. А.
Модель компактного представления харакреистик голоса диктора **11**

Киселевский О. С.
Динамическое моделирование организационных систем **13**

Климашевич З. Д., Клепацкий Е. Э., Линкевич А. О., Болабсов Р. А.
Многофункциональная робототехническая платформа с адаптивными режимами управления **15**

Крищенович В. А., Захарьев В. А.
Робастная верификация диктора в зашумлённых условиях на основе x-vectors **17**

Нестеренков С. Н., Шмаргун А. Э.
Повышение эффективности командной работы в распределённых проектах за счёт интеграции адаптивной Kanban-доски с механизмами реального времени и гибкой настройки рабочих процессов **19**

Русакович Е. С., Деменковец Д. В.
Элемент системы мониторинга промышленных производств **21**

Сержанов А. К.
Система поддержки принятия решений для управления процессом риформинга сырьевой нафты **23**

Свито И. Л., Тихоновецкий С. А.
Различные подходы к расчету сердечников трансформаторов **25**

Стешенко П. П., Иванюкович А. В., Свито И. Л., Журавлёв В. И.
Система управления станцией катодной защиты **27**

Хаджинов М. К., Павлова А. В., Стасевич Н. А.
Сервопривод со встроенным в наблюдатель комбинированным управлением **29**

Секция «Интеллектуальные информационные технологии» 31

Акмырадова А. Я.
Теоретические основы технологий дополненной и виртуальной реальности с использованием искусственного интеллекта в инженерии **33**

Арефин В. А.
Архитектура Retrieval-Augmented Generation как основа построения локальных ИИ-систем с корпоративными базами знаний **35**

Батура М. П., Кулевич А. О.
Система искусственного интеллекта анализа данных тематических сайтов с использованием RAG технологий, LLM и графа знаний **37**

Боброва Н. Л., Богомаз К. С.
Современные методы решения задач логистики с помощью систем поддержки принятия решений **39**

Вакарь Е. И., Коростелёв А. Д., Романюк М. В.
Методы оценки доверия и прозрачности решений ИИ-систем **41**

Герман О. В., Герман Ю. О.
Физические парадоксы и сверхмалые величины **43**

Гончаров С. В., Войтешенко И. С.
К вопросу назначения водителей на рейсы городского общественного транспорта на основе обучения с подкреплением **45**

Гракова Н. В.
Применение технологий искусственного интеллекта для решения задач управления в образовательной сфере **47**

Дичковский В. А., Ковалёв М. В.
Оценка производительности программ для локального запуска больших языковых моделей **49**

Дубинина Е. О., Захаров В. В.
Разработка интеллектуального плагина с локальной моделью генерации кода для языка SC-код **51**

Завалей В. А., Орлов Д. В., Скиба И. Г.
Оптимизация RAG-архитектур для предметно-ориентированных справочных систем вузов **53**

Зотов Н. В.
Модель управления агентами в общей семантической памяти интеллектуальных систем . . **55**

Зотов Н. В.
Средства разработки агентов в программной платформе интеллектуальных систем **57**

Иванова М. Д.
Анализ методов распознавания профессий в русскоязычных текстах **59**

Ивашенко В. П.
Метрики смысловых подпространств **61**

Ивашенко В. П.
Применение автокодировщиков и метода адаптируемого коэффициента обучения для понижения размерности в смысловом пространстве **63**

Каминский А. В., Петровский Н. А.
Сравнение подходов к построению хромаграммы для анализа аудио **65**

Клещукевич А. А., Владимирский П. Н.
Интеллектуальный асситент для персонализации профессиональной ориентации учащихся **67**

Ковалев М. В., Кривенков А. В., Мазаник Е. Н., Терешкова А. С., Шнейдеров Е. Н., Шункевич Д. В.
Применение технологий искусственного интеллекта при разработке и внедрении электронных образовательных ресурсов **69**

Колистратов М. В., Шинкарев И. А.
Разработка мобильного приложения для обнаружения повреждений на автомобилях **71**

Кретов А. А., Сащеко Е. А.
Интеллектуальная система персонализации обучения учащихся на примере геометрии . . **73**

Крез К. С., Ефименко Д. Д., Войнилович Н. Ю.
Разработка модуля для предобработки документов обнаружения заимствований в тексте . **75**

Крупенич Е. Г., Вергель И. В.
Обзор и классификация угроз информационным системам со стороны квантовых алгоритмов **77**

Крупенич Е. Г., Лукьянов А. А.
Анализ подходов к построению криптографических схем, устойчивых к квантовым атакам **79**

Курочка К. С., Овчинина О. В.
Использование алгоритмов машинного обучения для генерации нарративов в интерактивных 2D-приложениях **81**

Курочка К. С., Зайцева Л. П., Гуменников Е. Д.
Методика предварительной обработки данных и подготовки датасета для анализа препаратов жидкостной цитологии с использованием машинного обучения **83**

Кузнецов К. С.
Применение методов машинного обучения для калибровки МЭМС-акселерометров в микроконтроллерных инерциальных системах **85**

Лазаревич И. М., Мосин Н. А., Савицкий Р. М.
ИИ как инструмент обоснования стратегических и тактических решений в системе ЖКХ Республики Беларусь **87**

Нестеренков С. Н., Лебедевич А. В.
Реализация рекомендательной системы в маркетплейсе на основе эмбеддингов **89**

Логвинович М. В., Коростелёв А. Д., Марков А. Н.
Применение машинного обучения в управлении складскими запасами **93**

Малащук М. Ю.
Интеллектуальные алгоритмы для автоматизированной оценки кандидатов по фотоизображению **95**

Мальцев В. Л., Возмитель В. В., Матяс Е. В.
Обнаружение программ-вымогателей на основе комбинированной нейронной сети и анализа энтропии данных **97**

Мамченко К. А., Ситников А. В., Романюк М. В.
Построение интеллектуальных чат-ботов на основе моделей глубокого обучения **99**

Мигалевич С. А., Скалозуб К. А., Тимошек Е. С.
Применение технологий машинного обучения для реализации естественно-языкового взаимодействия с интеллектуальными системами **101**

Нестеренков С. Н., Фурсанов С. А., Камышев С. В.
Классификация и обзор методов параметро-эффективной адаптации больших языковых моделей (LLM) **103**

Нестеренков С. Н., Осадчий О. Э.
Автоматизация корректировки текста при помощи LLM **105**

Нестерович Г. В., Фурсанов С. А.
Сравнительный анализ методов дообучения LLM в условиях ограниченности для данных специализированных областей. **107**

Орлов Д. В., Коростелёв А. Д., Марков А. Н.
Применение искусственного интеллекта в рамках борьбы с дезинформацией: современные методы и решения **109**

Раптунович Д. И.
Сравнительный анализ математических моделей прогнозирования цен на рынке недвижимости Республики Беларусь **111**

Сальников Д. А., Ерофеев А. С.
Автоматизация работы с базами знаний на основе гибридного подхода **113**

Сальников Д. А., Крищенович В. А.
Аспекты защиты персональных данных в интеллектуальных медицинских системах **115**

Скалозуб К. А., Ярмош А. Д., Романюк М. В.
Алгоритмы интеллектуального поиска информации в распределенных базах данных с применением методов машинного обучения **117**

Таранчук В. Б.
Сопоставление программных реализаций, алгоритмы и инструменты настройки нейронных сетей **119**

Тимошенко В. А., Бегляк Е. В., Мигалевич С. А.
Прогнозирование рисков с использованием искусственного интеллекта **121**

Ходжиметов Э. Ш., Захарьев В. А.
Архитектуры глубоких нейронных сетей для анализа медицинских изображений **123**

Чернявский К. Э., Марцинкевич В. А.
Нейросетевые модели семантического поиска в предметно-ориентированных коллекциях текстов **125**

Шоть К. С.
Применение технологий персонализации в мобильных приложениях на основе анализа пользовательских данных с использованием методов машинного обучения **127**

Шушкевич Д. В.
Анализ векторов атак на системы социального восстановления ключей в системах суверенной идентификации (SSI) и роль пороговой криптографии в их митигации **129**

Adzinets D., Sychev Y., Alooeff E.
Analysis of the Architecture Design Approach Based on Requirements **131**

Dmitry Pakutnik
LLM-Based Intelligent System for Quality Control of Student Assignments **133**

Dmitry Pakutnik
Zero-Shot Automatic Assignment Grading with Large Language Models **135**

Xia Enduo, Fan Linda, He Hongyan
Network Traffic Analysis Based on Deep Learning Algorithms **137**

Секция «Проектирование встраиваемых систем» **139**

Босько Н. В., Луцик Ю. А.
Исследование времени выполнения частичной реконфигурации FPGA Xilinx **141**

Бурко Л. А., Иванюк А. А.
Генерация случайных чисел на основе схемы ФНФ конфигурируемого кольцевого осциллятора **143**

Кайкы М. Н., Иванюк А. А.
Временная модель RS-защёлки, реализованной как источник случайности на ПЛИС типа FPGA **145**

Кривальцевич Е. А., Вашкевич М. И.
Исследование эффектов квантования параметров нейронной сети на основе обучаемого разделимого преобразования **147**

Осипов А. С., Петровский Н. А.
IP-компонент конвейерного ускорителя свёрточных нейронных сетей для СнК серии Zynq-7000 **149**

Петровский Д. А., Иванюк А. А.
Исследование характеристик метода постобработки случайных данных на базе многовходового сигнатурного анализатора **151**

Субботенко О. Р., Вашкевич М. И.
FPGA реализация двухслойной нейронной сети прямого распространения для распознавания изображений **153**

Трубач К. И., Иванюк А. А.
Сравнение аппаратных и программных средств расчёта основных характеристик физически неклонируемых функций **155**

Черемисинова Л. Д.
Графовые модели задач, возникающих при декомпиляции КМОП СБИС **157**

Секция «Математическое моделирование и компьютерная графика» **159**

Белаш В. А.
Нейросетевые подходы к восстановлению трехмерной позы человека по двумерным данным **161**

Белобокова Ю. А., Колистратов М. В.
Маркировка растровых изображений для защиты от фальсификации **163**

Винник М. А., Кривец К. А., Ловшенко И. Ю., Новиков П. Э., Корсак К. В., Степанов А. А.
Математическое моделирование влияния параметров материалов металлизации на эксплуатационные характеристики элементов цифровых интегральных схем **165**

Журавлёв В. И., Наумович Н. М., Стежко И. К., Колбун В. С.
Тепловое моделирование микросхемы усилителя мощности X диапазона **167**

Журавлёв В. И., Свито И. Л., Петровский И. И., Шилин Л. Ю.
Моделирование силового преобразователя
с учётом температурных зависимостей **169**

Кукин Д. П., Купчина Е. В., Панасик А. А.
Разработка алгоритма восстановления компьютерных изображений в условиях недостаточной освещенности AIR-Net **171**

Логинова И. П.
Представление данных трассы параллельных процессов в виде таблиц для работы с базой данных SQLite **173**

Панасик А. А., Кукин Д. П., Купчина Е. В.
Обзор современных методов сегментации изображений **175**

Семенченя Т. С., Курочка К. С.
Математическая модель напряженно-деформированного состояния поясничного отдела позвоночника человека на основе анализа КТ-изображений **177**

Свинарский М. В.
Математическое моделирование траектории полета одноступенчатой твердотопливной баллистической ракеты малой дальности **179**

Цегельник В. В.
Аналитические свойства решений двух динамических систем с хаотическим поведением . . **181**

Шилин Л. Ю., Хаджинова К. А.
Моделирование синхронизированных генераторов Лоренца **183**

Pilipchuk L. A.
Decomposition of Basic Graphs in Fractional Linear Programming Problems with Linear Constraints **185**

Секция «Автоматизированные системы обработки информации» **187**

Белоусова Е. С., Вербило Н. А., Филиппов А. Н.
Сценарии реагирования на кибератаки в образовательном киберполигоне CyberLab **189**

Бранцевич П. Ю.
Автоматизация принятия решений в системах вибрационного контроля **191**

Быстрова В. А., Боброва Н. Л.
Анализ технологий и методов для автоматизированного выявления заболеваний растений **193**

Буцкевич Е. М., Кравченко Е. Д., Белоусова Е. С.
Методологии обеспечения катастрофоустойчивости в виртуализированных инфраструктурах: сравнительный анализ подходов **195**

Дубровский А. В.
Совершенствование алгоритмической и аппаратной составляющих компьютерного анализа фракционного состава методом мини-дистилляции **197**

Жуков А. М., Навроцкий А. А.
Использование Big Data для оптимизации ИТ-решений: комплексный анализ технологий и методов **199**

Зайцев А. С., Богуш Р. П.
Автоматизация создания курсов на основе учебных планов в системе Moodle **201**

Крагель В. А., Ярмош А. Д., Скиба И. Г.
Защита трафика распределенных систем на основе протокола Mutual TLS **203**

Краснопрошин Д. В., Вашкевич М. И.
Рекуррентная нейронная сеть на основе LSTM-блоков с механизмом внимания для распознавания эмоций в речи **205**

Кузнецов Е. В., Приходько Н. Н., Чехомов Е. Г., Косарева Е. М.
Алгоритм анализа средней длины слова
в словаре В.И. Даля **207**

Логинова И. П.
Организация ввода-вывода при выполнении параллельных вычислений в системе FLC2 . . **209**

Лукашевич М. М., Венгеренко В. В., Воронов А. А.
Прототипирование подготовки данных из видеоряда и сегментации для генерации 3D модели головы человека по 2D изображениям **211**

Матвеев Н. С., Орлов Д. В., Скиба И. Г.
Аналог шифрсхемы Эль-Гамаля на основе эллиптических кривых **213**

Мельник А. А.
Сравнение СУБД для хранения структур древовидных данных **215**

Нестеренков С. Н., Иванов А. А.
Информационная система для сбора, анализа и визуализации данных о наблюдениях за безопасностью на морских объектах нефтегазовой индустрии **217**

Нестеренков С. Н., Лазук И. С.
Анализ текстового содержимого главной страницы веб-сайта для определения его тематики при помощи LLM **219**

Новыш И. М., Герман Ю. О.
Анализ методов решения задачи размещения камер наблюдения в помещениях с перегородками **221**

Петров С. Н., Пакуль Е. С., Корчинский А. А.
Криптографический функционал мессенджера Peregrine на базе национальных стандартов шифрования Республики Беларусь **223**

Почебут А. С.
Принципы построения и методы балансировки нагрузки в высокопроизводительных сетях центров обработки данных **225**

Почебут А. С.
Методы оптимизации маршрутизации в высоконагруженных сетях в центрах обработки данных **227**

Прудникова И.В.
Интеллектуальная поддержка принятия управленческих решений в сфере кадрового делопроизводства **229**

Семижон Е. А., Герман Ю. О., Герман О. В.
Сравнительный анализ современных методов решения логико-комбинаторных задач **231**

Серкевич Д. С., Герман Ю. О.
Анализ стеганографических методов **233**

Славинский Г. А., Боброва Н. Л.
Сравнительный анализ моделей данных: реляционная, графовая и сетевая модели **235**

Струкова А. А., Шилин Л. Ю.
Сравнение алгоритмов взаимодействия компонентов в существующих фреймворках и разработка двухсвязной модели взаимодействия на стороне клиента **237**

Сьянов Д. А.
Анализ квантильного и гауссовского подходов в моделировании редакционных процессов . **239**

Хаджинова Н. В., Агель А. А., Пашковец М. В.
Экспериментальная оценка метода Adversarial Training на задаче бинарной классификации **241**

Хаджинова Н. В., Кадлубай В. А.
Когнитивно-доверительная модель гибридных систем «человек–агент» **243**

Хаджинова Н. В., Михнюк А. И., Засим А. О.
Минимизация задержек в распределенных мультиагентных системах с использованием граничных вычислений **245**

Хаджинова Н. В., Служалик В. Ю., Хаджинова К. А.
Алгоритм локальной реоптимизации маршрутов на динамических графах **247**

Хаджинова Н. В., Сманцер Н. С., Соркин В. О.
Иммунный подход к обнаружению компрометации в мультиагентных системах **249**

Хатеневич А. А., Перцев Д. Ю.
Алгоритмы сегментации при обработке спутниковых снимков **251**

Хлебанова К. О.
Безопасные облачные хранилища биомедицинских данных **253**

Чаган Н. Ф., Белоусова Е. С.
Угроза ботнетов для IoT-устройств **255**

Черемисинова Л. Д.
Решение сложных вычислительных задач на базе кластерного компьютера **257**

Wang Jing, Yu.German
Statistical Analysis of Pixel Distribution for Image Distortion Detection **259**

Mukha V. S.
I-optimal design for Affine Multiple Regression **261**

Orazdurdyyeva G. O., Bekiyeva M. B., Bekiyev A. R.
Statistical Evaluation of Network Traffic Variations Using Hypothesis Testing Methods **263**

Tang Yi, Yu.O.German
An Improved Q-Learning Algorithm with Optimized Initialization and Annealed Boltzmann Exploration **265**

Секция «Технологии цифровой экономики» **267**

Бекиева М. Б., Ораздурдыева Г. О., Бекиев А. Р., Дурдыев С. М.
Применение блокчейн-технологий для обеспечения прозрачности цепочек поставок **269**

Верняховская В. В., Усова В. А.
Использование искусственного интеллекта в digital-продвижении **271**

Дайнович О. Л., Ситников А. В., Мигалевич С. А.
Анализ потребительского поведения на основе методов цифрового маркетинга и машинного обучения **273**

Галяк И. П.
Программное средство анализа новостного сентимента на криптовалютном рынке **275**

Илюкевич А. А.
Гибридная адаптивная модель управления ИТ-компанией в условиях роста **277**

Клименко И. В.
Объяснимый искусственный интеллект и метод SHAP как основа прозрачного прогнозирования в агропромышленном комплексе Республики Беларусь **279**

Кропотин Д. Д.
Влияние искусственного интеллекта на цифровой маркетинг **281**

Марченкова Е. Е.
Визуальные коммуникации в рекламе: культурные особенности и эффективность восприятия **283**

Нагулевич Р. С.
Совершенствование методов управления ликвидностью банка **285**

Нестеренков С. Н., Асепков Д. И.
Использование возможностей современных мессенджеров в программных продуктах на примере Telegram Mini Apps **287**

Русецкий А. Д.
Устойчивое сельское хозяйство: план действий для оценки уровня сельских хозяйств Беларуси на основе ключевых показателей эффективности (KPI) **289**

Саблина А. И.
Геймификация как инструмент управления пожизненной ценностью клиента **291**

Смоляков В. А.
Программное средство учёта денежных средств физических лиц для повышения финансовой грамотности **293**

Научное издание

Информационные технологии и системы (ИТС 2025)

Материалы Международной научной конференции
(Республика Беларусь, Минск, 19 ноября 2025 года)

Information Technologies and Systems (ITS 2025)

Proceeding of the International Conference
(Republic of Belarus, Minsk, 19th November 2025)

В авторской редакции
Ответственный за выпуск *М. В. Ковалёв*
Компьютерная верстка *Н. В. Гракова*
Дизайн обложки *А. В. Жмырко*

Подписано в печать 25.11.2025. Формат 60×84 ⅛. Бумага офсетная. Гарнитура «BookAntiqua».
Отпечатано на ризографе. Усл. печ. л. 36,04. Уч.-изд. л. 36,0. Тираж 45 экз. Заказ 197.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2024,
№2/113 от 07.04.2024, №3/615 от 07.04.2014.
Ул. П. Бровки, 6, 220013, г. Минск