



XXII



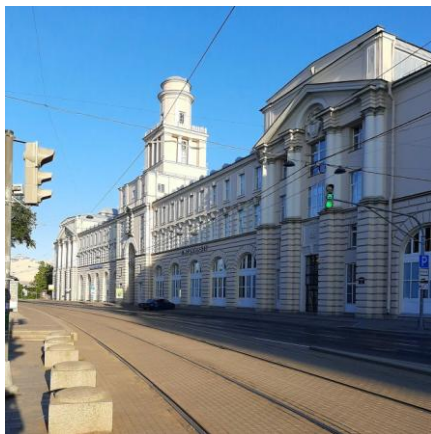
ИТМО

**НАЦИОНАЛЬНАЯ
КОНФЕРЕНЦИЯ
ПО ИСКУССТВЕННОМУ
ИНТЕЛЛЕКТУ
С МЕЖДУНАРОДНЫМ
УЧАСТИЕМ**



**СПб
ФИЦ
РАН**

КИИ-2025



**ТРУДЫ КОНФЕРЕНЦИИ
Том 3**

6-10 октября 2025 г.
Санкт-Петербург

XXII

Российская ассоциация
искусственного интеллекта

Федеральный
исследовательский центр
«Информатика и
управление» РАН

Национальный
исследовательский
университет ИТМО

Санкт-Петербургский
Федеральный
исследовательский
центр РАН

НАЦИОНАЛЬНАЯ КОНФЕРЕНЦИЯ ПО ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ С МЕЖДУНАРОДНЫМ УЧАСТИЕМ

КИИ-2025

ТРУДЫ КОНФЕРЕНЦИИ
Том 3

6-10 октября 2025 г.
Санкт-Петербург,
Национальный исследовательский
университет ИТМО

Санкт-Петербург
СПб ФИЦ РАН
2025

УДК 004.8+004.89+004.82+004.032.26(045)+004.9

ББК 32.813

Д 22

Организаторы конференции:

Российская ассоциация искусственного интеллекта

Федеральный исследовательский центр «Информатика и управление» РАН

Национальный исследовательский университет ИТМО

Санкт-Петербургский Федеральный исследовательский центр РАН

Д 22 Двадцать вторая Национальная конференция по искусственному интеллекту с международным участием, КИИ-2025 (Санкт-Петербург, 6-10 октября 2025 г.). Труды конференции. В 3-х томах. Т. 3. – СПб: Изд-во СПб ФИЦ РАН, 2025. – 461 с.

ISBN 978-5-6052274-4-1

Двадцать вторая Национальная конференция по искусственному интеллекту с международным участием КИИ-2025 продолжает традицию советских (российских) конференций, организуемых Российской ассоциацией искусственного интеллекта.

В третьем томе трудов публикуются доклады участников конференции, представленные на следующих секциях:

Секция 8 «Машинное обучение, нейросетевые методы, нейроинформатика»,

Секция 9 «Инструментальные средства и технологии проектирования интеллектуальных систем»,

Секция 10 «Прикладные интеллектуальные системы».

ББК 32.813

Рецензенты: академик РАН, ИПУ РАН *С.Н. Васильев*,
д.т.н., ФИЦ ИУ РАН *О.Г. Григорьев*

ISBN 978-5-6052274-4-1

© Авторы, 2025

© Российская ассоциация искусственного интеллекта, 2025

© Издательство СПб ФИЦ РАН, 2025



ВАЛЕРИЙ БОРИСОВИЧ ТАРАСОВ

16 февраля 1955 – 22 июля 2021

В этом году конференция посвящена 70-летию со дня рождения Валерия Борисовича Тарасова – выдающегося ученого в области искусственного интеллекта, ведущего отечественного специалиста по направлениям семиотического моделирования, многоагентных систем, нечетких систем, мягких вычислений и когнитивных измерений.

СОПРЕДСЕДАТЕЛИ КОНФЕРЕНЦИИ

Соколов И.А., акад. РАН, ФИЦ ИУ РАН, Москва

Васильев В.Н., член-корр. РАН, НИУ ИТМО, Санкт-Петербург

Ронжин А.Л., д.т.н., проф. РАН, СПб ФИЦ РАН, Санкт-Петербург

ПРОГРАММНЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

Сопредседатели Программного комитета

Кобринский Б.А., д.м.н., проф., ФИЦ ИУ РАН, Москва

Котенко И.В., д.т.н., проф., СПб ФИЦ РАН, Санкт-Петербург

Заместители председателя Программного комитета

Грибова В.В., член-корр. РАН, ИАПУ ДВО РАН, Владивосток

Забейжайло М.И., д.ф.-м.н., проф., ФИЦ ИУ РАН, Москва

Ответственный секретарь Программного комитета

Подвесовский А.Г., к.т.н., доц., БГТУ, Брянск

Члены программного комитета

Аверкин А.Н., к.ф.-м.н., доц., ФИЦ ИУ РАН, Москва

Афанасьева Т.В., д.т.н., РЭУ им. Г.В. Плеханова, Москва

Бобцов А.А., д.т.н., проф., НИУ ИТМО, Санкт-Петербург

Болодурина И.П., д.т.н., проф., Оренбургский ГУ, Оренбург

Боргест Н.М., к.т.н., доц., Самарский НИУ им. акад. С.П. Королева, Самара

Борисов В.В., д.т.н., проф., филиал НИУ МЭИ, Смоленск

Бухановский А.В., д.т.н., проф., НИУ ИТМО, Санкт-Петербург

Васильев С.Н., акад. РАН, ИПУ РАН, Москва

Визильтер Ю.В., д.ф.-м.н., проф., ГосНИИАС, Москва

Виноградов Д.В., д.ф.-м.н., ФИЦ ИУ РАН, Москва

Гаврилова Т.А., д.т.н., проф., СПбГУ, Санкт-Петербург

Гладков Л.А., к.т.н., доц., ИКТИБ ЮФУ, Таганрог

Городецкий В.И., д.т.н., проф., АО «Эврика», Санкт-Петербург

Еремеев А.П., д.т.н., проф., НИУ «МЭИ», Москва

Желтов С.Ю., акад. РАН, ГосНИИАС, Москва

Загорулько Ю.А., к.т.н., ИСИ СО РАН, Новосибирск

Ильин В.А., д.ф.-м.н., проф., НИУ ИТМО, Санкт-Петербург

Калюжная А.В., к.т.н., доц., НИУ ИТМО, Санкт-Петербург

Карпов А.А., д.т.н., проф., СПб ФИЦ РАН, Санкт-Петербург

Ковалев С.М., д.т.н., проф., РГУПС, Ростов-на-Дону

Колесников А.В., д.т.н., проф., БФУ, Калининград

Колоденкова А.Е., д.т.н., доц., Самарский НИУ, Самара

Кузнецов О.П., д.т.н., проф., ИПУ РАН, Москва

Кузнецов С.О., д.ф.-м.н., проф., НИУ ВШЭ, Москва

Лебедев О.Б., д.т.н., доц., ВАГШ ВС РФ, Москва

Лукашевич Н.В., д.т.н., проф., МГУ им. М.В. Ломоносова
Макаров Д.А., к.т.н., ФИЦ ИУ РАН, Москва
Мисник А.Е., к.т.н., доц. БРУ, Республика Беларусь, Могилев
Михеенкова М.А., д.т.н., проф., ФИЦ ИУ РАН, Москва
Мошкин В.С., к.т.н., доц., УлГТУ, Ульяновск
Насонов Д.А., к.т.н., доц., проф., НИУ ИТМО, Санкт-Петербург
Пальчунов Д.Е., д.ф.-м.н., доц., ИМ СО РАН, Новосибирск
Палюх Б.В., д.т.н., проф., ТвГТУ, Тверь
Панов А.И., д.ф.-м.н., доц., МФТИ, Москва
Редько В.Г., д.ф.-м.н., проф., НИИСИ РАН, Москва
Ройзензон Г.В., к.т.н., доц., ФИЦ ИУ РАН, Москва
Рыбина Г.В., д.т.н., проф., НИЯУ МИФИ, Москва
Смирнов И.В., д.ф.-м.н., доц., ФИЦ ИУ РАН, Москва
Стефанюк В.Л., д.т.н., проф., ИППИ РАН, Москва
Сулейманов Д.Ш., акад. АН РТ, ИПС АН РТ, Казань
Тельнов Ю.Ф., д.э.н., проф., РЭУ, Москва
Уткин Л.В., д.т.н., проф., СПбПУ, Санкт-Петербург
Финн В.К., д.т.н., проф., ФИЦ ИУ РАН, Москва
Хачумов В.М., д.т.н., проф., ФИЦ ИУ РАН, Москва
Шалфеева Е.А., д.т.н., ИАПУ ДВО РАН, Владивосток
Яковлев К.С., к.ф.-м.н., ФИЦ ИУ РАН, Москва

ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ

Сопредседатели организационного комитета

Борисов В.В., д.т.н., проф., филиал НИУ МЭИ, Смоленск
Заколдаев Д.А., к.т.н., доц., НИУ ИТМО, Санкт-Петербург

Члены организационного комитета

Благосклонов Н.А., ФИЦ ИУ РАН, Москва
Волошина Н.В., к.т.н., доц., НИУ ИТМО, Санкт-Петербург
Воробьева А.А., к.т.н., доц., НИУ ИТМО, Санкт-Петербург
Давыдов В.В., к.т.н., НИУ ИТМО, Санкт-Петербург
Десницкий В.А., к.т.н., доц., СПб ФИЦ РАН, Санкт-Петербург
Левигун Д.С., к.т.н., СПб ФИЦ РАН, Санкт-Петербург
Кириллова Е.А., д.э.н., доц., филиал НИУ «МЭИ», Смоленск
Попов И.Ю., к.т.н., доц., НИУ ИТМО, Санкт-Петербург
Синяевский Ю.В., к.т.н., филиал НИУ «МЭИ», Смоленск
Солопов Р.В., к.т.н., доц., филиал НИУ «МЭИ», Смоленск
Чечулин А.А., к.т.н., доц., СПб ФИЦ РАН, Санкт-Петербург

ИНФОРМАЦИОННАЯ ПОДДЕРЖКА КОНФЕРЕНЦИИ

ООО «Лаборатория информационных технологий», Смоленск

ПРЕДИСЛОВИЕ

Двадцать вторая Национальная конференция по искусственному интеллекту с международным участием КИИ-2025 продолжает традицию советских (российских) конференций, организуемых Российской ассоциацией искусственного интеллекта (РАИИ).

Конференция посвящена памяти Валерия Борисовича Тарасова, 70-летие со дня рождения которого отмечается в этом году – выдающегося ученого в области искусственного интеллекта, который внес существенный вклад в развитие методов семиотического моделирования, в теорию агентов и многоагентных систем, теорию нечетких систем, мягких вычислений и когнитивных измерений. Валерий Борисович Тарасов был участником Учредительного съезда Советской ассоциации искусственного интеллекта (ныне РАИИ). С 1992 по 2000 год являлся вице-президентом РАИИ, с 2000 года – членом Научного совета РАИИ. Он был одним из основателей Российской ассоциации нечетких систем (ныне Ассоциации нечетких систем и мягких вычислений).

Федеральный проект «Искусственный интеллект» направлен на решение междисциплинарных проблем, сформулированных в Национальной стратегии развития искусственного интеллекта, включая исследования как в области фундаментальной и прикладной науки, так и в образовательной сфере. Эти проблемы постоянно находятся в фокусе внимания членов РАИИ. Их отражение можно видеть в докладах конференции, организаторами которой являются Российская ассоциация искусственного интеллекта, Федеральный исследовательский центр «Информатика и управление» РАН, Национальный исследовательский университет ИТМО, Санкт-Петербургский Федеральный исследовательский центр РАН.

Тематика конференции охватывает следующие основные направления искусственного интеллекта: инженерия знаний; интеллектуальный анализ данных; моделирование рассуждений; интеллектуальный анализ текстов, большие языковые модели; нечеткие модели, мягкие измерения и вычисления, биоинспирированные методы; интеллектуальные агенты, роботы, компьютерное зрение; интеллектуальное управление и поддержка принятия решений; машинное обучение, нейросетевые методы, нейроинформатика; инструментальные средства и технологии проектирования интеллектуальных систем; прикладные интеллектуальные системы.

Представлены пленарные доклады видных ученых и специалистов в области искусственного интеллекта. Из поданных на конференцию 138 секционных докладов Программным комитетом после рецензирования отобраны 102 доклада.

В конференции принимают участие более 280 ученых, исследователей и специалистов из академических институтов, университетов, отраслевых институтов и организаций Москвы, Санкт-Петербурга, Белгорода, Брянска, Владивостока, Донецка, Иркутска, Казани, Калининграда, Красноярского края, Махачкалы, Московской области, Новосибирска, Пермского края, Ростова-на-Дону, Самары, Смоленска, Таганрога, Твери, Ульяновска, Уфы, Ханты-Мансийска, Челябинска, Ярославля, Могилёва (Республика Беларусь), Суйчжоу, (Китай), Ташкента (Узбекистан), Хайдарабада (Индия).

Программный комитет КИИ-2025

УДК 629.735:004.942

doi: 10.15622/rcai.2025.064

**ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ
ПРИ КОНЦЕПТУАЛЬНОМ ПРОЕКТИРОВАНИИ
САМОЛЕТА**

Э.О. Адигвэ (*elpromzeegm@gmail.com*)

Н.М. Боргест (*mail@borgest.ru*)

Самарский национальный исследовательский университет
им. С.П. Королёва, Самара

Рассматривается применение искусственного интеллекта (ИИ) в области проектирования летательных аппаратов, а в частности, использование машинного обучения (МО) для создания концепции самолета на основе данных прежних проектов. В исследовании использована аналогия эволюции живой материи и артефакта, в частности на примере формирования начального концепта самолетов гражданского и военного назначения. Проведен анализ исследований в области автоматизированного проектирования и производства самолета с целью выявления тенденции в построении их концепт-проектов. Инструменты МО и ИИ рассматриваются в качестве помощника при проектировании летательных аппаратов, которые могут позволить ускорить процесс принятия проектных решений.

Ключевые слова: искусственный интеллект, машинное обучение, летательный аппарат, эволюция, концепт-проект.

Введение

Ускоренная, за счет накопленных знаний о природе вещей, эволюция артефактов во многом схожа с относительно медленной эволюцией живой материи. Менее чем за полтора столетия авиационная отрасль подобно метаморфозе бабочки преобразовалась из достаточно простых на сегодняшний взгляд первых прототипов летательных аппаратов до современных воздушных судов. Основным принципом эволюции является принцип естественного отбора, предполагающий, что благоприятные черты помогают выжить и размножиться [Darwin, 1859].

В статье рассматривается развитие артефактов, в частности, пассажирских и военных самолетов – как результат прогресса транспортных средств, движимый насущной человеческой потребностью и теми добытыми знаниями, которые позволили ускорить эволюцию артефактов. В табл. 1 и 2 приведены фрагменты баз данных (БД) самолетов, которые использовались в работе [Засыпкин и др., 2011]. Общее количество реализованных проектов артефактов значительно меньше, чем это было и есть в арсенале природы. В данном исследовании оно составило не многим более 100 самолетов, созданных за 100 лет.

Таблица 1

Характеристики пассажирских самолетов (фрагмент базы данных)

Год и модель	Тип двигателя	Размах крыла, м	Макс. взлет. вес, кг	Скорость, км/ч	Материал	Кол-во пассажиров
1903 Флайер -1	Поршневой	12	340	14	Дерево	1
1929 К-4	Поршневой	16.8	2420	145	Метал, Дерево, полотно	3
1934 ХАИ-1	Поршневой	14.9	2600	258	Фанера	6
1946 Ил-12	Поршневой	31.7	17500	344	Метал	32
1971 Ту-144	Турбореактивный	28	195000	2200	Метал	150
2016 МС-21	Турбовентиляторный	35.9	85000	870	Метал, композит	211

Таблица 2

Характеристики истребителей (фрагмент базы данных)

Год и модель	Тип двигателя	Размах крыла, м	Макс. взлет. вес, кг	Скорость, км/ч	Материал
1914 С-16	Поршневой	8.8	676	125	Дерево и полотно
1926 И-1	Поршневой	10.8	1510	264	Дерево с фанерой
1936 И-14	Поршневой	11.3	1540	449	Метал
1940 Як-1	Поршневой	10	2884	592	Метал
1966 МиГ-25	Турбореактивный	14	34920	3000	Сталь, титан, дюра-люминиевый сплав
2008 МиГ-29К	Турбореактивный	12	17770	2300	Метал

Используемые в таблицах атрибуты и их значения косвенно характеризуют причину существенного изменения внешнего вида воздушных судов. Эволюция воздушных судов вызвана развитием технологий, а также внедрением новых идей и концепций в стремлении достичь совершенства.

1. Краткий обзор работ

ИИ находит применение в различных сферах, в том числе в решении инженерных задач, включая проектирование летательных аппаратов. Исследование использования больших языковых моделей (БЯМ) в автоматизации проектирования показало их перспективы [Pan et al, 2025].

ИИ может выступать как вспомогательный элемент в коллективных средах систем автоматизированного проектирования (САПР) [Hasby et al, 2024], которые фокусируются на детализации, но слабо поддерживают концептуальное проектирование. МО эффективно применяется для прогнозирования технического обслуживания, контроля качества и планирования производства [Manta-Costa et al, 2024]. Есть основание полагать, что применение МО может помочь в когнитивном, совместном, концептуальном проектировании [Goel et al, 2012]. Совмещение экспертных знаний и МО, в так называемый «гибридный интеллект», помогает провести анализ отказов в индустрии 4.0 [Mokhtarzadeh et al, 2024]. Автоматизированный контроль сборки на основе генерированных данных обеспечивает точность и экономию времени [Zhu et al, 2024], а генерация сценариев с помощью ИИ используется для стратегического планирования [Ferrer et al, 2025].

ИИ способствует повышению автономности и эффективности производственных систем, снижению затрат, что ускоряет переход к Индустрии 4.0 и далее к Индустрии 5.0 [Lin et al, 2024], [Gallab et al, 2024]. Различные подходы к проектированию, такие как технологический, социально-культурный и социально-экономический эволюционируют [Figoli et al, 2022], [Brand et al, 2011], при этом по-прежнему отмечается важная роль человека в проектировании [Rampino, 2018]. ИИ на разных этапах проектирования помогает с генерацией идей, расширением творческого спектра и устранением стереотипов путем предложения новых методов решения проблем [Figoli et al, 2022]. Практический результат перехода от текстового описания к физическому объекту был получен с использованием глубокого обучения (ГО) [Dosovitskiy et al, 2021]. ИИ способен сгенерировать желаемый прототип сложных объектов на основе текстовых описаний изображения с использованием модели генеративных состязательных сетей [Ramesh et al, 2021].

В авиационном производстве БЯМ продемонстрировали точность в технических задачах процесса проектирования, выборе материалов, поиске информации об инструментах, и при этом выявлены риски дезинформации [Liu et al, 2025].

Таким образом, ИИ и МО могут эффективно интегрироваться в работу инженеров-конструкторов, способствуя созданию инновационных и эффективных решений.

2. Искусственный интеллект и машинное обучение в проектировании летательных аппаратов

Анализ данных – это процесс сбора, изучения, и оценки информации с целью выявления закономерностей и ускорения принятия решений [Kinza, 2024]. Эффективность алгоритмов МО для выполнения специализированных задач возрастает с увеличением объема входных данных [Gerschütz et al, 2023]. МО, как и анализ, основанный на статистике, способно извлечь тренды из данных [L’Heureux et al, 2017]. ГО способно решить более сложные задачи, обрабатывая большое количество данных [Rafiei et al, 2023].

На основе оцифрованной БД изображений пассажирских самолетов, произведенных за последние сто лет, и с помощью нейросети GPT-4, сгенерированы прототипы концептов будущих моделей самолета. Для генерирования прототипов анализируются данные с целью извлечения признаков конструкции, далее анализируются эволюционные тренды и генерируются текстовые описания концептов, а затем создаются визуальные прототипы будущих самолетов (рис. 1 и 2).

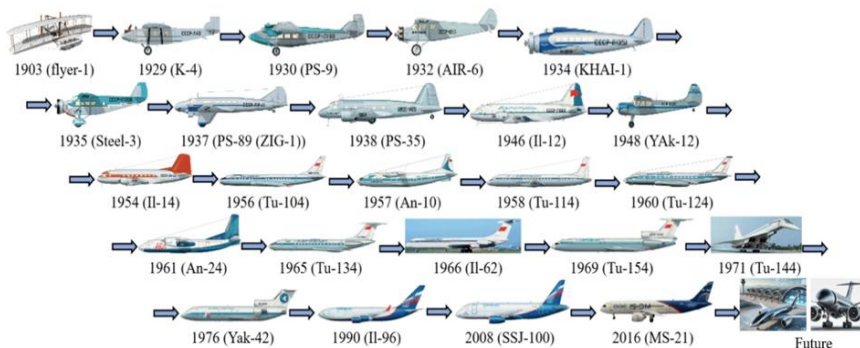


Рис. 1. Ретроспектива пассажирских самолетов и прототипы сгенерированных ИИ концептов

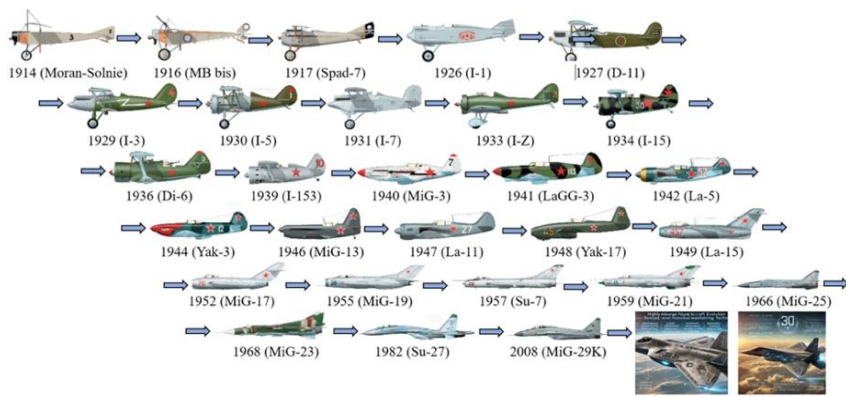


Рис. 2. Ретроспектива истребителей и прототипы сгенерированных ИИ концептов

Для сравнительного анализа, помимо изображений самолетов, было добавлено текстовое описание и информация о летно-технической характеристике каждой модели самолета, включенной в базу (см. фрагменты в табл. 1). На основании обновленной БД сгенерированы прототипы пассажирских самолетов (рис. 3). Эти результаты свидетельствуют о способности МО анализировать и извлекать данные с целью достижения желаемых результатов.

Улучшенные прототипы отличаются новыми инновациями при соблюдении аэродинамических требований. ИИ прогнозирует конструкции из композитных материалов, водородные двигатели и аэродинамическую форму в виде летающего крыла.



Рис. 3. Прототипы сгенерированных ИИ будущих концептов пассажирских самолетов

Аэродинамическая конфигурация и концепция самолета может быть выбрана с помощью так называемого «робота-проектанта» [Боргест и др., 2015], который помогает при обосновании выбора параметров самолета на этапе предварительного проектирования. Робот-проектант также используется в качестве инструмента для комплексного отслеживания процесса проектирования элементов самолета с выполнением расчетов и контролем принимаемых решений на всех этапах проектирования [Боргест и др., 2021].

Рассмотрены концепты интерьера будущего самолета на основе БД предыдущих проектов (рис. 4). Эти концепты представляют собой эргономичные кресла, панорамные и умные окна, интеллектуальные дисплеи и освещение. ИИ предсказывает возможность создания беспилотных пассажирских самолетов.



Рис. 4. Сгенерированные ИИ концепты интерьера салона и кабины пилотов пассажирских самолетов

В БД истребителей, помимо изображений, аналогично добавлено текстовое описание и информация, включающая типы и количество оружия. На основании этой дополненной БД, сгенерированы прототипы истребителей с адаптивными крыльями и покрытием с низкой радиолокационной заметностью. Это улучшает аэродинамику, повышает маневренность и топливную эффективность, снижает вероятность обнаружения радиолокационных систем. В конструкции предусмотрены автономные дроны и лазерное вооружение (рис. 5).



Рис. 5. Прототипы сгенерированных ИИ истребителей

Заключение

Сравнительный анализ изменений природных объектов и артефактов показал схожесть процессов эволюции как совершенствование функционала и адаптации к условиям среды. Использование МО для создания концептов летательных аппаратов на основе оцифрованных данных прошлого опыта позволяет рассмотреть новые идеи проектных решений. Интеграция ИИ и МО в проектирование авиационной техники может существенно изменить темпы развития в данной отрасли. Очевидно, что ИИ может быть использован лишь в качестве помощника с целью поиска новых идей. Результаты, полученные ИИ, требуют дополнительной верификации с помощью физических моделей, учитывающих аэродинамику, массу, прочность и другие параметры.

Список литературы

- [Боргест и др., 2015] Боргест Н.М., Власов С.А., Громов А.А., Коровина М.Д., Шустов Д.В. Робот-конструктор: на пути к реальности // Онтология проектирования. – 2015. – № 4(18). – С. 429-449. – doi: 10.18287/2223-9537-2015-5-4-429-449.
- [Боргест и др., 2021] Боргест Н.М., Григорьев В.А., Кузьмичёв В.С. Искусственный интеллект в проектировании авиационной техники и роль школы профессора В.Г. Маслова в процессе его развития // Вестник Самарского университета. Аэрокосмическая техника, технологии и машиностроение. – 2021. – Т. 20, № 3. – С. 171-190. – DOI: 10.18287/2541-7533-2021-20-3-171-190.
- [Засыпкин и др., 2011] Засыпкин Ю.В., Костырченко Г.В., Кузьмин Ю.В., Остапенко Ю.А., Симонов А.Д., Соболев Д.А. История российской авиационной промышленности. Серийное производство самолетов, 1910-2010 гг. / под общей ред. Д.А. Соболева. – М.: Российское авиационное общество (РУСАВИА), 2011. – 432 с.
- [Brand et al, 2011] Brand R., & Rocchi S. Rethinking value in a changing landscape and business transformation. In Philips Design. – 2011.
- [Darwin, 1859] Charles Darwin, On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life, London, 1859.

- [**Dosovitskiy et al, 2021**] Dosovitskiy A., Springenberg J.T., Tatarchenko M., & Brox T. Learning to Generate Chairs, Tables and Cars with Convolutional Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2017. – 39(4). – P. 692-705.
- [**Ferrer et al, 2025**] Ferrer i Picó J., Catta-Preta M., Trejo Omeñaca A., Vidal M., Monguet i Fierro J.M. The Time Machine: Future Scenario Generation Through Generative AI Tools // Future Internet. – 2025. – 17, 48.
- [**Figoli et al, 2022**] Fabio Antonio Figoli, Francesca Mattioli, Lucia Rampino. Artificial intelligence in the design process // The Impact on Creativity and Team Collaboration. Serie di architettura e design FrancoAngeli. – 2022.
- [**Gallab et al, 2024**] Gallab M., Di Nardo M. & Naciri L. Navigating contemporary challenges and future prospects industry evolution // Discover Applied Sciences. – 2024. – 6, 259. – doi.org/10.1007/s42452-024-05913-2.
- [**Gerschütz et al, 2023**] Gerschütz B., Goetz S., Wartzack S. AI4PD—Towards a Standardized Interconnection of Artificial Intelligence Methods with Product Development Processes // Applied Science. – 2023. – 13. – doi.org/10.3390/app13053002.
- [**Goel et al, 2012**] Goel A.K., Vattam S., Wiltgen B., Helms M. Cognitive, collaborative, conceptual and creative—Four characteristics of the next generation of knowledge-based CAD systems: A study in biologically inspired design //Computer Aided Design. – 2012. – 44. – P. 879-900.
- [**Hasby et al, 2024**] Fariz Muharram Hasby, Dradjad Irianto, Lucia Diawati. Novel Computer-Aided Design-Based Collaboration Framework for the Conceptual–Embodiment Design Phase // Designs. – 2024. – 8, 107.
- [**Howell, 1965**] Francis Clark Howell. Early man. Time-life books. – New York, 1965.
- [**Kinza, 2024**] Kinza Y. Data Analytics (DA). – 2024, [online]. Available: <https://www.techtarget.com/searchdatamanagement/definition/data-analytics>.
- [**Lin et al, 2024**] Lin C.-T. and Lu H.-J. An Intelligent Product-Driven Manufacturing System Using Data Distribution Service // in IEEE Access. – 2024. – Vol. 12. – P. 16447-16461. – doi: 10.1109/ACCESS.2024.3359228.
- [**Liu et al, 2025**] Beiming Liu, Zhizhuo Cui, Siteng Hu, Xiaohua Li, Haifeng Lin, Zhengxin Zhang. LLM Evaluation Based on Aerospace Manufacturing Expertise: Automated Generation and Multi-Model Question Answering // arXiv. – 2025. – doi.org/10.48550/arXiv.2501.17183.
- [**L'Heureux et al, 2017**] L'Heureux A., Grolinger K., Elyamany H.F. and Capretz M.A.M. Machine Learning with Big Data: Challenges and Approaches // in IEEE Access. – 2017. – Vol. 5. – P. 7776-7797. – doi: 10.1109/ACCESS.2017.2696365.
- [**Manta-Costa et al, 2024**] Manta-Costa A., Araújo S.O., Peres R.S. and Barata J. Machine Learning Applications in Manufacturing—Challenges, Trends, and Future Directions // in IEEE Open Journal of the Industrial Electronics Society. – 2024. – Vol. 5. – P. 1085-1103. – doi: 10.1109/OJIES.2024.3431240.
- [**Mokhtarzadeh et al, 2024**] Mokhtarzadeh M., Rodríguez-Echeverría J., Semajski I., Gautama S. Hybrid intelligence failure analysis for industry 4.0: a literature review and future prospective // Journal of intelligent manufacturing. – 2024. – doi: 10.1007/s10845-024-02376-5.

- [**Pan et al, 2025**] Jingyu Pan, Guanglei Zhou, Chen-Chia Chang, Isaac Jacobson, Jiang Hu, and Yiran Chen. A Survey of Research in Large Language Models for Electronic Design Automation. – 2025. – <https://doi.org/10.48550/arXiv.2501.09655>.
- [**Rafiei et al, 2023**] Rafiei M., Raitoharju J. and Iosifidis A. Computer Vision on X-Ray Data in Industrial Production and Security Applications: A Comprehensive Survey // in IEEE Access. – 2023. – Vol. 11. – P. 2445-2477. – doi: 10.1109/ACCESS.2023.3234187.
- [**Ramesh et al, 2021**] Ramesh A., Pavlov M., Goh G., & Gray S. DALL·E: Creating Images from Text. – <https://openai.com/blog/dall-e/>. 2021.
- [**Rampino, 2018**] Rampino L. Evolving Perspectives in Product Design: From Mass Production to Social Awareness. – FrancoAngeli, 2018.
- [**Zhu et al, 2024**] Zhu X., Mårtensson P., Hanson L. et al. Automated assembly quality inspection by deep learning with 2D and 3D synthetic CAD data // Journal of intelligent manufacturing. – 2024. – doi.org/10.1007/s10845-024-02375-6.

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

С.В. Алубин (*sergei.alubin@yahoo.com*)

Университет ИТМО, Санкт-Петербург

Исследована актуальная и востребованная задача повышения качества образовательного процесса в учебных заведениях. Традиционные методы преподавания не учитывают индивидуальные особенности и пожелания учащихся, что сказывается на их результатах успеваемости. В статье предложена адаптивная система обучения, в которой исследователи сосредоточились на поиске решения вопросов организации более эффективного процесса обучения, стимулирования заинтересованности и успеваемости студентов с помощью современных технологий искусственного интеллекта. В рамках данной работы проведен эксперимент и предложен персонализированный подход обучения с использованием нейросетей. В качестве данных для эксперимента использованы результаты анкетирования студентов и тесты, выполненные учащимися в рамках учебного курса. Предложенные методы были реализованы в рамках эксперимента с участием более 500 студентов и получены положительные результаты. Разработана архитектура с использованием сверточных и рекуррентных нейронных сетей. По результатам эксперимента получен рост успеваемости студентов до 91,8% и персонализации учебного процесса до 95,3%. Данная методика может быть применена в рамках создания гибкой системы обучения, с целью повышения качества образовательных услуг и успеваемости студентов.

Ключевые слова: искусственный интеллект, нейронные сети, образовательный процесс, персонализированное обучение.

Введение

В последнее время спрос на различные образовательные услуги неизменно растет. Помимо государственных учебных заведений существует и ежегодно отрывается значительное количество частных компаний по ведению обучающих курсов. В частности, выручка российского рынка онлайн-

высшего образования в 2024 году по оценкам аналитиков достигла 4,5 млрд рублей¹. Одной из актуальных проблем в настоящее время является качество оказываемых услуг в сфере образования. Довольно часто распространена ситуация, когда студенты не довольны выбранным курсом обучения, стилем подачи материала, либо в процессе обучения у учащихся пропадает интерес. Таким образом, особое значение приобретает вопрос выработки подходов и методов для повышения качества обучения, мотивации учащихся, их успеваемости и вовлеченности в образовательный процесс. В научных исследованиях на данную тему многие авторы предлагают различные варианты прогнозирования успеваемости учащихся, однако отсутствуют решения, в которых в значимой степени учитываются психологических факторы, практические навыки и предпочтения студентов [1, 2].

Как показала практика, в традиционном процессе обучения без использования нейросети, на второй половине курса успеваемость и заинтересованность студентов существенно снижается вследствие психологических и поведенческих факторов.

В рамках повышения качества образовательного процесса предлагается разработать персонализированный подход к обучению для каждого студента.

Персонализированное обучение – это метод, который адаптирует образовательный процесс к индивидуальным потребностям и способностям каждого учащегося, учитывая его индивидуальные приоритеты, цели и психологический тип.

Реализация адаптивного обучения сформулирована как организация учебного материала под персональные приоритеты и потребности каждого учащегося [3, 4].

Цель исследования данной работы – с помощью технологий искусственного интеллекта разработать инструмент, который смог бы повысить уровень образовательного процесса, посещаемость и успеваемость студентов, стимулировать их мотивацию к обучению.

В рамках проведения эксперимента поставлены следующие задачи:

1. Разработать архитектуру нейросети в рамках персонализации программы обучения для каждого студента, с целью повышения мотивации и успеваемости в учебе.
2. Внедрить нейросеть в образовательную платформу и систему оценок успеваемости.
3. Провести индивидуальное анкетирование студентов.
4. Реализовать эксперимент и оценить эффективность применения нейросети.
5. Сформулировать выводы и подготовить статистику результатов.

¹ РБК Тренды <https://trends.rbc.ru/trends/education/676437649a79478a21d74da1>.

Уникальность и научная новизна заключается в том, что особое значение в исследовании уделялось созданию эффективной нейросети и определению индивидуального подхода, при котором студенты смогут полностью себя реализовать на курсе обучения при сохранении мотивации в течение всего курса и показать высокий уровень успеваемости. В значительной степени учитываются психологические факторы, влияющие на заинтересованность студента в образовательном процессе. В рамках анкетирования определяется психотип студента и предлагается индивидуальный подход для коммуникации.

Практическая ценность исследования: заключается в разработке технического программного обеспечения и методик с целью повышения качества образовательных услуг. Прикладная польза от внедрения данной методики заключается в повышении заинтересованности и мотивации учащихся. Предлагаемые инструменты позволяют в значительной мере повысить средний выпускной балл студентов, уровень подготовки учащихся по своей специализации, и в свою очередь, учебному заведению выпускать высококвалифицированных и востребованных специалистов.

В рамках эксперимента реализовано следующее:

1. Создание архитектуры нейросети для реализации персонального обучения.
2. Проведение детального анкетирования студентов.
3. Внедрение нейросети в учебный процесс.
4. Получение и анализ результатов.

В исследовании предложено предварительное анкетирование всех студентов перед началом учебного процесса. Выявление приоритетов, определение более интересных для студента разделов курса и черт характера учащегося. Данная информация анализируется нейросетью и позволяет предложить персональную программу обучения для каждого студента в отдельности, учитывая его индивидуальность, психологический тип, личные предпочтения и реализовать более эффективную линию поведения преподавателя с каждым конкретным студентом.

По результатам первичного анкетирования прогнозируется потенциальная успеваемость студента. По ответам на анкету определяются показатели, которые наибольшим образом влияют на мотивацию и успеваемость конкретного учащегося, формируется его психологический тип. Далее учебным заведением собирается индивидуальная программа – персонализированный курс, который состоит из теоретического материала, презентаций, практических заданий и тестов по темам, которые наиболее интересны и мотивируют студента в учебном процессе. Определяется эффективная и комфортная линия поведения преподавателя по отношению к учащемуся на основе психологического портрета и характера студента.

Затем проводится промежуточное анкетирование студентов в процессе обучения, что позволяет скорректировать подход к каждому студенту, предложить дополнительный материал к персонализированному курсу, изменить линию отношения со студентом, понять факторы для повышения мотивации студента к учебе.

После окончания курса проводится контрольное анкетирование, которое позволяет получить понимание – насколько учащийся доволен пройденным курсом обучения. Для дообучения нейросети и совершенствования системы учитываются пожелания и рекомендации студента.

Методы исследования

Используемые архитектуры нейросетей

В рамках эксперимента использовалась гибридная нейросеть с последовательной архитектурой:

- сверточная нейронная сеть (Convolutional Neural Network, CNN);
- рекуррентная нейронная сеть (Recurrent Neural Network, RNN).

Данная схема была выбрана вследствие ее простоты, поскольку последовательные соединения гораздо проще проектировать и обучать.

Нейронная сеть состоит из следующих элементов: входного слоя (256 нейронов); трех сверточных слоев (128, 64 и 32 фильтра); двух полносвязных слоев (128, 64 нейрона) и выходного слоя (16 нейронов). В нейросети применялись функции активации LSTM (Long Short-Term Memory). С целью обучения нейросети был использован фреймворк PyTorch машинного обучения (ML) с открытым исходным кодом, созданный на базе языка программирования Python и библиотеки Torch. Данная структура позволила получить значительную точность прогнозирования успеваемости и персонализированные рекомендации для студентов.

Экспериментальное исследование

Для обучения нейросети использовались данные анкетирования и результаты успеваемости студентов на курсах в рамках групповых и индивидуальных занятий. Эксперимент начинался в мае 2023 года и завершился весной 2025 года. За данный период времени было проанализировано более 500 анкет студентов различной возрастной категории от 18 до 55 лет, средний возраст учащихся – 25 лет, женщин – 57%, мужчин – 43%, находящиеся на всей территории России. Преимущественно обучение проводилось онлайн, с использованием специализированной платформы оценки успеваемости студентов. Все ученики проходили курс «Тестирование программного обеспечения», рассчитанный на 6 месяцев обучения и один месяц на сдачу дипломной работы.

Сбор данных для эксперимента

Предварительное анкетирование:

Проводился перед началом обучения, ожидания студента от прохождения курса.

- демографические данные: возраст и пол;
- имеющийся опыт работы по специализации обучения;
- информация о приоритетах студента на курсе;
- выборка разделов курса, где студент проявляет наибольшую заинтересованность;
- вопросы психологического характера для определения психотипа студента.

Ответы на данные вопросы позволяют сформировать индивидуальную программу курса обучения из интересующих студента разделов и тем. Предложить дополнительный материал для обучения. Определить более эффективную линию поведения преподавателя со студентом, исходя из его психотипа и характера, с целью более комфортных и доверительных отношений. Преподаватель должен являться наставником и авторитетом для студента.

Промежуточный опрос:

- оценка сложности практических заданий;
- удовлетворенность общения с преподавателем;
- необходимость дополнительного материала по курсу;
- удовлетворенность формой и стилем подачи материала;
- необходимость дополнительных практических заданий;
- общая оценка курса студентом.

Дополнительно оценивались данные:

- статистика посещаемости студента на занятиях и активность на онлайн платформе;
- результаты и оценки выполненных практических работ.

Позволяют откорректировать программу обучения. Добавить дополнительный материал, понять степень удовлетворенности студента материалом курса, психологической атмосферой и увлеченностью студента материалом.

Завершающий опрос, по окончании курса:

- оценка предоставленного учебного материала;
- удовлетворенность формой и стилем подачи материала;
- оценка сложности практических заданий;
- удовлетворенность общения с преподавателем;
- общая оценка курса студентом.

Дополнительно оценивались данные:

- статистика посещаемости студента на занятиях и активность на онлайн платформе;
- результаты и оценки выполненных практических работ.

В рамках исследования группы студентов были разделены на две: экспериментальная и контрольная. Обе группы учащихся проходили один курс обучения у тех же преподавателей.

Группа студентов в рамках эксперимента проходила обучение с применением гибкой системы с использованием нейросети. Группа для контроля обучалась по традиционной программе.

В анкетировании применялась шкала Ликерта по пяти балльной системе, где 1 – низкая оценка, 5 – самая высокая оценка.

В целях сохранения личных данных студентов применялось шифрование и протокол передачи данных SSL. Персональные данные студентов были скрыты в соответствии с Федеральным законом № 152-ФЗ «О персональных данных»².

Оценка показателей

В табл. 1 представлена информация о показателях точности рекомендаций предложенных нейросетью.

Таблица 1

№	Название нейросети	Точность рекомендаций (%)
1	Сверточная нейросеть CNN	88,5
2	Рекуррентная нейросеть RNN	89,7
3	Гибрид CNN и RNN	91,2

Прогнозируемая оценка успеваемости определялась по следующей формуле:

$$Z = 0,16A + 0,25B + 0,11C + 0,31D + 0,17E. \quad (1.1)$$

A – навыки по выбранной специализации;

B – данные психологических опросов;

C – приоритет учащегося для изучения материала;

D – удовлетворенность содержанием материала курса;

E – дисциплина посещения уроков и выполнения заданий;

Z – прогнозируемая итоговая успеваемость;

В формуле (1.1) указаны размеры весов, полученные экспериментальным путем в процессе обучения. Сеть корректирует внутренние параметры, чтобы минимизировать разницу между желаемым результатом и соб-

² Федеральный закон № 152-ФЗ от 27 июля 2006 года «О персональных данных» (с изм. на 08 августа 2024 года). <https://docs.cntd.ru/document/901990046>.

ственным прогнозом. Входные веса задавались случайно и корректировались в процессе обучения нейросети. По наиболее результативным выходным показателям были определены оптимальные значения весов, влияние каждого фактора в отдельности на успеваемость студентов. Случайная инициализация помогает начать с разных точек оптимизации. Процесс уточнения значения весов происходит через обратное распространение ошибки (backpropagation) и оптимизацию.

Общая сумма всех весов равна – 1.

Переменные указываются от математического нуля и до 1.

Где 1 – это максимальный показатель, например – дисциплина посещения занятий в течении всего курса 100%, если посещение занятий на 50%, то показатель равен – 0,5.

Навыки по специализации (A) – указываются имеющиеся навыки студента в изучаемой области, переводятся в значение от 0 и до 1. Определяется среднее число, отражающее все показатели. Данный показатель важен для более уверенного вхождения на курс обучения.

Психологическое анкетирование (B) – определяется психотип, черты характера учащегося с индексом в диапазоне от 0 и до 1. Здесь учитываются манера общения, сдержанность, умение слушать, предрасположенность к конфликтам и спорам или их отсутствие, темперамент студента. Для анализа данных, в том числе, применяются типологии личности Юнга и Левина.

Приоритет учащегося для изучения материала (C) – рассматриваются предпочтения и цели студента в изучении материала и наработки практического опыта. Важно выделить конечную цель участия студента на данном курсе и выбрать более приоритетные, индекс от 0 до 1.

Удовлетворенность содержанием материала курса (D) – показатель увлеченности и заинтересованности студента в индивидуальной программе обучения с приоритетом на интересующие учащегося темы. Индекс от 0 до 1, как среднее значение всех показателей.

Дисциплина посещения занятий и выполнения заданий (E) – показатель посещаемости студентом занятий, сданных работ и тестов с индексом от 0 до 1. Рассчитывается среднее значение, учитывая все элементы [5].

Итоговая оценка успеваемости (Z) – прогнозируемая оценка учащегося, рассчитывается как совокупность всех показателей от 0 и до 1. Затем успеваемость студента переводится в проценты, например, с показателем 0,75–75%.

В рамках обучения нейросети было определено, что наибольшие веса и значимость имеют следующие факторы: первичный опыт в изучаемой специализации; наличие подробного, понятного и интересного для студента учебного материала; психологические факторы, влияющие на студента в процессе его обучения; комфортная и доверительная коммуникация с преподавателем.

Переменные в формуле (1.1) были проанализированы и определены с помощью следующих методик:

Логистическая регрессия – определяет, какие переменные и факторы активнее влияют на результат успеваемости учащихся. Оценивается вероятность улучшения качества образования при изменении входящих факторов по следующей формуле:

$$P(y = 1) = \frac{1}{1 + e^{-(b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n)}}$$

где

- y – целевая переменная (0 или 1),
- x1, x2 ... xp – признаки (посещаемость, оценки),
- b0, b1 ... bp – коэффициенты модели,
- e – основание натурального логарифма.

Многофакторный дисперсионный анализ (MANOVA) – позволяет одновременно проанализировать влияние нескольких независимых факторов.

Методы кластеризации (K-means и DBSCAN) – анализируют эффективность преподавания. Студенты группируются по схожим характеристикам и показателям успеваемости, по каждой группе определяются успешные стратегии и факторы, максимально влияющие на результат обучения.

С учетом реализованных выше методик и использования различных параметров в работе и обучении нейросети, были определены оптимальные факторы для получения наилучших значений прогнозирования успеваемости учащихся.

Результаты эксперимента

В табл. 2 отражены показатели студентов в контрольной и экспериментальной группах студентов в конце курса обучения.

Таблица 2

№	Параметры	Контрольная группа, в процентах	Экспериментальная группа, в процентах
1	Мотивация к обучению	68,5	90,2
2	Персонализация обучения	63,1	95,3
3	Посещаемость	61,8	88,1
4	Эффективность коллективной работы	65,4	85,6
5	Психологическая устойчивость	68,1	90,5
6	Успеваемость, оценки выполненных работ	70,3	91,8
7	Удовлетворённость курсом обучения	76,7	95,1

На рис. 1 показана зависимость уровня различных компетенций студентов в контрольной и экспериментальной группах.



Рис. 1. Анализ показателей различных компетенций студентов в контрольной и экспериментальной группах

На рис. 2 показана зависимость успеваемости студентов в контрольной и экспериментальной группах – прогнозируемая в начале курса обучения, в середине и в конце.



Рис. 2. Анализ успеваемости студентов в контрольной и экспериментальной группах

На графике и в таблицах отражена положительная тенденция, которая отражает эффективность внедрения персонализации обучения в рамках проводимого эксперимента.

Преимущества индивидуализированного обучения с использованием нейросети способствует повышению мотивации, успеваемости студентов и сокращению времени на изучение учебного материала.

Обсуждение результатов

В результате экспериментов было выявлено незначительное влияние пола студентов, поэтому данный показатель был исключен из анализа. В то же время значительное влияние на успеваемость и мотивацию студента оказали следующие показатели:

- первичный опыт в изучаемой специализации;
- возможность выбрать индивидуальную программу обучения исходя из личных интересов студента (разделы и темы, предоставление дополнительного материала для изучения);
- психологический портрет учащегося (позволяет определить персонализированный подход к студенту как к личности и наладить комфортный диалог в процессе обучения);
- дисциплина выполнения заданий и посещения занятий.

В итоге, данные показатели, в совокупности, значительно повышают доверие студента к преподавателю, заинтересованность в изучаемом материале, личную мотивацию в завершении курса обучения и получение отличных оценок.

Заключение

По результатам проведенного эксперимента можно с уверенностью сказать, что гибкая система обучения с применением гибридной нейросети CNN и RNN показывают высокий уровень персонализации обучения и повышения мотивации и успеваемости учащихся.

Основные результаты:

- точность рекомендаций нейросети составляет 91,2%, что показывает ее высокую эффективность и практическую ценность;
- отмечается повышение мотивации студентов на 21,7%;
- показатель персонализации обучения показал значительный рост на 32,3%;
- уровень психологического комфорта учеников в процессе обучения увеличился на 22,4%;
- успеваемость учащихся в среднем улучшилась на 21,5%.

Залог успеха состоит в том, что система с помощью детальных входных данных и обучения нейросети позволяет разработать индивидуальный подход к каждому студенту таким образом, чтобы максимально удовлетворить его стремления, интересы, учесть наработанные навыки, подстроиться под психологический тип учащегося для взаимного комфортного общения между студентом и преподавателем. В итоге, подав-

ляющее число студентов удовлетворены пройденным курсом, подачей материала и персональным отношением. Позитивные отзывы также предоставлены многими студентами в виде рекомендательных писем с высокой оценкой работы преподавателей и качеством подачи учебного материала курса.

Таким образом, технологии искусственного интеллекта позволяют предоставить обратную связь и адаптировать учебную программу индивидуально для каждого студента, что в значительной мере сказывается на повышении их успеваемости.

Результаты исследования показали, что за искусственным интеллектом будущее и новые разработки алгоритмов позволят более эффективно использовать преимущества нейросетей в образовательном процессе. Использование нейросети в персонализированном обучении открывает новые возможности для студентов и преподавателей. Предложенные решения могут быть применены для реализации гибких образовательных систем в учебных заведениях. Необходимо изучить использование предложенной методики для различных специализаций обучения и образовательных институтов.

Список литературы

- [Савина, 2022] Савина Н.В. Персонализированное образование: взгляд студентов // Проблемы современного педагогического образования. – 2022. – С. 321-324.
- [Булаева, 2022] Булаева М.Н., Зубкова Я.В., Мельников Д.Д. Персонализированный подход в образовании // Проблемы современного педагогического образования. – 2022. – С. 4-6.
- [Анисимов, 2024] Анисимов А.В., Сазанова Е.В. Использование искусственного интеллекта в адаптивном обучении. // Молодежь и наука: актуальные проблемы фундаментальных и прикладных исследований. VII Всероссийская национальная научная конференция молодых ученых. – 2024. – С. 464-467.
- [Нечкин, 2024] Нечкин В.Н. Применение нейронных сетей в адаптивном обучении. Научный аспект. – 2024. – С. 3465-3471.
- [Huang, 2024] Huang D. Artificial intelligence driving innovation in higher education management and student training mechanisms // Applied Mathematics and Nonlinear Sciences. – 2024. – doi: 10.2478/amns-2024-0835.

УДК 004.93

doi: 10.15622/rcai.2025.066

ПРИМЕНЕНИЕ МЕТОДОВ 2D СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ НА ОСНОВЕ АРХИТЕКТУРЫ ТРАНСФОРМЕРА В ЗАДАЧЕ ДИСТАНЦИОННОГО ИССЛЕДОВАНИЯ СТРУКТУРЫ АРХЕОЛОГИЧЕСКИХ ПАМЯТНИКОВ¹

А.В. Вохминцев (*vav@csu.ru*)^A

В.Р. Аббазов (*abbavar@yandex.ru*)^B

М.А. Романов (*std.romanov.ma@gmail.com*)^A

Т.С. Вохминцева (*Atska85@mail.ru*)^B

М. Хатер (*Mostafa.khater2024@xzhmu.edu.cn*)^{B,C}

А.В. Мельников (*MelnikovAV@uriit.ru*)^B

^A Челябинский государственный университет, Челябинск

^B Югорский государственный университет, Ханты-Мансийск

^C Университет Суйчжоу, Суйчжоу, Китай

Данная работа посвящена разработке методов семантической сегментации археологических памятников с применением методов машинного обучения и 2d цифровых моделей археологического ландшафта по данным аэрофотосъемки и дистанционного зондирования Земли, был проведен анализ современных моделей для детектирования объектов и сегментации изображений, рассмотрены архитектуры на основе модели трансформера, такие как InternImage, BEiT Pre-Training of Image Transformers, проведено обучение моделей с использованием коллекции данных об археологических памятниках бронзового века Южного Зауралья с оценкой на валидационном наборе, полученные результаты представлены и обсуждены.

Ключевые слова: семантическая сегментация, трансформеры, цифровая модель рельефа, археологический памятник.

¹ Работа выполнена при финансовой поддержке РНФ (проект № 23-11-20007).

Введение

Данное исследование посвящено проблеме разработки системы для обнаружения и исследования археологических памятников на основе методов машинного обучения, картографирования, ГИС, геофизики и цифровых моделей археологического ландшафта. В период 80-х – 90-х гг. XX века проведена аэрофотосъемка с использованием малой авиации территории Южного-Зауралья и Северного Казахстана, при исследовании материалов съемок группой археологов под руководством проф. Здановича Г.Б. был открыт комплекс укрепленных поселений, датируемых XXI-XVIII вв. до н.э. бронзового века, который впоследствии стали относить к синташтинской археологической культуре. Самым известным обнаруженным укрепленным поселением является Аркаим, однако для исследователей наибольший интерес представляют менее известные поселения, например, археологические памятники вблизи поселений Степное и Левобережное. В результате археологических экспедиций по проекту РНФ № 23-11-20007 проведены поисковые работы и собрана цифровая коллекция данных об утраченных памятниках археологии, открытых в XIX-XX вв. без привязки к системе координат, поврежденных памятниках со снятым верхним почвенным слоем, а также нечитаемых на местности объектов в известных археологических памятниках (грунтовых захоронениях, ритуальных и хозяйственных сооружениях без выраженных в рельефе признаков). Возле многих из укрепленных поселений обнаружены отдельные неукрепленные поселения и курганные могильники. Комплекс археологических памятников синташтинской культуры, так называемая “Страна городов” является уникальным объектом для проведения междисциплинарных исследований с использованием методов геофизики, математического моделирования и искусственного интеллекта. Дистанционные методы исследования позволяют сохранить культурное наследие для будущих поколений, а также решать важные практические задачи, например, определение границ памятника и точных мест для забивки шурфов с целью получения археологического материала без раскопки поверхности памятника. В исследовании [Vokhmintcev et al., 2023a] предложена методология для картирования и съемки археологических объектов на основе методов машинного обучения, методов геофизики и картографирования. Ключевым компонентом данной методологии являются методы 2d сегментации данных, в данной статье предлагается оригинальный метод 2d сегментации археологических памятников на основе архитектуры трансформера. Согласно бенчмаркинг лидирующие позиции при решении поставленной задачи принадлежат методам сегментации на основе архитектуры трансформера, можно отметить следующие модели InternImage [Wang et al., 2023a], ONE-PEACE [Wang et al., 2023b], BEiT: BERT Pre-Training of Image Transformers [Bao et al., 2023], ViT-Adapter [Chen et al., 2023]. Давайте рас-

смотрим данные модели более подробно, InternImage – это крупномасштабная модель, основанная на свёрточных нейронных сетях и разработанная для задач компьютерного зрения, включая классификацию изображений, обнаружение объектов и семантическую сегментацию. Архитектура данной сети содержит базовый блок модели Deformable Convolution v3 (DCNv3), слой нормализации, функцию активации GELU и полносвязную сеть. Блок DCNv3 улучшает стандартную деформируемую свёртку, вводя обучаемые смещения и коэффициенты, которые позволяют ядру свёртки динамически приспосабливаться к структуре входного изображения. Это обеспечивает эффективное улавливание как локальных, так и глобальных зависимостей в данных. Процесс обработки данных в модели разделен на несколько этапов с последовательным уменьшением пространственного разрешения, что позволяет извлекать и обрабатывать признаки на разных уровнях абстракции с высоким значением метрики mask AP, модель протестирована на разных эталонных коллекциях данных, показывает выдающиеся результаты для коллекции COCO test-dev. Модель ONE-PEACE представляет собой универсальную архитектуру для обучения представлений, способную интегрировать и согласовывать данные различных модальностей: изображения, аудио и текста. Для каждой модальности модель использует специальные адаптеры, преобразующие исходные данные в последовательности признаков. Процесс обработки модальности изображений использует иерархический MLP-сте́м (hMLP) для разбивки изображения на сегменты размером 16×16 . Mask2Former применяется в качестве головы модели, обеспечивая подход к семантической сегментации с поддержкой различных задач сегментации. Модель сначала обучается на наборе данных COCO-Stuff для лучшего понимания контекстов объектов, а затем дообучается на целевом наборе данных ADE20K, достигая выдающихся результатов для метрики mIoU. BEiT – модель, которая изначально нацелена на обработку естественного языка (NLP), но затем разработчики адаптировали ее для задач в области компьютерного зрения. В основе архитектуры BEiT находится идея моделирования маскированного изображения (Masked Image Modeling MIM). Работа модели BEiT основана на случайном процессе маскирования определенной части сегментов исследуемого изображения. Далее модель обучается восстанавливать соответствующие визуальные токены, используя контекст оставшихся непрерывных сегментов. Модель BEiT основана на классической архитектуре трансформера при обработке последовательности сегментов изображения, что позволяет эффективно моделировать долгосрочные зависимости между различными частями изображения и интегрировать глобальные признаки объектов. Основным преимуществом модели BEiT является способность работать на глобальном уровне, предсказывая дискретные визуальные токены вместо предсказания пиксель-

ных значений маскированных сегментов, что приводит во многих известных методах 2d сегментации к фокусировке на локальном уровне сегментации, который затрагивает низкоуровневые детали изображения. Для получения визуальных токенов используется дискретный вариационный автоэнкодер dVAE [Vahdat et al., 2018], который осуществляет процесс токенизации изображения в последовательность дискретных представлений изображения. Данная особенность модели BEiT позволяет концентрироваться на высокоуровневых признаках объектов и семантической информации. Модель ViT-Adapter является улучшенной модификацией архитектуры трансформера Vision Transformer (ViT) [Dosovitskiy et al., 2018], которая позволяет повысить эффективность решения задачи 2d семантической сегментации. Модель ViT-Adapter состоит из трех компонентов: модуль 1 – пространственного приоритета (Spatial Prior Module, SPM), модуль 2 – инжектор пространственных признаков, модуль 3 для извлечения многоуровневых признаков. Модуль 1 содержит нескольких свёрточных слоёв и слоёв max-pooling, которые позволяют извлекать локальные пространственные признаки из изображения. Модуль 1 создает многоуровневые пространственные признаки с различным разрешением: 1/8, 1/16 и 1/32 от размера изображения. Модуль 2 интегрирует локальные пространственные признаки из SPM в одноуровневые токены ViT. Модуль 3 осуществляет реконструкцию многоуровневых признаков из исходных одноуровневых признаков ViT. Работа модели трансформера ViT-Adapter основана на концепции многоэтапного взаимодействия признаков. Слои трансформера разделяются на несколько блоков, на каждом этапе обработки осуществляется обмен информацией между локальными и глобальными признаками из SPM через Модуль 2 и Модуль 3. В данной работе согласно бенчмаркинг (<https://paperswithcode.com/task/semantic-segmentation>) для компьютерного моделирования выбраны модели InternImage и BEiT, проведено компьютерное моделирование отобранных моделей применительно к задаче 2d семантической сегментации, которое показало необходимость внесения изменений в архитектуру трансформеров. Статья организована следующим образом: в разделе 1 представлено описание коллекции данных археологических памятников бронзового века Южного Зауралья, в разделе 2 предложены оригинальные методы 2d семантической сегментации на основе архитектуры трансформера, в разделе 3 представлены результаты компьютерного моделирования.

1. Признаки дешифрирования археологических памятников бронзового века

Коллекция данных археологических памятников бронзового века Южного Зауралья создана на основе материалов экспедиций и полевых работ, проведенных специалистами учебно-научного центра изучения проблем

природы и человека ЧелГУ и историко-археологического музея-заповедника «Аркаим», начиная с 80-х годов прошлого века по настоящее время. Материалы для коллекции данных подобраны так, чтобы максимально представить типичные памятники степной зоны Зауралья: 22 укрепленных поселения эпохи бронзы, 16 неукрепленных поселений эпохи бронзы, 12 курганов и курганных могильников разных эпох, 5 средневековых курганов «с усами», 4 могильника с гантелевидными и подковообразными насыпями. Археологические памятники были обследованы с применением аэрофотосъемки, тахеометрической съемки и электроразведки: трехмерные данные были получены с использованием тахеометра сканеров глубины LiDAR в виде ортофотопланов и облаков точек, тахеометра Trimble 3300 и системы индукционного профилирования АЭМП-14 соответственно. Аэрофотоснимки были сделаны в масштабе 1:14 000 с высоким разрешением для всей территории Кизильского района Челябинской области. На основе собранных данных была построена цифровая модель археологического ландшафта опорных участков вокруг поселений Степное и Левобережное и территории археологического микрорайона в среднем течении р. Синташта, введены описания структур данных, соответствующих классам археологических объектов на языке графического описания объектного моделирования. В данной работе в качестве источников данных для обучения моделей были использованы: результаты локальной аэрофотосъемки с использованием малой авиации и квадрокоптера DJI Mini 2 (с 2022-2024 гг.) и результаты дистанционного зондирования Земли с зарубежных и российских спутников Sentinel-2 (с 2015 г.), Ресурс-П, (с 2013 по 2021 гг.), Канопус-В, (с 2013 по 2023 гг.) с пространственным разрешением 10 м, от 0.7 до 1.5 м, 2.1 м соответственно. Аэрофотоснимки прошли обработку в ПО Agisoftmetashape перед включением в коллекцию данных. В 2023 г. было открыто два новых укрепленных поселения Верхнеуральское и Нижнеуспенское, что стало большим событием в российской археологии: новые памятники бронзового века заполняли собой «белое пятно» на карте так называемой «Страны городов» между поселениями Степное и Черноречье и остальными объектами. В этой работе используются следующие классы археологических объектов: жилище и могильники.

2. Методы 2d семантической сегментации на основе архитектуры трансформера

Для реализации методов 2d семантической сегментации археологических памятников бронзового века Южного Зауралья предлагается подход, в котором можно выделить следующие шаги:

Шаг 1. Формирование набора снимков территорий, содержащих археологические памятники. Источником данных является существующая база аэрофотосъемки и космоснимков из коллекции данных.

Шаг 2. Создание разметки полученного набора. Разметка включает в себя классификацию археологических объектов по установленным классам и создание масок для задачи 2d семантической сегментации.

Шаг 3. Расширение размеченного набора данных в 10 раз с использованием различных вариантов аугментаций, основанных на повороте и переносе частей изображений, а также мозаичной технологии.

Шаг 4. Разделение аугментированного на шаге 3 набора снимков на обучающую, валидационную и тестовую выборку.

Шаг 5. Выполнение компьютерного моделирования современных моделей для 2d семантической сегментации изображений InternImage и BEiT на основе реальных данных и созданного на Шаге 4 набора снимков. Обучение данных моделей на тестовом наборе данных с оценкой моделей на валидационном наборе с целью предотвращения переобучения.

Шаг 6. Оценка моделей, отобранных на шаге 5 на тестовой выборке с использованием метрики IoU.

Далее рассмотрим подробно Шаг 5 подхода, в работе для 2d сегментации были использованы модели InternImage (см. Листинг 1) и BEiT (см. Листинг 2). Для задачи семантической сегментации InternImage используется в качестве энкодера, в качестве декодера используются такие модели как UperNet или SegFormer. Модифицированные шаги методов 2d сегментации отмечены символом *, были внесены изменения в процесс маскирования сегментов исследуемого изображения и восстановления визуальных токенов на основе контекста непрерывных сегментов.

Листинг 1

1:	procedure InternImage-UPerNet*(I)
2:	Входные данные: Изображение I / Выходные данные: Карта сегментации S
3:	// Этап 1. Извлечение признаков – backbone InternImage
4:	X_feat ← Преобразовать I в тензор входных признаков (X0)
5:	StageOutputs ← новый пустой словарь // хранение {X1,X2,X3,X4}
6:	for s_idx in 1,...,4 do // Индекс стадии (s_idx соответствует X1, X2, X3, X4)
7:	X_current_stage_processing ← X_feat // Рабочий тензор для текущей стадии
8:	for l_idx in 1,...,КоличествоБазовыхБлоков[s_idx] do // Индекс базового блока в текущей стадии
9:	X_residual_connection ← X_current_stage_processing
10:	// Вычисление смещений и модульных множителей DCNv3 на основе X_current_stage_processing

11:	Offsets, Modulators \leftarrow Вычислить DCN Параметры($X_{current_stage_processing}$)
12:	$H \leftarrow DCNv3(X_{current_stage_processing}, Offsets, Modulators)$ // H – промежуточный результат
13:	$*H \leftarrow LayerNorm(H); H_{ffn} \leftarrow FFN(H)$
15:	$X_{current_stage_processing} \leftarrow H_{ffn} +$ $X_{residual_connection}$ // Обновление тензора после блока
16:	end for // конец цикла по базовым блокам
17:	$StageOutputs[s_idx] \leftarrow X_{current_stage_processing}$
18:	if $s_idx < 4$ then // Если это не последняя стадия извлекающего признаки блока
19:	// Понижение разрешения для входа следующей стадии
20:	$X_{feat} \leftarrow Downsample(X_{current_stage_processing})$
21:	else // Для последней стадии (X_4) нет понижения разрешения; этот X_{feat} (X_4) пойдет в PPM
23:	$X_{feat} \leftarrow X_{current_stage_processing}$
24:	end if // конец условия понижения разрешения
25:	end for // конец цикла по стадиям
26:	// Извлечение именованных признаков стадий для удобства
27:	$X_1 \leftarrow StageOutputs[1,..., X_4 \leftarrow StageOutputs[4];$
31:	// Этап 2. Декодер UPerNet
32:	// 2.1 Pyramid Pooling Module (PPM)
33:	$PPM_features \leftarrow PyramidPoolingModule(X_4)$
34:	// 2.2 Feature Pyramid Network (FPN)
35:	// FPN: объединение признаков с разных уровней через lateral connections и upsampling
36:	$FPN_features \leftarrow FPN(\{PPM_features, X_3, X_2, X_1\})$
37:	// 2.3 Объединение признаков
38:	// FPN_ – это коллекция карт признаков с разных уровней FPN
39:	$*FusedFeatures \leftarrow$ Конкатенация&Сверточное слияние($FPN_features$)
40:	// Этап 3. Классификация и сегментация
41:	$SegmentationMap_logits \leftarrow Conv1x1(FusedFeatures)$ // Получение логитов для каждого класса
42:	$*SegmentationMap_upsampled \leftarrow Upsample(SegmentationMap_logits,$ до разрешения входного изображения I)
43:	// Этап 4. Финальный результат
44:	$*S \leftarrow \text{argmax}(SegmentationMap_upsampled, \text{по_каналам})$ // Выбор класса с максимальной вероятностью для каждого пикселя
45:	return S ; end procedure.

Листинг 2

1:	procedure BEIT-UPerNet*(I)
2:	// Этап 1. Представление I и предварительная обработка
3:	X_tensor ← Преобразовать I в тензор входных признаков
4:	Patches ← Разделить X_tensor на сетку из N патчей размером P×P (16×16) // N = (H_img/P) * (W_img/P)
5:	TokenSequence ← [] // Пустой список для последовательности эмбедингов
6:	for каждый Патч_i in Patches do
7:	Embedding_i ← ЛинейноеПреобразование(Патч_i) // Patch Embedding
8:	TokenSequence.Добавить(Embedding_i)
9:	end for // TokenSequence теперь имеет размерность N x D, где D – размерность эмбединга
10:	TokenSequence ← Добавить Абсолютные Позиционные Эмбединги(TokenSequence) // К каждому токenu
11:	// Этап 2. BEIT Encoder (Vision Transformer)
12:	// L – общее количество слоев трансформера
13:	// FeatureExtractionLayers – индексы слоев, после которых извлекаются признаки
14:	EncoderStageOutputs_raw ← новый пустой список // для {токены стадии1, ..., токены стадии4}
15:	CurrentTokenSequence ← TokenSequence // Рабочая последовательность токенов
16:	for l_idx in 1,...,L do // Индекс слоя трансформера
17:	InputToBlock ← CurrentTokenSequence
18:	// Multi-Head Self-Attention с относительным позиционным смещением (bias)
19:	AttentionOutput ← MultiHeadSelfAttention(InputToBlock, относительные позиционные bias)
20:	Residual1 ← InputToBlock + AttentionOutput // Первое skip
21:	Normalized1 ← LayerNorm(Residual1) // Первая нормализация
22:	// Feed-Forward Network
23:	*FFNOutput ← FFN(Normalized1)
24:	*Residual2 ← Normalized1 + FFNOutput // Второе skip
25:	CurrentTokenSequence ← LayerNorm(Residual2)
26:	InputToBlock_l ← CurrentTokenSequence
27:	*AttentionOut_l ← MultiHeadSelfAttention(InputToBlock_l, с относительными позиционными bias)
28:	Sum1_l ← InputToBlock_l + AttentionOut_l // Остаточное соединение
29:	Norm1_l ← LayerNorm(Sum1_l)
30:	FFNOut_l ← FFN(Norm1_l)
31:	CurrentTokenSequence ← Norm1_l + FFNOut_l // Остаточное соединение

32	if l_idx находится в FeatureExtractionLayers then
33	PatchTokens_l ← ОтброситьCLSTокенЕслиЕсть(CurrentTokenSequence)
34	EncoderStageOutputs_raw.Добавить(PatchTokens_l)
35	end if
36	end for // конец цикла по слоям трансформера
37	// Преобразование извлеченных последовательностей токенов в 2D карты признаков
38	// H_feat = H_img/P, W_feat = W_img/P
39	// D_stage_i - размерность эмбединга на выходе i-ой выбран- ной стадии
40	X1 ← ReshapeTo2D(EncoderStageOutputs_raw[0], H_feat, W_feat)
	// Карта признаков (H feat, W feat, D stage 1) . . .
43	X4 ← ReshapeTo2D(EncoderStageOutputs_raw[3], H_feat, W_feat)
	// Карта признаков (H feat, W feat, D stage 4)
44:	// Этап 2. Декодер UPerNet
45:	// 2.1 Pyramid Pooling Module (PPM)
46:	*PPM_features ← PyramidPoolingModule(X4)
47:	// 2.2 Feature Pyramid Network (FPN)
48:	// FPN: объединение признаков с разных уровней через lateral connections и upsampling
49:	FPN_features ← FPN({PPM_features, X3, X2, X1})
50:	// 2.3 Объединение признаков
51:	// FPN – это коллекция карт признаков с разных уровней FPN
52:	FusedFeatures ← Конкатенация&Сверточное_Слияние(FPN_features)
53:	// Этап 3. Классификация и сегментация
54:	SegmentationMap_logits ← Conv1x1(FusedFeatures) // Получение логитов для каждого класса
55:	*SegmentationMap_upsampled ← Upsample(SegmentationMap_logits, до разрешения входного изображения I)
56:	// Этап 4. Финальный результат
57:	*S ← argmax(SegmentationMap_upsampled, по_каналам) // Выбор класса с максимальной вероятностью для каждого пикселя
58:	return S; end procedure.

3. Компьютерное моделирование

В данном разделе представлены и обсуждены результаты компьютерного моделирования, для тестов применялся компьютер на базе Intel Core i7-9800X с графическим процессором NVIDIA GeForce RTX 2080 Ti 11264MiB. Оценим точность и сходимость предложенных методов на примере коллекции данных об археологических памятниках бронзового века Южного Зауралья, которая состоит из 9205 кадров в обучающей выборке и 3229 в тестовой. Результаты 2d семантической сегментации с использованием различных вариантов моделей InternImage и BEiT

для классов жилище и могильник приведены на рис. 1 и 2 соответственно. Для обучения модели InternImage использованы следующие параметры обучения: число эпох 100, оптимизатор AdamW, функция потерь Cross Entropy Loss, Learning Rate – LR, базовый learning rate – 0.00006, использовано полиномиальное затухание, был применен warmup со следующими параметрами: тип – линейный, количество итераций warmup – 5 (по эпохам), начальное значение – $1e^{-6}$ от основного LR. Для обучения были использованы модели InternImage-S (Small) и InternImage-B (Base).

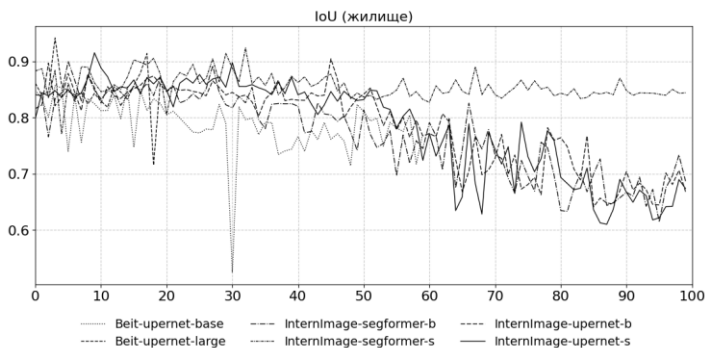


Рис. 1. Качество 2d семантической сегментации для класса жилище

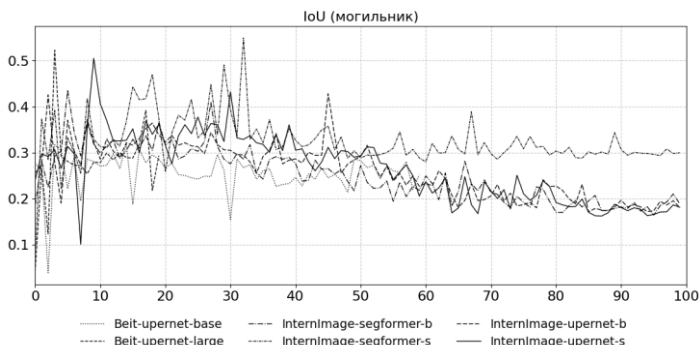


Рис. 2. Качество 2d семантической сегментации для класса могильник

Лучшие значения метрик IoU для разных моделей InternImage и BEiT представлены в табл. 1. Примеры сегментированных масок для неукрепленного поселения в среднем течении реки Синташта в Челябинской области приведены на рис. 3, масками отмечены объекты класса жилище. Для обучения модели BEiT использованы параметры: количество эпох обу-

чения для модели base – 60, для модели large – 20, использованы предобученные модели microsoft/beit-base-finetuned-ade-640-640 и microsoft/beit-large-finetuned-ade-640-640 соответственно, оптимизатор – AdamW, размер патча – 16, функция потерь – Cross Entropy Loss, learning rate – $5e^{-5}$. Дополнительно отметим, предложенные модели InternImage и BEiT производят 2d семантическую сегментацию лучше, чем базовые версии моделей: для InternImage-S разница составляет 0.15 пункта, InternImage-B – 0.09 пункта, Beit-upernet-base – 0.06 пункта, Beit-upernet-large – 0.08 пункта.

Таблица 1

Название модели	IoU, среднее (классы жилище и могильник)	IoU класс могильник	IoU класс жилище
InternImage-upernet, Small (s)	0.710	0.505	0.915
InternImage-segformer, Small (s)	0.737	0.549	0.924
InternImage-upernet, Base (b)	0.666	0.428	0.904
InternImage-segformer, Base (b)	0.668	0.436	0.900
Beit-upernet-base	0.587	0.320	0.857
Beit-upernet-large	0.732	0.522	0.942

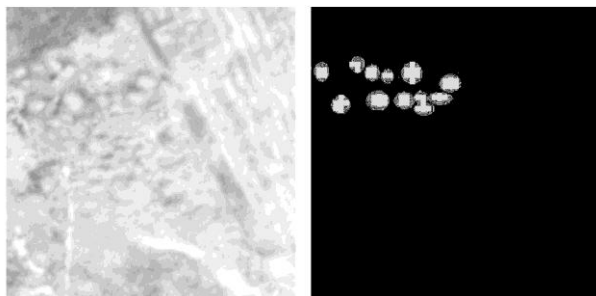


Рис. 3. Примеры сегментированных масок для неукрепленного поселения

В процессе обучения моделей для класса могильник получены невысокие значения по метрике качества IoU, с одной стороны это объясняется признаками дешифрирования данного класса [Vokhmintsev et al., 2023a], а с другой стороны связано с архитектурой трансформера BEiT. Решение данной проблемы видится в использовании для обучения моделей результатов магнитометрической съёмки, которая позволит идентифицировать могильники бронзового века с существенно более высокой точностью по сравнению с другими геофизическими методами и методами дистанционного зондирования Земли.

Заключение

В работе предложены модифицированные варианты методов InternImage и BEiT на основе архитектуры трансформера для задачи 2d семантической сегментации археологических памятников бронзового века Южного Зауралья на основе цифровой модели археологического ландшафта, была использована коллекция аэрофотоснимков и космических снимков с высоким пространственным разрешением. Полученные результаты позволяют автоматизировать процесс сегментации структуры археологического памятника, предложенные методы позволили с высокой точностью определять границы жилищ в поселениях бронзового века и как следствие определять точные места для забивки шурфов для получения археологического материала, не прибегая к раскопке поверхности всего памятника.

Список литературы

- [Bao et al., 2023] Bao H., Dong L., Piao S., Wei F. BEiT: BERT Pre-Training of Image Transformers // In: Proc. The 10-th International Conference on Learning Representations (ICLR), virtual event, 2022. – P. 1-16. – <https://arxiv.org/pdf/2106.08254v1>.
- [Chen et al., 2023] Chen Z., Duan Y., Wang W., He J., Lu T., Dai J., Qiao Y. Vision Transformer Adapter for Dense Predictions // The 11-th International Conference on Learning Representations (ICLR), Kigali, Rwanda, 2023. – P. 1-20. – arXiv preprint arXiv:2205.08534.
- [Dosovitskiy et al., 2018] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale // The 9-th International Conference on Learning Representations (ICLR), virtual event, 2021. – P. 1-22. – arXiv:2010.11929v2. [8].
- [Vahdat et al., 2018] Vahdat A., Andriyash E., Macreedy W.G. DVAE#: Discrete Variational Autoencoders with Relaxed Boltzmann Priors // The 22th Annual Conference on Neural Information Processing Systems (NIPS), Canada 2018. – 7.
- [Vokhmintcev et al., 2023a] Vokhmintcev A.V., Melnikov A.V., Romanov M.A. and etc. Research System of Archaeological Sites Using Deep Learning // Pattern Recognition and Images Analysis. – 2023. – No. 3. – P. 304-315. – doi: 10.1134/S105466182470038X.
- [Vokhmintcev et al., 2024b] Vokhmintcev A.V., Khristodulo O.I., Melnikov A.V. and etc. Application of Dynamic Graph CNN* and FICP for Detection and Research Archaeology Sites // International Conference on Analysis of Images, Social Networks and Texts. Ed. by D.I. Ignatov, M. Khachay, A. Kutuzov and etc. Lecture Notes in Computer Science. – 2024. – No. 14486. – P. 294-308. – doi:10.1007/978-3-031-54534-4_21.
- [Wang et al., 2023a] Wang W., Dai J., Chen Z., Huang Z., Li Z., Zhu X., Hu X., Lu T., Lu L., Li H., Wang X., Qiao Y. InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 2023. – P. 14408-14419. – <https://arxiv.org/abs/2211.05778>.
- [Wang et al., 2023b] Wang P., Wang S., Lin J., Bai S., Zhou X., Zhou J., Wang X., Zhou C. ONE-PEACE: Exploring One General Representation Model Toward Unlimited Modalities [Электронный ресурс] // Computer Vision and Pattern Recognition 2023. URL: <https://arxiv.org/abs/2305.11172> (дата обращения 19.04.2025).

УДК 004.82

doi: 10.15622/rcai.2025.067

СИСТЕМА ПОДДЕРЖКИ СОЗДАНИЯ КОРПУСА РУКОПИСНОГО НАСЛЕДИЯ ДРЕВНЕЙ РУСИ

Д.В. Демидов (*dvdemidov@mephi.ru*)

Национальный исследовательский ядерный университет «МИФИ»,
Москва

В работе описывается система поддержки разработки корпуса рукописного наследия Древней Руси и технология распознавания рукописей. Представляется концепция сращивания в единую среду средств оптического распознавания символов и машинного обучения с лингвистическим обеспечением и средствами поддержки работы инженеров и филологов. Целью совместной работы коллектива является получение цифрового корпуса рукописей, доступного для проведения палеографических и текстологических исследований.

Ключевые слова: оптическое распознавание; корпусная лингвистика; древние рукописи; эволюционное моделирование.

Введение

Идея создания корпуса рукописного наследия Древней Руси (далее Корпус) возникла в Институте Русского Языка им. В.В. Виноградова РАН, а её реализация началась в 2021 г. в стенах НИЯУ МИФИ совместными усилиями коллектива инженеров и филологов [Демидов и др., 2024a]. Источниками для Корпуса являются рукописные Служебные Миней XI–XVII вв., доступные в виде набора отсканированных разворотов книг. По памятникам этого типа можно проследить изменения в самом языке и многое почерпнуть об истории средневековой Руси. Однако изучение этого массива информации на бумажном носителе или в виде нераспознанных изображений, полученных в результате сканирования, чрезвычайно трудоёмко – поиск, различные виды анализа выполняются вручную.

Целью исследовательской группы является создание инструментария для работы со славянскими рукописями, а также их электронного издания. Это требует разработки идеологии и интерфейса «автоматизированного рабочего места филолога», а также применения системы распознавания. К сожалению, современные коммерческие системы и веб-сервисы оптиче-

ского распознавания не рассчитаны на работу с таким материалом, а специализированные платформы распознавания на основе моделей машинного обучения типа transkribus.org оказываются слабо применимы на практике.

Это связано с такими особенностями рассматриваемых источников как:

- древность самого языка, вышедшего из светского употребления;
- обилие графической информации помимо собственно символов (рис. 1): буквицы, надстрочные и диакритические знаки, вязь, крюковые ноты (межстрочные пометы-указания для певцов);
- сокращения слов, на что указывают титла и буквотитла;
- высокая дефектность при ветхом носителе: затемнение бумаги/пергамена, размытие чернил, потёртости, пятна, капли воска, утрата фрагментов, печати, карандашные пометы;
- геометрические искажения (неровные строки, вариация почерков), искажения, возникшие при сканировании (повороты, чёрные рамки).

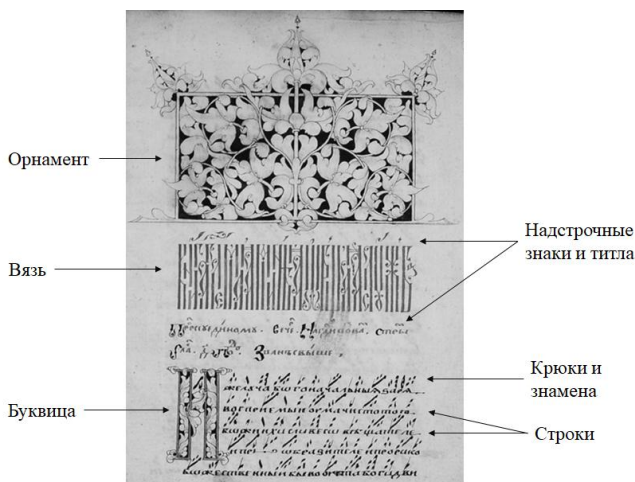


Рис. 1. Пример страницы Служебной Миней

Все эти обстоятельства требуют подготовки системы распознавания славянских рукописных Миней с необходимым лингвистическим и палеографическим обеспечением, выделением отдельных графических элементов букв для распознавания почерков и машинной датировки, каталогизацией источников для текстологических исследований.

Используемый рукописный материал характеризуется широким временным охватом (порядка 700 лет). За этот период существенно менялись все составляющие грамматики языка – азбука, лексика (орфография), морфология, синтаксис. Практически с нуля развилась пунктуация – по-

началу письмо было сплошным, без пробелов. В нотированных рукописях менялась и крюковая нотация (три поколения). В плане оптического распознавания современный русский язык представляется удобным статическим объектом, но древний славянский язык требует учёта фактора времени при обработке рукописей разных эпох: грамматика, крюковая нотация, графические традиции оформления становятся функциями от времени. С одной стороны, это ведёт к усложнению программных средств, а с другой, позволяет учитывать особенности разных периодов и открывает возможности для решения задач датировки рукописей. Встаёт вопрос об адаптивности как программного, так и лингвистического обеспечения для работы с подобным материалом.

Важно отметить, что у филологической группы объект исследования хоть и весьма широк, но определён изначально. К нему относятся:

- фрагменты книг для совершения богослужений: служебник, минея, триодь, паремийник, служебное евангелие, следованная псалтырь и др.;
- лексика, пунктуация, синтаксис церковнославянского языка;
- почерки;
- крюковая нотация.

У инженерной группы ситуация иная. В начале проекта предметом исследования были алгоритмы предобработки изображений текстов: устранение искажений, фильтрация, бинаризация и др. [Васелюк и др., 2024], алгоритмы сегментации текстовых зон, строк, букв и буквиц, надстрочных символов [Воронин и др., 2022], [Демидов и др., 2024b]. Известные подходы, изложенные, например, в [Гонсалес и др., 2012], [Визильтер и др., 2010] исследовались на применимость, шёл отбор наиболее перспективных методов для решения частных задач, таких как обнаружение линий разреза разворотов, текстурная сегментация орнаментов. Исследовалась задача поиска фрагментов текстов: полнотекстовый и нечёткий поиск по выражению [Маннинг и др., 2011], осложнённый наличием надстрочных знаков и титлов. Позднее предметом исследования стали модели классификации символов, модели синтеза размеченных данных, методы аугментации и доразметки. В фокусе этой работы – технология распознавания рукописей, эволюционирующая по мере создания Корпуса рукописного наследия Древней Руси.

1. Модель системы поддержки создания корпуса

Рассмотрим систему поддержки построения корпуса с точки зрения системного подхода. Отметим, что в целевой состав Корпуса входят:

1. Каталог рукописей с текстами.
2. Каталог графических объектов, представляющих отдельный интерес: орнаменты, вязь, отдельные буквы и лигатуры, буквицы.

3. Графические образы разворотов, страниц, строк, символов, букв, надстрочных знаков, крюков, знамен и др., сопоставленные с текстовой и признаковой информацией, используемой для поиска.
4. Морфологический словарь.

Следуя [Волкова и др., 2006], модель системы можно представить кортежем $S = \langle A, R, B, Z \rangle$, где A – множество изображений, включая как сканированные изображения рукописей, так и все производные изображения и фрагменты, продуцируемые средствами распознавания; B – множество текстов рукописей, в т.ч. нотируемых, продуцируемых средствами распознавания или привносимых извне; R – множество отношений между фрагментами изображений и фрагментами текстов, устанавливаемых в ходе работы над Корпусом; Z – множество целей системы, например, стремление обработать все входные изображения, максимизация метрик качества распознавания, установление соответствия между фрагментами изображений и фрагментами текстов и др. В более развитой форме модель системы можно представить кортежем $S = \langle Z, STR, TECH, COND \rangle$, где помимо целей Z имеется множество структур STR , реализующих цели, технология $TECH$ – методы, алгоритмы, модели, конвейеры обработки рукописей и страниц, и $COND$ – условия функционирования системы (внутренние и внешние), например, реализация проекта в академической среде учебного вуза.

1.1. Эволюционный подход к развитию системы поддержки создания корпуса

Первый прототип средств поддержки создания корпуса можно охарактеризовать как наивный: реализовывались классические подходы к распознаванию и апробировались вручную на небольшом количестве образцов. После создания средств каталогизации и первичной обработки страниц стал ясен масштаб Корпуса – сейчас коллектив активно работает с 245 рукописями, содержащими порядка 110 тысяч страниц.

Первичной целью было создание скелетного прототипа системы распознавания и параллельное улучшение отдельных алгоритмов. В табл. 1 приведены некоторые шаги развития алгоритмов и средств поддержки построения корпуса. Можно отметить общую тенденцию к усложнению и ветвлению алгоритмов в ходе активного применения метода проб и ошибок, накопление алгоритмов-кандидатов для повторного использования, иерархизацию конвейера обработки изображений: добавлялись новые этапы, некоторые этапы делились на подэтапы, результаты одного этапа определяли набор последующих этапов. Новые версии компонентов распознавания проверялись на всё большем объёме данных. Если видимые результаты внушали оптимизм, то изменения внедрялись. Однако, автоматическое применение алгоритмов ко всему объёму корпуса никак не облегчало оценку результатов, которая по-прежнему проводилась вруч-

ную и выборочно. К сожалению, новые версии программных средств не всегда давали однозначный результат: на каком-то материале удавалось получить прирост качества, а на ином результаты становились сильно хуже, причём далеко не всегда это было сразу заметно.

Таблица 1

Фаза	Предобработка	Сегментация	Классификация
1	Медианная фильтрация полутонового изображения; Глобальная бинаризация	Разбиение изображения на строки и символы методом профилей	Простая признаковая модель символа (моменты, профили)
2	Контрастирование; Адаптивная бинаризация; Нейросетевая бинаризация	Сегментация букв, вязи; Сегментация строк; Сегментация и аннотация символов внешним средством	Ручная разметка символов; Классификация букв, сегментированных вручную
3	Морфологические преобразования; Разделение методов предобработки в зависимости от цели и потребностей следующего шага конвейера.	Текстурная сегментация орнаментов, вязи; Сегментация крюков и знамен; Полуавтоматическая разрезка разворотов (с валидацией и коррекцией)	Расширенная признаковая модель символов (более 100 признаков); Нейросетевой классификатор крюков и знамен (разметка ручная)
4	Обнаружение геометрических искажений страницы (поворот, трапеция) и редактор для их устранения	Встроенные средства сегментации и аннотирования	Синтез данных для обучения на основе имитации шрифтов и реальных фонов страниц; Нейросетевой классификатор строк

Изначально спроектировать систему «от и до» с учётом всех этих аспектов, не представлялось возможным в силу фундаментальных причин:

1. Зависимость языка от периода времени и необходимость учёта фактора времени в средствах обработки.

2. Неполнота представлений команды разработки об объекте исследования. Ясно, что по мере углубления и уточнения этих представлений, система также должна эволюционировать.

3. Многообразие и неформализуемость языка в силу его естественно-го происхождения (несмотря на осознанное проектирование славянского алфавита под сложившуюся фонетику). Это приводит к сильному сцепле-

нию и взаимозависимости отдельных стадий анализа изображения. Как морфологический анализ нельзя завершить до завершения синтаксического из-за лексической омонимии, так и сегментацию нельзя завершить до завершения классификации из-за омонимии знаков и их фрагментов.

4. Имеющийся объём данных изначально требовал применения статистического подхода наряду со структурным и др., но без размеченных данных это было сильно затруднено. Вручную удалось разметить порядка 15000 образцов символов.

1.2. Критерии оценки успешности функционирования системы

В силу отмеченных фундаментальных причин нет возможности сразу создать идеальную систему, как нет возможности сформулировать единственный критерий успешности её функционирования. Это сложный процесс человеко-машинного взаимодействия, направленный на достижение совокупности неявных характеристик, обуславливающих практическую применимость материалов Корпуса.

Целесообразно изначально выстраивать критерии оценки качества решения каждой из подзадач, изучая взаимное влияние стадий обработки и добиваясь глобальной оптимизации. При этом движение в сторону оптимума возможно как путём изменения настроек компонентов, фиксируя их состав, так и за счёт усложнения внутренней структуры системы. Более того, сами критерии также должны эволюционировать по мере развития алгоритмов и появления новых метрик, вычисление которых ранее не было возможным. В задаче создания Корпуса наблюдались, по крайней мере, следующие фазы эволюции критериев оценки:

1. Качественная оценка оператором визуализированных результатов распознавания страниц.
2. Количественные статистические показатели с разным охватом (по-странично, по всем рукописям) для выявления грубых ошибок и оценки достигнутого прогресса.
3. Количественные показатели с потенциалом использования в качестве обратной связи для подстройки параметров алгоритмов и гиперпараметров моделей и как метрики для машинного обучения.

Критерии оценки можно разделить по подзадачам и автоматизируемым процессам: сегментация, классификация, поиск, обучение, валидация. В этот список можно добавить такие метапроцессы как валидация, оптимизация и перепроектирование (внесение инноваций), которые в ходе эволюции должны будут стать встроенными в систему и направлять дальнейшую эволюцию. Вместо количественных метрик качества можно рассматривать качественные в сочетании с такими подходами как нечёткие алгоритмы и нечёткие модели [Борисов и др., 2007]. Переход от системы, решающей конкретную задачу, к инструментальной системе и затем к интеллектуальной инструментальной системе, демонстрирующей целена-

правленную деятельность, давно известен на примере развития экспертных систем до оболочек и инструментальных комплексов ([Рыбина, 2014]). Конкретно для задачи распознавания изображений страниц древних рукописей уместны следующие критерии оценки, сгруппированные по этапам обработки:

1. Сегментация. Статистическая схожесть результатов сегментации реальных страниц рукописей с результатами сегментации синтезированных страниц или размеченных реальных страниц. Например, тяготение сегментов символов к срединным линиям, небольшой разброс высот символов, сходство текстурных признаков зон текста и пр.

2. Обучение и классификация. Традиционные точность, полнота, F-мера, близость матрицы неточностей к диагональной, непереобученность (устойчивость показателей распознавания на выборках различного масштаба).

3. Валидация. Снижение с течением времени количества ручных корrekтив в ходе валидации полученных результатов. Уместно для алгоритмов всех этапов от разрезки разворотов и устранения искажений до правок аннотаций или распознанного текста.

Чтобы в ходе развития системы трудоёмкость ручной валидации снижалась, а обратные связи автоматизировались в большей степени, необходимо ввести такие критерии качества, как простота модификации конвейера обработки и системы в целом, а также толерантность к вносимым изменениям (возможность отката в случае неудовлетворительного результата итерации).

2. Технология распознавания рукописей

Помимо средств автоматизации операций создание Корпуса требует системы хранения информации. Здесь уместны и транзакционная СУБД, и объектное хранилище, и документно-ориентированные постреляционные представления. В силу эволюционного характера деятельности по созданию Корпуса в основу поддерживающей эту деятельность технологии целесообразно положить принципы BASE (базовая доступность, мягкое состояние, потенциальная согласованность) [Pritchett, 2008] в противовес принципам ACID транзакционных баз данных (атомарность, согласованность данных, изоляция транзакций, устойчивость) [Gray, 1981]. Здесь под согласованностью данных подразумевается не свойство данных, хранящихся в БД, а согласованность выделенных фрагментов изображений с текстовыми фрагментами. Достижение этой согласованности является одной из целей системы. Для реализации принципа мягкого состояния усложнялись и жизненные циклы рукописей и страниц.

2.1. Жизненный цикл ключевых сущностей

Жизненный цикл рукописи:

1. Рукопись создана – в каталоге создана запись с библиографической информацией о рукописи.
2. Загружены изображения – сканированные изображения разворотов по отдельности или цельным документом.
3. Предложены линии разрезов разворотов на страницы.
4. Линии разреза валидированы – пользователь принял предложения или внёс коррективы линий (подвинул, удалил, добавил).
5. Разрезка осуществляется – автоматический процесс на основе согласованных линий разреза разворотов.
6. Рукопись разрезана на страницы, страницы доступны для просмотра и дальнейшей работы.
7. Запущена постраничная обработка рукописи – массовая операция по предобработке, сегментации и классификации символов.
8. Рукопись обработана, производные результаты доступны для оценки и корректировки.
9. Рукопись опубликована, результаты распознавания доступны для пользователей Корпуса.

Жизненный цикл страницы:

1. Новая страница (загружена или получена в результате разреза).
2. Страница в очереди на обработку.
3. Бинаризация – формируется монохромное изображение страницы для удобства последующих стадий обработки.
4. Сегментация – формируется иерархическое представление геометрической структуры страницы.
5. Классификация – типизация сегментов и классификация символов.
6. Визуализация результатов – сборка текста и генерация выходного изображения для качественной оценки результатов распознавания.
7. Страница обработана, конвейер обработки успешно пройден.
8. Валидация – если в процессе обработки были выявлены ошибки распознавания, пользователь может вручную исправить эти ошибки и перезапустить процесс обработки с учётом исправлений.
9. Страница валидирована, результаты работы могут пополнить обучающие выборки соответствующих подзадач и задействованы в дообучении/повторном обучении моделей.

В случае возникновения ошибок во время обработки, страница возвращается на предыдущий этап для повторной обработки.

2.2. Структура программных средств

Структурно система поддержки создания корпуса включает подсистему управления корпусом, подсистему распознавания, подсистему обучения и инженерную библиотеку (рис. 2). Для работы с корпусом в подсисте-

теме управления предусмотрен веб-портал, доступный в интернет по адресу slavcorp.ru. Основная функциональность: каталогизация рукописей, поиск рукописей по каталогу, полнотекстовый поиск фрагмента текста, редактирование морфологического словаря с поисковыми возможностями, валидация разрезов разворотов страниц, валидация сегментации текстовых зон на строки, валидация распознанного текста. Ролевая модель включает три категории: читатель, редактор (филолог или инженер), администратор.

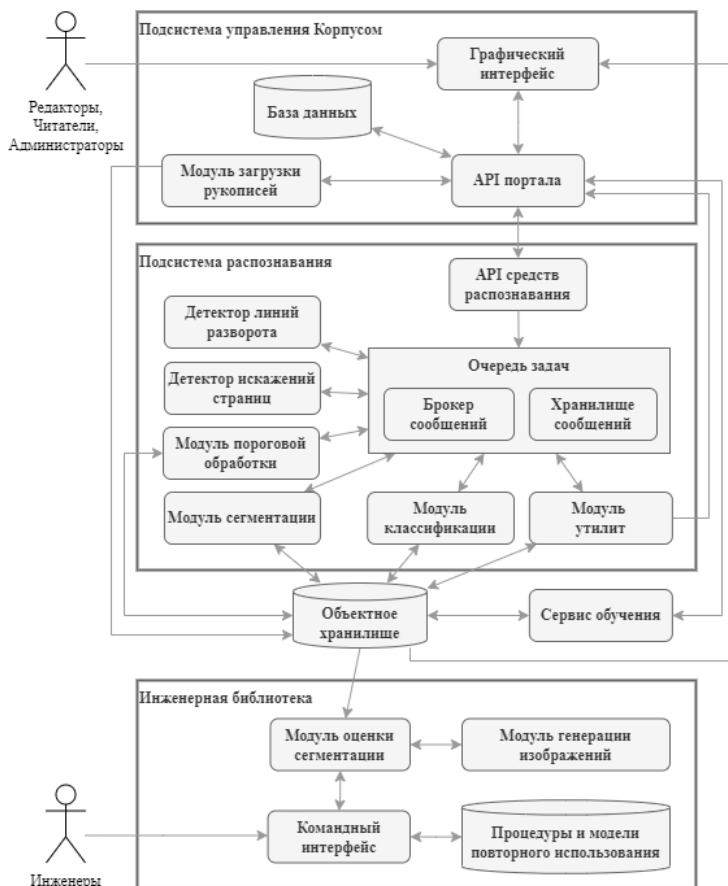


Рис. 2. Архитектура программных средств

Подсистема распознавания включает такие модули как детектор линий разворотов, детектор искажений страниц, модули пороговой обработки, сегментации, классификации и вспомогательных утилит. Доступ к моду-

лям предоставляется через программный интерфейс с применением очереди задач для асинхронной обработки массовых операций. Результаты обработки размещаются в объектном хранилище.

Подсистема обучения отвечает за обучение моделей классификации.

Инженерная библиотека поддерживает поисковую и адаптационную деятельность инженера. На основе материала, производимого подсистемами распознавания и обучения, экспериментально апробируются гипотезы для последующих параметрических, алгоритмических или структурных изменений системы с целью её совершенствования. В библиотеку входят повторно используемые процедуры и модели обработки, оценки, визуализации и др. Доступ к компонентам библиотеки осуществляется разными способами (записные книжки, командный и программный интерфейсы). Это открывает возможности структурного синтеза подпрограмм и сборочного программирования ([Тыгун, 1984], [Лаврищева, 2025]) и обеспечивает модифицируемость конвейера обработки рукописей.

Входными данными системы являются изображения текстов и в тех случаях, когда для рукописи удаётся найти уже готовый результат, сам текст. Также вводится метаинформация о рукописях, данные для морфологического словаря. Корректирующие и валидирующие действия можно отнести к управляющим воздействиям. Выходными данными являются тексты рукописей, распознанные и валидированные с помощью системы, поисковые выдачи, словарь в машиночитаемом формате.

Ручные и частично автоматизируемые операции: валидация результатов работы различных алгоритмов (зачатки механизма отбора для эволюционного процесса); коррекция промежуточных данных, служащих входными данными для следующих этапов; подстройка параметров алгоритмов/обучение моделей; принятие решений о внесении новых алгоритмов/модификаций существующих алгоритмов.

Пока точкой принятия решений об изменениях системы является человек (инженер или филолог), однако путём добавления оценочных модулей и формализации целевой функции создаются предпосылки делегирования части решений самой системе. Как только циклы обратной связи замыкаются, то более поздние стадии обработки получают возможность влиять на более ранние стадии в следующей итерации. Как отмечается в [Емельянов и др., 2003], обратные связи препятствуют такому действию системы, которое выводит её за установленные пределы. Версионированность повторно используемых компонентов гарантирует управляемость процесса эволюции системы. Таким образом наряду с машинным обучением и постепенным усложнением системы становятся применимы подходы эволюционной кибернетики.

3. Перспективы развития

Развитие системы поддержки создания корпуса рукописного наследия Древней Руси можно рассматривать в трёх направлениях.

Корпусная лингвистика. При наполнении Корпуса открываются перспективы построения лингвистических моделей древнего языка и синтаксически-размеченного корпуса с привязкой к периоду времени. Так как на раннем этапе развития русской письменности слова не разделялись графически, то задачи морфологического и синтаксического анализа требуют отдельной нетривиальной процедуры членения, в отличие от современного языка. В терминах трёхмерной стратификационной модели языка [Шаляпина, 2007], где ось членения на языковые единицы рассматривается многоярусной, переход с морфологического на синтаксический ярус мыслится итеративным, а не скачкообразным, при этом лексический ярус является результатом нахождения наиболее непротиворечивого варианта членения по отношению к морфологическому и синтаксическому ярусам. Эта задача не нова для глифowych языков типа китайского, японского или корейского, где нет графического разделения слов, а членение выполняется с опорой на морфологию, и аналогична задаче преобразования непрерывного речевого потока в текст, однако для раннего церковнославянского языка ставится, по всей видимости, впервые.

Обобщение на другие языки. Результатом обобщения разрабатываемой технологии может стать разделение системы поддержки создания корпуса на систему управления корпусом в виде относительно универсальной оболочки и надсистему со средствами визуального проектирования конвейеров обработки изображений текстов древних языков на основе эволюционного моделирования. Прежде всего рассматриваются славянские языки, языки бывшего СССР, например, древнеармянский и среднеармянский (V–XVII), где особый интерес представляет коллекция рукописей в Матенадаране (г. Ереван). В Отделе языков народов Азии и Африки Института Востоковедения РАН (до 2020 г. группа Шаляпиной З.М.) изучают доклассический и постклассический древнекитайский (V до н.э. – V в. н.э.), старояпонский и средневековый японский язык (VIII–XVII) и также испытывают значительные трудности из-за слабой применимости средств оптического распознавания текстов.

Расширение до экосистемы. Предоставление доступа к отдельным программным средствам в виде веб-сервисов для исследователей; создание мобильного приложения с поиском в Корпусе по фотографии фрагмента текста, «узнаванием» рукописи по фото. В образовательных и исследовательских целях интересна разработка средств автоматического перевода с церковнославянского на современный русский язык и наоборот.

Работа выполнена при поддержке программы Приоритет-2030 НИЯУ МИФИ.

Список литературы

- [Борисов и др., 2007] Борисов В.В., Круглов В.В., Федулов А.С. Нечёткие модели и сети. – М.: Горячая линия–Телеком, 2007. – 284 с.: ил.
- [Васелюк и др., 2024] Васелюк А.С., Демидов Д.В., Потемкина А.А. Фильтрация и бинаризация изображений текстов славянских рукописных книг // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во С.-Петерб. ун-та, 2024. – С. 43-50.
- [Визильтер и др., 2010] Визильтер Ю.В., Желтов С.Ю., Бондаренко А.В., Ососков М.В., Моржин А.В. Обработка и анализ изображений в задачах машинного зрения. – М.: Физматкнига, 2010. – 672 с.
- [Волкова и др., 2006] Волкова В.Н., Денисов А.А. Теория систем: учеб. пособие. – М.: Высшая школа, 2006. – 511 с.: ил.
- [Воронин и др., 2022] Воронин К.А., Демидов Д.В., Плетнёва А.А. Автоматическая сегментация строк в текстах на церковнославянском языке // Гибридные и синергетические интеллектуальные системы: [Электронный ресурс]: научное электронное издание. – Калининград: Издательство БФУ им. И. Канта, 2022. – С. 286-292.
- [Гонсалес и др., 2012] Гонсалес Р., Вудс Р. Цифровая обработка изображений. – 3-е изд. – М.: Техносфера, 2012. – 1104 с.
- [Демидов и др., 2024a] Демидов Д.В., Кравецкий А.Г., Ларионов А.А., Плетнёва А.А. Создание Корпуса рукописного наследия Древней Руси // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во СПбГУ, 2024. – С. 72-81.
- [Демидов и др., 2024b] Демидов Д.В., Миронова В.В. Предобработка изображений букв в славянских рукописных книгах // Труды международной конференции «Корпусная лингвистика – 2023». – СПб.: Изд-во СПбГУ, 2024. – С. 82-88.
- [Емельянов и др., 2003] Емельянов В.В., Курейчик В.В., Курейчик В.М. Теория и практика эволюционного моделирования. – М.: Физматлит, 2003. – 432 с.
- [Лаврищева, 2025] Лаврищева Е.М. Программная инженерия и технологии программирования сложных систем. – М.: Изд-во Юрайт, 2025. – 432 с.
- [Маннинг и др., 2011] Маннинг К.Д., Рагхаван П., Шютце Х. Введение в информационный поиск: пер. с англ. – М.: ООО «И.Д. Вильямс», 2011. – 528 с.
- [Рыбина, 2014] Рыбина Г.В. Основы построения интеллектуальных систем. – М.: Финансы и статистика, 2014. – 432 с.
- [Тыугу, 1984] Тыугу Э.Х. Концептуальное программирование. – М.: Наука. Главная редакция физико-математической литературы, 1984. – 256 с.
- [Шаляпина, 2007] Шаляпина З.М. Трёхмерная стратификационная модель языка и его функционирования: к общей теории лингвистических моделей. – М.: Вост. лит., 2007. – 480 с.
- [Gray, 1981] Gray J. The Transaction Concept: Virtues and Limitations // Very Large Data Bases, 7th International Conference, September 9–11, 1981, Cannes, France, Proceedings. IEEE Computer Society, 1981. – P. 144-154.
- [Pritchett, 2008] Pritchett, D. BASE: An Acid Alternative: In partitioned databases, trading some consistency for availability can lead to dramatic improvements in scalability // Queue. – 2008. – Vol. 6, Iss. 3. – P. 48-55.

УДК 004.838.8

doi: 10.15622/rcai.2025.068

ПОИСК СХОДСТВА ВЫИГРЫШНЫХ СТРАТЕГИЙ В ЗАДАЧЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

Д.А. Добрынин (*rabota1@mail.ru*)

Федеральный исследовательский центр
«Информатика и управление» РАН, Москва

В работе описывается метод получения обобщенных стратегий игры для задач интеллектуального анализа данных, использующий идеи сходства и индуктивного вывода. В качестве модели используется широко известная игра крестики-нолики. Для получения данных может быть использовано либо полное дерево игры крестики-нолики, либо набор из случайных игр. Описываются принципы получения сходства стратегий игры, алгоритм получения обобщенных стратегий. Приводятся некоторые результаты компьютерного моделирования. Показано, что применение данного метода позволяет получить обобщенные выигрышные стратегии. Полученные стратегии обладают хорошей объяснимостью и могут быть легко интерпретированы человеком.

Ключевые слова: машинное обучение, крестики-нолики, принятие решений, индуктивный вывод.

Введение

На сегодняшний день активно развиваются технологии машинного обучения в сфере информационных технологий. Для разработки и отладки методов машинного обучения активно используется теория игр, которая изучает стратегии принятия решений в конфликтных ситуациях, что делает ее полезной для развития методов ИИ.

Игра крестики-нолики является одной из старейших и поэтому хорошо изученных в теории игр [Гарднер, 1998]. Ее часто используют для тестирования различных методов машинного обучения. Несмотря на внешнюю простоту правил и небольшую размерность поля для игры, игра крестики-нолики имеет довольно большое дерево решений – порядка пятисот пяти-

десяти тысяч узлов. Полное дерево решений содержит порядка 25 правил (с учетом симметрии), некоторые из которых достаточно нетривиальны [Крестики-нолики, 2025].

Игра крестики-нолики имеет простые правила. Крестики ходят первыми. Игроки по очереди ставят на свободные клетки поля 3×3 знаки (один игрок всегда крестики, другой всегда нолики). Игрок, который выстроил в ряд 3 своих фигуры по вертикали, горизонтали или большой диагонали, выигрывает. Если игроки заполнили все 9 клеток и оказалось, что ни у одного из них нет выигрышной позиции, партия считается закончившейся в ничью.

На сегодняшний день игра крестики-нолики реализована почти для всех известных методов машинного обучения: Q-learning, обучение с подкреплением [Sutton, 2018], нейронные сети [Rajani, 2011] и глубокие нейронные сети. Обзор некоторых методов представлен в [Littman, 2004]. Сравнение методов для игры, основанных на нейронных сетях представлено в [Aditya, 2020].

На сегодняшний день одной из ключевых тенденций развития ИИ является требование объясняемости моделей ИИ [Mittelstadt, 2018]. Согласно определению, объясняемость означает способность понимать и оценивать решения и рассуждения, лежащие в основе прогнозов моделей ИИ. Объясняемость должна помочь конечному пользователю систем, основанных на ИИ, выявить ошибки путем анализа логики принятия решений. Если логика принятия решений скрыта, как для нейронных сетей, то возникает кризис доверия конечных пользователей к решениям, принимаемым системой ИИ. Особенно важно наличие объясняемости для областей применения, критических к возможным ошибкам, например медицине, энергетике, транспорту и им подобных.

Целью данной работы является разработка поиска сходства выигрышных стратегий для игры крестики-нолики с использованием идей сходства и индуктивного вывода. На этих принципах построен широко известный ДСМ метод [Финн, 2020], ВКФ метод [Виноградов, 2024] и некоторые другие. Задача получения полного набора правил игры крестики-нолики в данной работе не ставилась.

Основная идея разрабатываемого метода – получить из большого количества частных случаев обобщенные правила для выигрышной стратегии. Обобщение правил не только сокращает их количество, но и позволяет оценить их человеком. Такой подход лежит в русле разработки объяснимого ИИ, когда интеллектуальная система может объяснить свои действия на понятном человеку языке.

Выигрышные стратегии можно получить разными способами. Одно из возможных решений – выполнение случайных действий во время игры. Если получится выигрыш, то такая стратегия является выигрышной. При выполнении случайных действий может потребоваться большое время, чтобы получить достаточное количество вариантов игр.

Другой способ – использовать заранее построенное дерево игры, которое уже содержит все возможные варианты действий. Отметим, что построить полное дерево игры можно только для простых игр, у которых мало количество действий. Игра крестики-нолики содержит всего 9 вариантов действий, поэтому для нее можно построить полное дерево игры [Крестики-нолики, 2025].

Преимуществом второго подхода при разработке алгоритмов поиска сходства выигрышных стратегий является полнота решений, которые содержит полное дерево игры. При использовании случайных игр некоторые варианты могут быть пропущены, если количество игр недостаточно велико.

Отметим, что задача изучения влияния симметрии на количество получаемых правил в данной работе не ставилась. Интуитивно понятно, что поле для игры крестики-нолики имеет несколько видов симметрии. Поэтому учет симметрии может существенно сократить общее количество правил. Заметим, что при использовании многих методов машинного обучения симметрия не учитывается, либо учитывается неявным образом [Aditya, 2020]. Например, использование в правилах таких понятий, как «центр», «угол» или «середина стороны» указывает на использование симметрии. Задача обобщения правил с учетом возможных видов симметрии является довольно сложной и требует отдельных исследований.

Использование дерева решений

Для игры крестики-нолики можно построить полное дерево решений [Крестики-нолики, 2025]. Алгоритм построения и свойства дерева решений рассматривались ранее в [Добрынин, 2024].

Всего в дереве игры крестики-нолики десять уровней (рис. 1). Крестики ходят на нечетных уровнях, а нолики ходят на четных.

Верхний узел (0-й) соответствует началу игры – свободному полю. Первый уровень соответствует ходу для крестиков в одну из возможных позиций (всего 9 позиций). Второй уровень содержит узлы, в которых выполнен ход за ноликов в одну из свободных клеток (в оставшиеся 8 позиций). Для каждого узла строятся узлы с возможным ходом в оставшихся свободных позициях. Выигрыш возможен, начиная с пятого уровня по девятый. Узлы с выигрышем являются конечными. Для них потомки в дереве не строятся.

Каждый узел дерева связан с нижними узлами (потомками) и с верхним (родительским) узлом. Для удобства извлечения информации из дерева все узлы одного уровня связаны в список. Используя связи между узлами дерева можно строить различные процедуры извлечения данных.

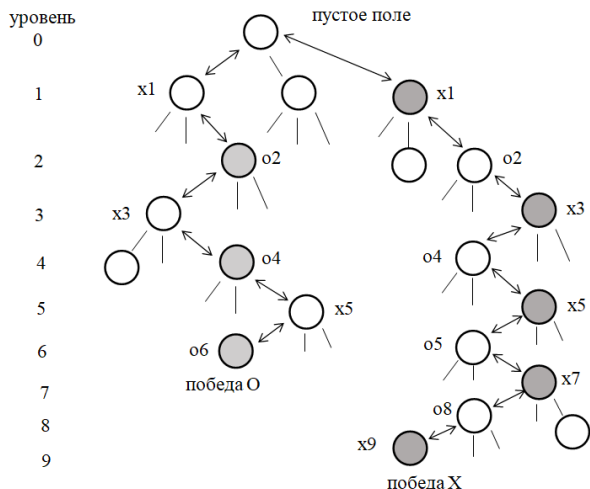


Рис. 1. Выигрышные стратегии в полном дереве игры.

Дерево игры крестики-нолики содержит все возможные ходы, поэтому оно является ценным источником информации об игре. Заметим, что полное дерево может быть построено только для простых игр, у которых количество вариантов ходов относительно невелико. Для игры крестики-нолики с девятью возможными вариантами ходов полное дерево содержит 549 946 узлов для 255 168 игр [Крестики-нолики, 2025].

Определим понятие выигрышной стратегии для полного дерева игры. Выигрышной стратегией будем называть последовательность ходов одного игрока, которая соответствует пути в дереве игры от первого узла до узла с выигрышем (рис. 1).

Игрок выполняет ходы, которые соответствуют состояниям узлов, входящих в стратегию, гарантированно приходит к выигрышу.

Каждый узел дерева может быть описан состоянием всех клеток игрового поля и ходом, который привел к нему.

$$x = (\{f_i\}, h), i = [1..9],$$

где f_i это состояние i -й клетки поля перед ходом,

$\{f_i\}$ – текущее состояние всех клеток,

h – ход, который привел к данному состоянию.

Обозначения и кодирование состояния клеток в виде битовых строк представлено в табл. 1. Битовое представление выбрано, исходя из простого представления операции сходства с помощью побитовой дизъюнкции.

Таблица 1

Обозначение	Кодирование	Назначение
х	0b001	крестик
о	0b010	нолик
. (точка)	0b100	свободная позиция
* (звездочка)	0b000	не важно (пустое значение \emptyset)

Ходы обозначаются:

х1..х9 это ходы крестиков,

о1..о9 это ходы ноликов.

Пример кодирования состояния поля:

.хх|оо.|... х1

перед ходом крестиков в клетку 1 (х1) получено состояние поля (рис. 2):

.	х	х	1	2	3	нумерация клеток поля
о	о	.	4	5	6	
.	.	.	7	8	9	

Рис. 2. Пример кодирования состояния поля.

Выигрышная стратегия Х может быть записана в виде кортежа, элементами которого являются состояния узлов дерева в цепочке от выигрыша до первого уровня:

$$X_j = (x_j, x_{j-2}, x_{j-4}, \dots, x_1),$$

где j – номер уровня дерева, на котором произошел выигрыш,

Отметим, что нулевой уровень соответствует пустому полю, поэтому его состояние можно не учитывать. Элементы кортежа записываются в обратном порядке – от j до 1, что удобно при поиске сходства. Для крестиков используются нечетные уровни, а для ноликов – четные.

Для рис. 1 можно записать две выигрышных стратегии – для ноликов и для крестиков:

$$O_6 = (o_6, o_4, o_2)$$

$$X_9 = (x_9, x_7, x_5, x_3, x_1).$$

Отметим, что элементы одного и того же уровня для стратегий O_6 и X_9 , соответствуют разным узлам дерева игры и потому различны.

Поиск сходства выигрышных стратегий

Для того, чтобы построить выигрышную стратегию, состоящую из обобщенных правил, необходимо находить сходство двух и более стратегий.

Для поиска сходства двух стратегий будем использовать операцию индукции. Эта операция ищет общие части элементов кортежей, которые определяют выигрышные стратегии.

Сходство двух стратегий определяется как

$$G = X_i \cap Y_i = (x_i \cap y_i \neq \emptyset, x_{i-2} \cap y_{i-2} \neq \emptyset, \dots, x_k \cap y_k \neq \emptyset),$$

где $x_{k-2} \cap y_{k-2} = \emptyset$

Искать сходство можно только для стратегий, соответствующих выигрышу на одном уровне дерева игры. Поэтому

$$G = X_i \cap Y_j = \emptyset, i \neq j.$$

Таким образом, пересечение двух стратегий содержит непустые пересечения элементов кортежа, начиная с уровня i . Если на уровне $k-2$ получается пустой элемент (пересечение пусто), то пересечения для остальных уровней не включаются в результат.

Для получения правил выполнения хода необходимо с помощью индукции получить гипотезы. Если поле состояний гипотезы вкладывается в поле текущего состояния игры, то может быть выполнен ход, присутствующий в данной гипотезе.

Гипотезы можно считать знаниями, поскольку они описывают общие зависимости, извлеченные из исходных данных (фактов).

Гипотеза есть пересечение двух и более примеров

$$g = x_i \cap x_j, i \neq j.$$

Правила пересечения приведены в табл. 2. Пересечение с пустым элементом \emptyset всегда дает пустой элемент. Пересечение элемента f_i с собой дает тот же элемент. Если элементы различны, получается пустой элемент.

Таблица 2

$f_i \setminus f_i$	x	o	свободно	\emptyset
x	x	\emptyset	\emptyset	\emptyset
o	\emptyset	o	\emptyset	\emptyset
свободно	\emptyset	\emptyset	свободно	\emptyset
\emptyset	\emptyset	\emptyset	\emptyset	\emptyset

Пример гипотезы:

$$.xx|***|*** x1$$

перед ходом крестиком в поле 1 состояние поля должно быть

.	x	x
*	*	*
*	*	*

нетрудно заметить, что после совершения хода крестиками они выиграют (будет получено три крестика в верхней строчке). При этом учитываются только состояние клеток 1-3, а значения остальных клеток не важны.

Полученные сходства выигрышных стратегий необходимо обобщить, что получить более компактные записи правил. Для этого можно использовать процедуру обобщения, предложенную ранее в [Добрынин, 2024]. Это циклическая процедура, заключающаяся в последовательном применении индукции, фильтрации и отбора минимальных гипотез.

Индукция – это операция получения общих частей исходных примеров во всех возможных комбинациях [Финн, 2020]. После применения индукции генерируется большое количество гипотез, которые представляют собой общие части исходных примеров. Фильтрация – отбор гипотез по определенным критериям, на основе выбранной эвристики. Фильтрация необходима для выделения значимых гипотез из большого количества всех возможных пересечений. Процедура получения минимальных гипотез значительно уменьшает количество гипотез, оставляя только те из них, которые обладают наиболее общими свойствами.

Получение обобщенных стратегий

Процедура получения выигрышных стратегий, содержащих обобщенные правила, представлена на рис. 3. Выигрышная стратегия может быть получена либо из полного дерева решений, либо проведением случайной игры. Далее происходит обработка каждого хода, принадлежащего стратегии. Схожие стратегии (с выигрышами на том же уровне) берутся из банка стратегий. Изначально банк стратегий пуст. Он заполняется по мере работы алгоритма.

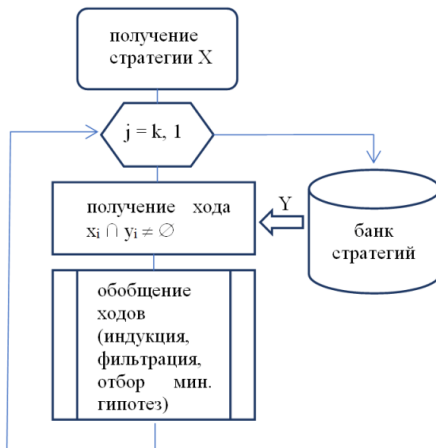


Рис. 3. Получение обобщенных стратегий

При поиске пересечения двух схожих стратегий происходит поиск сходства ходов, принадлежащих стратегии. Остаются только непустые пересечения ходов, как было показано ранее. Для получения обобщенного хода запускается циклическая процедура обобщения, состоящая из трех этапов – индукции, фильтрации и поиска минимальных гипотез. Процедура обобщения, подробно рассмотренная в [Добрынин, 2024], заменяется частные случаи правил на более общие. Это позволяет уменьшить общее количество правил и улучшает их объяснимость.

Сформированная обобщенная стратегия помещается в банк стратегий. Информация из этого банка используется для дальнейшего поиска обобщенных стратегий.

Отметим, что решающие правила, находящиеся в банке стратегий, связаны друг с другом. Последовательность правил от уровня 1 до уровня 9, образуют выигрышную стратегию для игрока. Правила для крестиков и для ноликов хранятся независимо друг от друга.

Критерии завершения процедуры получения обобщенных стратегий зависят от способа получения исходных стратегий. Если выигрышные стратегии берутся из полного дерева решений, то достаточно перебрать все элементы дерева, соответствующие выигрышным ситуациям. Если выигрышные стратегии берутся из случайных игр, то четкого критерия завершения процедуры не существует. Необходимо провести большое количество случайных игр, пока все возможные правила не будут получены.

Пример.

Получение обобщенных стратегий выигрыша для уровня 5 полного дерева игры крестики-нолики (выигрыш крестиков) для хода x1.

Исходных игр 45, представлены выигрышные позиции на уровне 5 (табл. 3)

Таблица 3

.xx oo ... x1	.xx ..o o.. x1	.o. x.. xo. x1	... x.o x.o. x1	..o .xo ..x x1
.xx o.o ... x1	.xx ..o .o. x1	.o. x.. x.o x1	... x.o x.o x1	..o .x .o.x x1
.xx o.. o.. x1	.xx ..o ..o x1	..o x.o .x.. x1	... x.. xoo x1	..o .x .ox x1
.xx o.. .o. x1	.xx ... oo. x1	..o x.o x.. x1	..oo x.. .x x1	... oxo ..x x1
.xx o.. .o x1	.xx ... o.o x1	..o x.. xo. x1	.o. ox .x x1	... ox .o.x x1
.xx .oo ... x1	.xx ... oo x1	..o x.. x.o x1	.o. .xo .x x1	... ox .ox x1
.xx .o. o.. x1	..oo x.. x.. x1	... xoo x.. x1	.o. .x .o.x x1xo .o.x x1
.xx .o. .o. x1	.o. xo .x.. x1	... xo .xo. x1	.o. .x .ox x1xo .ox x1
.xx .o.. .o x1	..o. x.o x.. x1	... xo .x.o x1	..o ox .x x1x .oox x1

Из табл. 3 хорошо видно, что представленные позиции описывают все возможные расстановки для верхней строки, левого столбца и одной диагонали поля игры.

Для полученных выигрышных позиций сгенерировано 90 выигрышных стратегий.

После обобщения стратегий, получаются следующие последовательности ходов для крестиков (табл. 4).

Таблица 4

№	уровень 5	уровень 3	уровень 1
1	.xx *** *** x1	.x. *** *** x3	... *** *** x2
2	.xx *** *** x1	..x *** *** x2	... *** *** x3
3	.** x** x** x1	.** x** .** x7	.** .** .** x6
4	.** x** x** x1	.** .** x** x6	.** .** .** x7
5	.** *x* **x x1	.** *x* .** x9	.** .** .** x5
6	.** *x* **x x1	.** .** .** x5	.** .** .** x9

Первая группа (стратегии 1 и 2) соответствует заполнению верхней строки. Вторая группа (стратегии 3 и 4) соответствует заполнению левого столбца. И наконец, третья группа (стратегии 5 и 6) соответствует заполнению диагонали.

В ходе применения процедуры обобщения стратегий их количество сократилось в 15 раз ($90/6=15$).

Эксперименты показывают, что для ходов в другие клетки получают аналогичные обобщенные стратегии, которые применимы к строкам, столбцам и диагоналям игрового поля.

Заключение

Получение стратегий игры, содержащих обобщенные правила, является важным методом интеллектуального анализа данных.

Обобщение решающих правил (замена многих частных случаев небольших количеством общих правил) позволяет сократить количество правил (в 15 раз для данных примеров) и улучшить их обзорность человеком. Данные исследования направлены на разработку объяснимого ИИ.

Разрабатываемый метод поиска стратегий игры основан на идеях схождения и использует операции с низкой вычислительной сложностью. Как было показано, стратегии игры крестики–нолики могут быть получены либо из полного дерева решений, либо проведением большого количества случайных игр. Численные оценки времени поиска стратегий в данной работе не проводились. Это задача для будущих исследований в этой области.

Другой интересной задачей будущих исследований является учет симметрии при поиске общих правил. Учет симметрии значительно сокращает общее количество решающих правил. Исследования в этой области требуют дальнейших разработок.

Список литературы

- [Гарднер, 1998] Мартин Гарднер Крестики-нолики: пер. с англ. И.Е. Зино. – М.: Мир, 1988. – 350 с. – ISBN 5-03-001234-6.
- [Крестики-нолики, 2025] Википедия, игра крестики-нолики [Электронный ресурс]. – URL: <https://ru.wikipedia.org/wiki/Крестики-нолики> (дата обращения: 26.05.2025).
- [Sutton, 2018] Richard S. Sutton, Andrew G. Barto Reinforcement learning: an introduction. Second edition. – MIT Press, Cambridge, MA, 2018.
- [Rajani, 2011] Rajani Nazneen & Dar Gaurav & Biswas Rajoshi & Ramesha C.K. Solution to the Tic-Tac-Toe Problem Using Hamming Distance Approach in a Neural Network. Intelligent Systems, Modelling and Simulation // International Conference on. – 2011. – P. 3-6. – DOI: 10.1109/ISMS.2011.70.
- [Littman, 2004] Littman M.L., Stone P. Game theory and machine learning // Proceedings of the Twenty-First International Conference on Machine Learning. – 2004.
- [Aditya, 2020] Paul Aditya. Randomized fast no-loss expert system to play tic tac toe like a human // Cognitive Computation and Systems. – 2020. – 2. – DOI: 10.1049/ccs.2020.0018.
- [Mittelstadt, 2018] Mittelstadt B. and Russell C. and Wachter S. Explaining Explanations in AI (November 4, 2018) // Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency (FAT* '19), January 29–31, 2019, Atlanta, GA, USA. ACM, New York, NY, USA. – doi/10.1145/3287560.3287574. – ISBN: 978-1-4503-6125-5. Available at SSRN: <https://ssrn.com/abstract=3278331>.
- [Финн, 2020] Автоматическое порождение гипотез в интеллектуальных системах / под общ. ред. В.К. Финна. – М.: Книжный дом «ЛИБРОКОМ», 2020. – 2-е изд. стереотип.
- [Виноградов, 2024] Виноградов Д.В., Якимова Л.А. Вероятностный подход к порождению стратегий для игр с полной информацией // Сборник научных трудов XII Международной научно-практической конференции Интегрированные модели и мягкие вычисления в искусственном интеллекте (ИММВ-2024, Коломна, 14-17 мая 2024 г.). В 2-х т. Т. 1. – Смоленск: Универсум, 2024. – С. 266-273.
- [Добрынин, 2024] Добрынин Д.А. Об одном методе обобщения правил для задач интеллектуального анализа данных // Сборник научных трудов XII Международной научно-практической конференции Интегрированные модели и мягкие вычисления в искусственном интеллекте (ИММВ-2024, Коломна, 14-17 мая 2024 г.). В 2-х т. Т. 1. – Смоленск: Универсум, 2024. – С. 274-284.

УДК 007.5:519.816:681.3.016

doi: 10.15622/rcai.2025.069

МОДЕЛЬ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ С ПРИМЕНЕНИЕМ АДАПТИВНОЙ СЕТИ И НЕЧЕТКОГО ВЫВОДА¹

А.П. Еремеев (*eremeev@appmat.ru*)

В.С. Петров (*PetrovVS@mpei.ru*)

Национальный исследовательский университет, «МЭИ», Москва

В работе предлагается модель глубокого обучения с подкреплением (Deep Reinforcement Learning, DRL) на основе адаптивной нейронной сети и системы нечеткого вывода. Ключевой целью является разработка высокопроизводительного агента DRL, обладающего повышенным уровнем объяснимости по сравнению с традиционными агентами, функционирующими как «черные ящики». Проводится тестирование и сравнительный анализ на примере задачи CartPole (тележка с шестом). Полученные результаты подтверждают потенциал гибридных моделей, сочетающих методы DRL и нечеткой логики, для создания эффективных и интерпретируемых систем искусственного интеллекта, критически важных в прикладных областях, требующих доверия и прозрачности.

Ключевые слова: искусственный интеллект, обучение с подкреплением, нейронная сеть, нечеткая логика, нечеткие системы, система реального времени.

Введение

Традиционно в системах нечеткого управления широко применяется объединение модели динамики объекта со средствами нечеткого логического вывода, что приводит к формированию динамических нечетких регуляторов [Presup et al., 2011]. Однако в ряде прикладных задач, таких как интерактивные программные среды или виртуальная реальность, построение точной математической модели объекта затруднено или вовсе невозможно. В таких условиях классические подходы, основанные на математическом моделировании, оказываются недостаточно эффективными [Reiners et al., 2021].

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 24-11-00285) <https://rscf.ru/project/24-11-00285/>.

Обучение с подкреплением (Reinforcement Learning, RL) представляет собой класс методов машинного обучения, ориентированных на выработку стратегий принятия решений в условиях неопределенности. Алгоритмы RL позволяют агенту постепенно улучшать поведение в процессе взаимодействия со средой, используя информацию о полученных наградах для корректировки будущих действий [Russell et al., 2010].

В последние годы методы глубокого обучения в сочетании с RL – методы DRL – продемонстрировали качественный прорыв: во многих задачах агенты на базе DRL показывают более высокие результаты, чем люди [Zhao et al., 2022]. Эти методы и модели на их основе успешно применяются, например: для навигации беспилотных летательных аппаратов [Walker et al., 2019], дообучения больших языковых моделей [Ouyang et al., 2022] и в других приложениях. Вместе с тем характерной проблемой таких моделей остается их ограниченная прозрачность, затрудняющая интерпретацию принятых решений.

Для устранения этого недостатка интенсивно развиваются методы и модели объяснимого искусственного интеллекта, направленные на создание моделей (и алгоритмов) с понятной и воспроизводимой логикой работы [Borys et al., 2023]. Особенно перспективными считаются гибридные нечеткие модели, поскольку они позволяют формализовать правила в виде понятных формулировок и визуализировать логические зависимости [Борисов и др., 2012], [Нечеткие гибридные системы, 2007], [Mencar et al., 2019], [Víaña et al., 2023].

Особый интерес представляют гибридные решения, объединяющие нечеткую логику и методы RL [Berenji, 1992]. Первые исследования по нечеткому RL (Q-обучению) [Er et al., 2004], [Jamshidi et al., 2015] показали эффективность в задачах управления, однако в них уделялось мало внимания вопросам прозрачности принимаемых решений. В работе [Kumar, 2020] предложен гибридный механизм, в котором базовые правила Такаги–Сугено–Канга (TSK) интегрируются с алгоритмом Deep Q-Network (DQN) [Mnih et al., 2015]. В этой схеме алгоритм RL отвечает за одновременную корректировку функций принадлежности и линейных коэффициентов вывода нечеткой модели, аналогично тому, как это делается в модели (системе) ANFIS (Adaptive Neuro-Fuzzy Inference System) [Jang, 1993], представляющей собой развитие TSK-моделей. Благодаря этому система самостоятельно подстраивает нечеткие правила под динамику окружающей среды, что подтверждается ее высокой практической эффективностью.

В предыдущих работах авторов рассматривались алгоритмы RL с нечеткой логикой и сравнивалась их эффективность [Eremeev et al., 2024], а также исследовались возможности самоорганизующейся нечеткой Q-сети [Еремеев и др., 2023], с целью применения в интеллектуальных системах поддержки принятия решений в реальном времени (ИСППР РВ).

Целью работы являются дальнейшие исследования нечетких систем с применением алгоритмов DRL в качестве решений для создания объяснимого искусственного интеллекта. Работа выполнена в рамках тематики научной группы кафедры Прикладной математики и искусственного интеллекта НИУ «МЭИ» по созданию инструментальных средств (методов, моделей, программ) конструирования ИСППР РВ для мониторинга и управления сложными техническими и другими системами.

1. Обучение с подкреплением с использованием адаптивной нейро-нечеткой модели

Интеграция в модели обучения с подкреплением и нечеткой логики была предложена [Berenji, 1992] на основе классической схемы Мамдани: при этом агент корректирует правила нечеткого вывода, получая сигналы вознаграждения за свои действия в среде. Первоначальные эксперименты, в частности, на задаче управления тележкой с шестом CartPole [Barto et al., 1983], где агент постепенно совершенствует стратегию балансирования, подтвердили практическую применимость данного подхода.

Дальнейшее развитие получило направление RL-TSK, в котором вместо простых констант в выводе правил используются аффинные функции TSK-модели. Такие гибридные модели разрабатывались для самых разных сценариев: от управления роботизированными манипуляторами до адаптации параметров в системах поддержки решений [Yan et al., 2001], а также для стратегий навигации и оптимизации ресурсов в динамических приложениях [Kumar, 2020].

В задачах управления и обучения с подкреплением одной из ключевых проблем является трудность интерпретации поведения агента. Стратегии, выработанные в процессе многократного взаимодействия со средой и оптимизации сложных критериев, обычно представляют собой черные ящики, что затрудняет их анализ и применение в критичных областях. Именно поэтому сейчас развивается направление объяснимого RL (Explainable Reinforcement Learning, XRL), фокусирующееся на создании моделей, обеспечивающих понятное объяснение принимаемых решений.

Одним из подходов в плане объяснимости в XRL является применение нечетких TSK-моделей. Благодаря возможности представлять правила в форме «если..., то...», они обладают хорошей наглядностью, а система типа универсального аппроксиматора позволяет приближать любые функции. В контексте RL это означает, что функция полезности агента $Q(s, a)$ может аппроксимироваться TSK-моделью, где каждое правило задает линейную или аффинную функцию входных переменных, а итоговой оценкой служит взвешенное по степеням истинности объединение результатов всех правил:

$$Q(s, a) = TSK(s, a), \quad (1)$$

где s – состояние среды, a – действие агента.

Гибридная модель продемонстрировала свою работоспособность как в дискретных средах, например, в клеточных автоматах, так и в непрерывных задачах, таких как балансировка маятника, управление посадкой лунного модуля (Lunar Lander) [Barto et al., 1983].

Обучение в этой модели строится по тому же принципу, что и в классический Q-learning: текущая оценка полезности корректируется с учетом разности во времени (temporal difference, TD). Формально обновление параметров происходит по правилу

$$Q(s, a) = Q(s, a) + \alpha(R(s') + \gamma Q(s', a) - Q(s, a)),$$

где s – текущее состояние среды, a – действие, выбранное в состоянии s , s' – следующее состояние среды после выполнения действия a , $R(s')$ – вознаграждение, полученное за переход в состояние s' , $Q(s, a)$ – текущее значение Q -функции для пары (s, a) , $Q(s', a)$ – оценка Q -функции для следующего состояния и действия, α – скорость обучения (learning rate), $0 < \alpha \leq 1$, γ – коэффициент дисконтирования будущих вознаграждений (discount factor), $0 \leq \gamma \leq 1$

Параметры нечеткой модели можно оптимизировать с помощью модифицированной версии градиентного спуска:

$$w_i = w_i + \alpha(R(s') + \gamma Q(s', a) - Q(s, a)) \frac{\delta Q}{\delta w_i}, \quad (2)$$

где w_i – параметры TSK-модели.

Подставляя выражение (1), получаем:

$$w_i = w_i + \alpha(R(s') + \gamma TSK(s', a) - TSK(s, a)) \frac{\delta TSK(s, a)}{\delta w_i}. \quad (3)$$

Аналогично классическим алгоритмам RL, устойчивость и эффективность обучения в нечетких алгоритмах RL могут быть значительно повышены за счет применения таких приемов, как повторное использование накопленного опыта (experience replay) и разделение функций выбора действий и оценки ценности в рамках моделей актор–критик.

Теоретическая обоснованность данного подхода базируется на известных результатах о сходимости Q-обучения при выполнении стандартных условий марковского процесса принятия решений (MDP) [Jaakkola et al., 1993]. Более того, как нейросети с различными функциями активации, так и нечеткие модели типа TSK доказали свою способность к универсальной аппроксимации. Благодаря этому TSK-модели могут эффективно заменить традиционные нейросетевые блоки в структуре DQN, сохраняя при этом фундаментальные свойства сходимости и точности аппроксимации.

Следовательно, нечеткая TSK-модель способна выполнять роль аппроксиматора функции полезности $Q(s, a)$, что подтверждает применимость алгоритма Q-обучения в сочетании с данной моделью. Это создает прочную теоретическую основу для дальнейших исследований в направлении адаптивных нечетких моделей в контексте обучения с подкреплением.

Модель (система) ANFIS, как отмечено ранее, является развитием концепции TSK-моделей и позволяет не просто задавать параметры правил вручную, а настраивать их автоматически с использованием алгоритма градиентного спуска. В отличие от подходов с жестко заданными параметрами условий (антецедентов) и следствий (консеквентов), ANFIS обеспечивает гибкую и адаптивную настройку, а также поддерживает XRL, что делает ее особенно подходящей для динамичных обучающих сред.

Для реализации заданной цели используется система ANFIS, интегрированная с нейросетевым компонентом. На первом этапе нейронная сеть извлекает информативные признаки из входных данных, после чего они преобразуются в степени принадлежности с помощью фаззификационного слоя. Далее выходы этого слоя комбинируются с условиями (антецедентами) нечетких правил путем их попарного перемножения. Полученные значения масштабируются с помощью обучаемых параметров, после чего агрегируются для формирования заключений (консеквентов) системы.

Когда требуется предсказывать несколько выходных значений одновременно (multi-output), применяется модифицированный вариант ANFIS – MANFIS (Multioutput ANFIS). В этой версии для каждого компонента выходного вектора формируется собственный набор нечетких правил, в то время как механизм дефаззификации остается идентичным классической реализации ANFIS. Схема системы ANFIS [Jang, 1993] представлена на рис. 1.

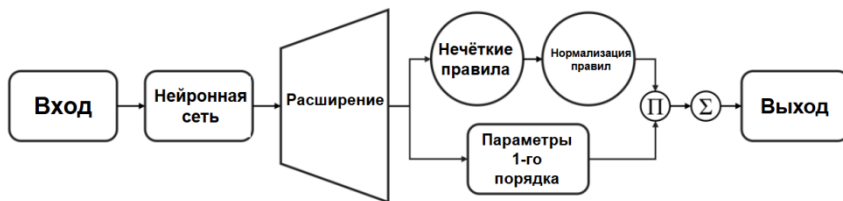


Рис. 1. Схема ANFIS

В данной модели (системе) используется нейронная сеть обобщенной структуры, параметры и глубина которой подбираются в зависимости от конкретной задачи. После первичной обработки входных данных нейросетью, результирующий выход трансформируется таким образом, чтобы его размер соответствовал числу заданных нечетких правил. Затем они подаются на модуль нечеткого вывода, где для каждого правила вычисляется степень его активации (firing strength), после чего значения нормализуются отдельно по каждому выходному каналу.

В случае применения ANFIS первого порядка, к каждой переменной применяется линейная функция – входное значение масштабируется с помощью обучаемого коэффициента и корректируется сдвигом (bias). Для ANFIS нулевого порядка используется упрощенная схема, в которой входной сигнал напрямую передается на следующий этап обработки, без преобразований. На заключительной стадии нормализованные коэффициенты срабатывания умножаются на соответствующие выходы от системы правил. Далее все полученные значения агрегируются путем суммирования, обеспечивая формирование итогового выхода модели с требуемыми размерностями.

2. Тестирование модели и сравнение результатов

В данном разделе приводится описание тестовой задачи CartPole, результаты моделирования и сравнение с моделями, разработанными в предыдущих работах.

2.1. Описание тестовой задачи CartPole

В задаче CartPole, называемой также задачей перевернутого маятника (inverted pendulum) (см. рис. 2) агенту необходимо управлять тележкой с прикрепленным к ней вертикальным шестом (маятником). Входной вектор среды состоит из четырех параметров: линейной скорости тележки; позиции тележки; угла отклонения шеста от вертикали; угловой скорости шеста. Действие, которое должен выбрать агент на каждом шаге, сводится к одному из двух вариантов – переместить тележку влево либо вправо.

Целью обучения является максимизация кумулятивной функции полезности для каждого состояния, что в идеале позволяет агенту накапливать наибольшее возможное вознаграждение – до 500 очков. На каждом временном шаге агент получает вознаграждение, равное 1, если система (тележка и шест) остается в допустимых пределах по всем параметрам. Таким образом, достижение общего вознаграждения в 500 свидетельствует о том, что агент успешно удерживал систему в равновесии в течение 500 последовательных шагов.

Принятие решений агентом основывается на выборе действия, соответствующего максимальному значению оценочной функции Q в текущем состоянии.

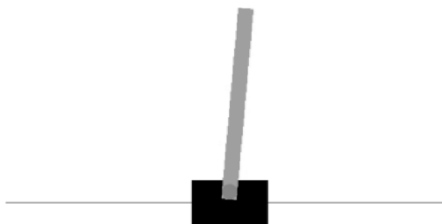


Рис. 2. Иллюстрация CartPole

2.2. Описание разработанной нечеткой модели

Нечеткие множества (довольно простые, но для решения данной задачи они подходят) для модельной задачи CartPole представлены на рис. 3.

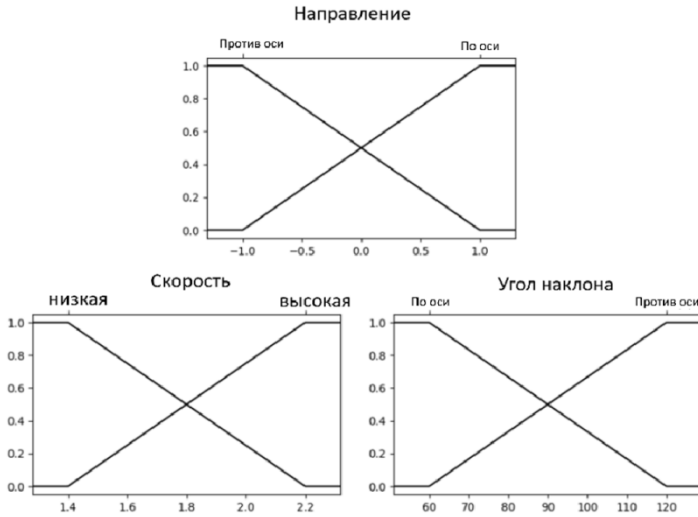


Рис. 3. Нечеткие множества для задачи CartPole

Нечеткие правила задаются следующим образом:

1. «Если направление против оси и угол наклона по оси и скорость высокая; то осевая тяга должна быть сильная, направление по оси»;
2. «Если направление против оси и угол наклона по оси и скорость низкая; то осевая тяга должна быть среднее, против оси»;
3. «Если направление тяги против оси и угол скорости против оси и скорость низкая; то осевая тяга должна быть слабая, направление против оси»;
4. «Если направление тяги против оси и угол абсолютной скорости против оси, и скорость высокая; то осевая тяга должна быть без изменений (0)»;
5. «Если направление тяги по оси и угол абсолютной скорости по оси и скорость высокая; то осевая тяга должна быть без изменений (0)»;
6. «Если направление тяги по оси и угол абсолютной скорости по оси, и скорость низкая; то осевая тяга должна быть слабая, по оси»;
7. «Если направление тяги по оси и угол абсолютной скорости против оси, и скорость низкая; то осевая тяга должна быть средняя, по оси»;

8. «Если направление тяги по оси и угол абсолютной скорости против оси, и скорость высокая; то осевая тяга должна быть сильная, по оси».

Непротиворечивость набора (базы) правил проверяется посредством выявления попарных пересечений antecedентов правил с несовпадающими консеквентами; установлено, что данный набор непротиворечив. Примеры формальной записи правил даны в работе [Еремеев и др., 2023].

Структура разработанной гибридной модели (системы) ANFIS-DQN представлена на рис. 4.

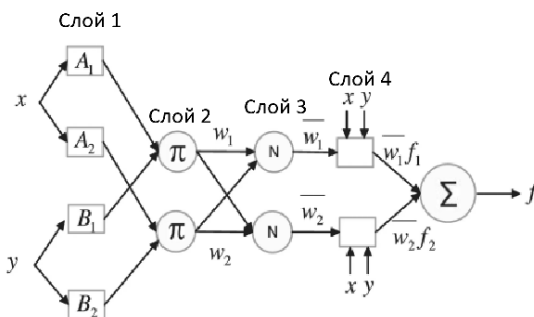


Рис. 4. Структура ANFIS-DQN

2.3. Результаты компьютерного моделирования и сравнительный анализ

Разработанная модель (система) ANFIS-DQN сравнивалась со стандартной моделью DQN, результаты приведены на рис. 5. Обе модели построены с применением механизма мягкого обновления целевой сети DQN. Количество обучаемых параметров в моделях сопоставимо: 10 290 у стандартной DQN и 10 293 в случае ANFIS-DQN. В целях обеспечения корректности сравнения обе модели обучались в идентичных условиях, с одинаковыми параметрами:

- скорость обучения (α) 0.001;
- коэффициент дисконтирования (γ) 0.99;
- размер батча для Experience Replay 128;
- число итераций 10 000.

Для оценки результатов моделей DQN и ANFIS-DQN (на рис. 5 обозначены, соответственно, как dqn и anfis) измеряется средняя награда по 10 тестовым средам каждые 2000 итераций. Цель – чтобы агент достигал средней награды 500 за 10 эпизодов.

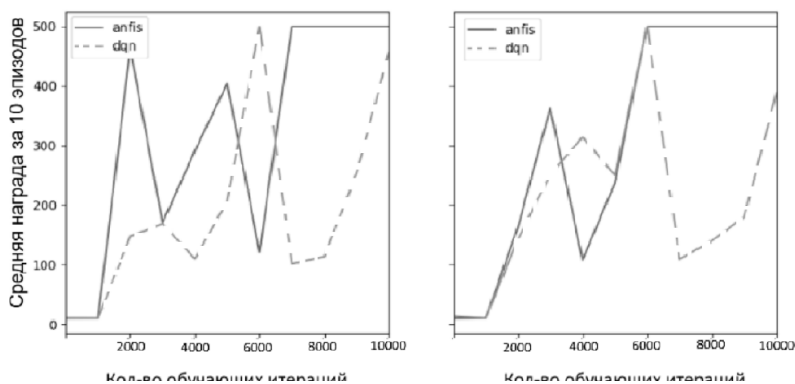


Рис. 5. Сравнение ANFIS-DQN и DQN

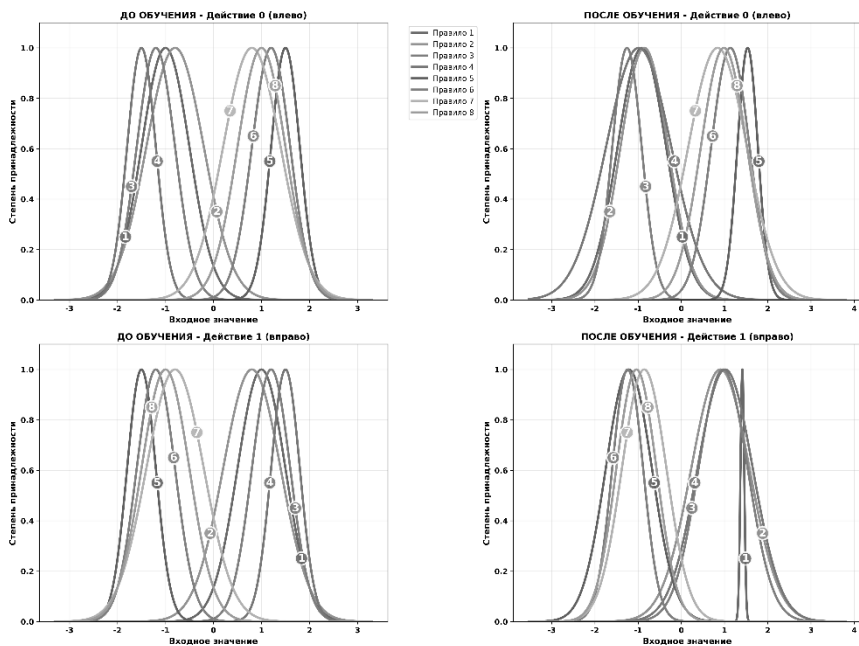


Рис. 6. Сравнение нечетки правил до и после обучения для перемещения тележки влево и вправо в среде CartPole

Из рис. 5 видно, что обе модели способны обучиться в среде, причем ANFIS-DQN обучается быстрее, чем DQN (изломы на кривых объясняются немонотонностью до определенного числа итераций соответствующих

функций оценки). Время обучения DQN – 45 секунд ANFIS-DQN – 164 секунды. По результатам моделирования можно заключить, что модель ANFIS-DQN может быть использована в ИС/ИСППР РВ. На рис. 6 приведены графики функций принадлежности нечетких правил. Видно, что некоторые кривые сместились влево или вправо. Это объясняется тем, что движение тележки в сторону, противоположную падению шеста, помогает скорректировать его положение. В работе [Еремеев и др., 2023] приведен пример тестирования на задаче «горный автомобиль» (mountain car), также показавший эффективность предложенного подхода. В дальнейшем планируется проведение экспериментов для более сложных сред (LunarLander, MuJoCo).

Заключение

В рамках данного исследования была разработана модель (система) ANFIS-DQN на основе DQN-модели обучения с подкреплением, дополненный адаптивной нейро-нечеткой системой вывода ANFIS, поддерживающая объяснимое RL. Разработанная модель (система) ANFIS-DQN прошла тестирование на задаче CartPole и продемонстрировала способность к успешному обучению. Выполнено сравнение ANFIS-DQN с классической реализацией DQN. Результаты показали, что нейро-нечеткая модель ANFIS-DQN обладает лучшей стабильностью и более быстрой сходимостью. В отличие от стандартных нейросетевых моделей, которые функционируют как «черный ящик», модель на базе нечетких правил обеспечивает интерпретируемость поведения как до, так и после обучения. Выполненная работа является частью комплексной задачи по созданию интегрированной среды для разработки перспективных ИС/ИСППР РВ.

Список литературы

- [Борисов и др., 2012] Борисов В.В., Круглов В.В., Федулов А.С. Нечеткие модели и сети. – 2-е изд., стереотип. – М.: Горячая линия-Телеком, 2012. – 284 с. – ISBN 978-5-9912-0283-1.
- [Еремеев и др., 2023] Еремеев А.П., Сергеев М.Д., Петров В.С. Интеграция методов обучения с подкреплением и нечеткой логики для интеллектуальных систем реального времени // Программные продукты и системы. – 2023. – № 4. – С. 600-606. – DOI: 10.15827/0236-235X.144.600-606.
- [Нечеткие гибридные системы, 2007] Батыршин И.З., Недосекин А.О., Стецко А.А., Тарасов В.Б. Нечеткие гибридные системы. Теория и практика: учебное пособие / под ред. Н.Г. Ярушкиной. – М.: Физматлит, 2007. – 208 с. – ISBN 978-5-9221-0786-0.
- [Barto et al., 1983] Barto A.G., Sutton R.S., Anderson C.W. Neuronlike adaptive elements that can solve difficult learning control problems // IEEE Transactions on Systems, Man, and Cybernetics. – 1983. – Vol. 13(5). – P. 834-846. – doi: 10.1109/TSMC.1983.6313077.

- [Berenji, 1992] Berenji H.R. A reinforcement learning–based architecture for fuzzy logic control // *International Journal of Approximate Reasoning*. – 1992. – Vol. 6(3). – P. 267-292. – doi: 10.1016/0888-613X(92)90021-C
- [Borys et al., 2023] Borys K., Schmitt Y.A., Nauta M., Seifert C., Krämer N., Friedrich C.M., Nensa F. Explainable AI in medical imaging: an overview for clinical practitioners – beyond saliency-based XAI approaches // *European Journal of Radiology*. – 2023. – Vol. 162(11). – P. 780-786. – doi:10.1016/j.ejrad.2023.110786.
- [Er et al., 2004] Er M.J., Deng C. Online tuning of fuzzy inference systems using dynamic fuzzy Q-learning // *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics*. – 2004. – Vol. 34(3). – P. 1478-1489. – doi: 10.1109/TSMCB.2004.826353.
- [Eremeev et al., 2024] Eremeev A.P., Sergeev M.D., Petrov V.S. Integrating reinforcement learning methods with neural networks and fuzzy logic in real-time intelligent systems // In: 2024 7th International Conference on Information Technologies in Engineering Education (Inforino), Moscow, Russian Federation, 2024. – P. 1-4. – doi: 10.1109/Inforino60363.2024.10551979.
- [Jaakkola et al., 1993] Jaakkola T., Jordan M., Singh S. Convergence of stochastic iterative dynamic programming algorithms // In: *Advances in Neural Information Processing Systems*. – 1993. – Vol. 6(1). – P. 1-8. – doi: 10.5555/297645.297652.
- [Jamshidi et al., 2015] Jamshidi P., Sharifloo A.M., Pahl C., Metzger A., Estrada G. Self-learning cloud controllers: fuzzy Q-learning for knowledge evolution // In: 2015 International Conference on Cloud and Autonomic Computing. – IEEE, 2015. – P. 208-211. – doi: 10.1109/ICCAC.2015.38.
- [Jang, 1993] Jang J.S.R. ANFIS: adaptive-network-based fuzzy inference system // *IEEE Transactions on Systems, Man, and Cybernetics*. – 1993. – Vol. 23(3). – P. 665-685. – doi: 10.1109/21.256541.
- [Kumar, 2020] Kumar S. Learning of Takagi-Sugeno fuzzy systems using temporal difference methods // Redmond: DigiPen Institute of Technology. – 2020. – 65 p.
- [Mencar et al., 2019] Mencar C., Alonso J.M. Paving the way to explainable artificial intelligence with fuzzy modeling. In: *Fuzzy Logic and Applications*. – Springer, 2019. – P. 215-227. – doi: 10.1007/978-3-030-19591-5_19.
- [Mnih et al., 2015] Mnih V., Kavukcuoglu K., Silver D., Rusu A.A., Veness J., Bellemare M.G., Graves A., Riedmiller M., Fidjeland A.K., Ostrovski G., Petersen S., Beattie C., Sadik A., Antonoglou I., King H., Kumaran D., Wierstra D., Legg S., Hassabis D. Human-level control through deep reinforcement learning // *Nature*. – 2015. – Vol. 518(7540). – P. 529-533. – doi: 10.1038/nature14236.
- [Ouyang et al., 2022] Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C.L., Mishkin P., Zhang C., Agarwal S., Slama K., Ray A., Schulman J., Hilton J., Kelton F., Miller L., Simens M., Askell A., Welinder P., Christiano P.F., Leike J., Lowe R.J. Training language models to follow instructions with human feedback // *Advances in Neural Information Processing Systems*. – 2022. – Vol. 35. – P. 27730-27744. – doi: 10.48550/arXiv.2203.02155.
- [Precup et al., 2011] Precup R.E., Hellendoorn H. A survey on industrial applications of fuzzy control // *Computers in Industry*. – 2011. – Vol. 62(3). – P. 213-226. – doi: 10.1016/j.compind.2010.10.001.

- [Reiners et al., 2021] Reiners D., Davahli M.R., Karwowski W., Cruz-Neira C. The combination of artificial intelligence and extended reality: a systematic review // *Frontiers in Virtual Reality*. – 2021. – Vol. 2(1). – P. 1-24. – doi: 10.3389/frvir.2021.72193.
- [Russell et al., 2010] Russell S.J., Norvig P. *Artificial Intelligence: A Modern Approach*. – M.: Pearson Education, Inc., 2010.
- [Viana et al., 2023] Viana J., Ralescu S., Cohen K., Ralescu A., Kreinovich V. Extension to multidimensional problems of a fuzzy-based explainable and noise-resilient algorithm // In: *Decision Making Under Uncertainty and Constraints: A Why-Book*. – Springer, 2023. – P. 289-296. – doi: 10.1007/978-3-031-27402-2_26.
- [Walker et al., 2019] Walker O., Vanegas F., Gonzalez F., Koenig S. A deep reinforcement learning framework for UAV navigation in indoor environments // In: *2019 IEEE Aerospace Conference*. – IEEE, 2019. – P. 1-14. – doi: 10.1109/AERO.2019.8742226.
- [Yan et al., 2001] Yan X., Deng Z., Sun Z. Competitive Takagi-Sugeno fuzzy reinforcement learning // In: *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01)* (Cat. No. 01CH37204). – IEEE, 2001. – P. 878-883. doi: 10.1109/CCA.2001.948137
- [Zhao et al., 2022] Zhao E., Yan R., Li J., Li K., Xing J. AlphaHoldem: High-performance artificial intelligence for heads-up no-limit poker via end-to-end reinforcement learning // In: *Proceedings of the AAAI Conference on Artificial Intelligence*. – AAAI, 2022. – Vol. 36(4). – P. 210-218. – doi: 10.1609/aaai.v36i4.20394.

УДК 004.89

doi: 10.15622/rcai.2025.070

НЕЙРОСЕТЕВАЯ МОДЕЛЬ ТЕМПОРАЛЬНОГО ОБЪЕДИНЕНИЯ КАДРОВ ВИДЕОЛЕКЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РЕКОНСТРУКЦИИ ИЗОБРАЖЕНИЯ

М.Е. Исмагулов (*m_ismagulov@ugrasu.ru*)^A

А.В. Мельников (*melnikovav@uriit.ru*)^{A,B}

^A Югорский государственный университет, Ханты-Мансийск

^B Югорский научно-исследовательский институт
информационных технологий, Ханты-Мансийск

В связи с ростом популярности формата видеолекций, возникают задачи мультимодальной обработки видеолекции в процессе которой можно получить конспект лекции или краткое содержание видеоматериала. При мультимодальной обработке видеолекции извлечение данных из видеоряда осложняется перекрытием контента лектором. Для решения этой задачи существует метод темпорального объединения кадров, данный метод широко применяется для реконструкции изображений в случаях движущихся объектов в кадре. Цель исследования разработка нейросетевой модели темпорального объединения видеокадров для восстановления областей доски. Отражен процесс разработки и обучения нейросетевой модели, в качестве метода обучения выбрано обучение с учителем. Основой датасета выбраны кадры видеолекции разбитые на сэмплы с эталонными изображениями доски. Модель построена на основе гибридной архитектуры, сочетающей сверточную нейронную сеть и рекуррентный слой с долгой краткосрочной памятью (LSTM). По результатам обучения были получены значения метрик accuracy: 0.7711, loss: 0.0773, метрика PSNR достигла 35 децибел, что является хорошим показателем восстановления изображений.

Ключевые слова: темпоральное объединение видеокадров, сверточные нейронные сети, рекуррентные нейронные сети, обучение нейронной сети, реконструкция изображения, обработка кадров видеолекции.

Введение

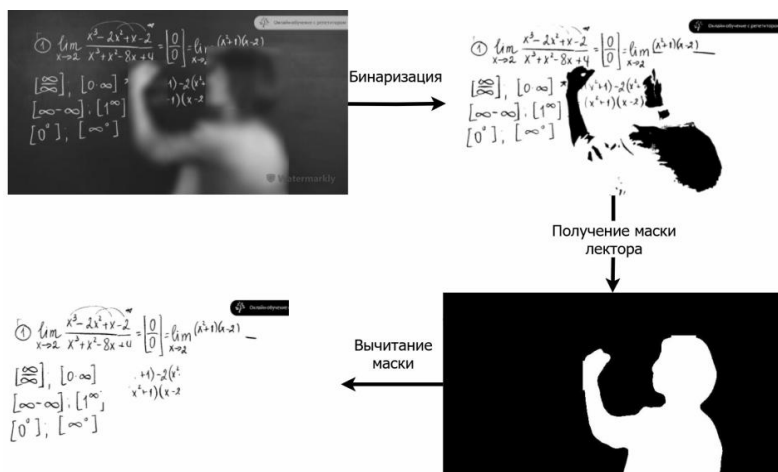
В настоящее время формат видеолекций приобретает большую популярность и активно развивается. Это проявляется в росте числа массовых открытых онлайн-курсов (МООК), вебинаров, а также в распространении практики видеозаписи очных лекций. Помимо роста популярности, увеличивается и объем такого контента. В связи с этим все острее ощущается потребность в сервисах, автоматически преобразующих видеолекции в текстовый формат для создания конспектов или аннотированных документов. В данной работе, как часть решения этой комплексной задачи, рассматривается алгоритм темпорального объединения кадров, направленный на реконструкцию изображения.

В процессе мультимодальной обработки видеолекции с лектором и доской, необходимо извлекать информацию с каждой из модальностей, как из аудиодорожки, так и из видеоряда, в процессе извлечения информации существует проблема того, что часть контента доски скрывается за лектором, и периодически открывается при перемещении лектора [Wang et al., 2022]. Получить единое представление доски без пробелов в автоматическом режиме становится нетривиальной задачей [Urala Kota et al., 2019]. Единое представление доски необходимо для последующего оптического распознавания символов (далее OCR), и получения текстового содержания видеомодальности. Если в изображениях, подаваемых на модель OCR, будут пробелы или незаполненные участки, то информация будет не полной и иметь ошибки. Таким образом, получить точное содержание видеомодальности становится труднодостижимым.

Для решения подобных проблем можно использовать методы темпорального объединения кадров видео [Urala Kota et al., 2018]. Темпоральное объединение кадров — это метод обработки видео или изображений, при котором кадры, близкие во времени (например, в пределах одной сцены), объединяются на основе их временной последовательности, чтобы восстановить, усилить или дополнить информацию, недоступную на отдельных кадрах [Zhang et al., 2024].

Идея заключается в следующем, если исключить (абстрагировать) лектора из кадра путем вычитания маски и оставить только надписи на доске, то получится область, в которой содержимое доски отсутствует, подробнее можно ознакомиться если обратиться к рис. 1. Абстрагирование необходимо для сосредоточения внимания на содержимом видеолекции, в данном случае, на содержимом доски [Kumar et al., 2020], [Исмагулов, 2024]. Получив очищенное, полное представление изображения доски, в дальнейшем можно передать изображение на модель OCR в результате обработки которой, можно получить текст, математические выражения в формате LaTeX или MathML. Объединив текстовое содержание видеомодальности и текстовое содержание аудиомодальности можно получить

документ, отражающий полное содержание видеолекции. Вне зависимости, от содержащихся на доске видов информации, таких как – текст, формулы, графики и т.д., задача модели состоит в корректном объединении кадров на уровне сцен, где каждая сцена отражает содержание доски на определенном этапе видеолекции. Корректность заключается в том, что на полученных изображениях доски отсутствуют артефакты, мешающие применению OCR, а также в заполнении пустых участков изображения.



Источник: видео TutorOnline (<https://youtu.be/19TSR9rrwxQ>), авторские права принадлежат TutorOnline

Рис. 1. Схема процесса абстрагирования лектора

Как видно из рис. 1, на последнем этапе обработки в области доски образуется пустая область, в процессе воспроизведения ролика лектор перемещается в кадре открывая разные области. За счет открывающихся областей, возможна реконструкция изображения доски в полном объеме. Для объединения кадров с разными открытыми областями доски и применяется метод темпорального объединения видеок кадров.

Построение датасета

В качестве основы датасета были выбраны несколько видеолекций по математике, с помощью библиотеки FFMPEG и языка Python из видеолекций были извлечены кадры, затем к каждому кадру была применена функция бинаризации изображения из библиотеки OpenCV.

Далее кадры были распределены по видеолекциям, для каждой видеолекции вручную производилась сортировка по сценам, где каждая сцена, это изменение положение кадра. Например, зуммирование, изменение ракурса камеры и т.д. Затем сцены были разбиты на сэмплы по 15 кадров,

для каждого сэмпла было подготовлено изображение истинности для реализации метода обучения с учителем [Feichtenhofer et al., 2018], [Wang et al., 2017]. Размер сэмпла в 15 кадров был выбран экспериментально для избежания ошибки out of memory, поскольку входной тензор формируется из кадров 1920 на 1080 пикселей. Подготовка изображения истинности была выполнена вручную, путем наложения бинаризованных кадров друг на друга в редакторе изображений GIMP 2.

Для обучения также были использованы маски лектора, полученные в результате обработки видеолекции моделью YOLO11 Small, для указания областей, которые необходимо восстанавливать. Структура датасета представлена на рис. 2, здесь представлена схема, отражающая иерархию каталогов датасета.

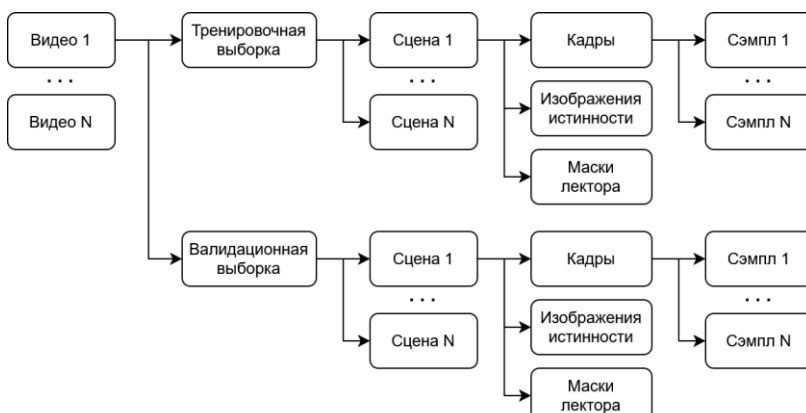


Рис. 2. Схема иерархии каталогов датасета

Помимо структуры датасета необходимо определить количественные и качественные характеристики, датасет состоит из 5 видео в каждом из которых 80 тренировочных сэмплов по 15 кадров и 16 валидационных сэмплов также по 15 кадров, если в сцене не хватает 15 кадров, то недостаточность компенсируется дополняющими изображениями (padding images), пустыми по содержанию. Разрешение изображений в датасете составляет 1920 на 1080 пикселей (исходное разрешение видеолекции). В итоге для каждого видео насчитывается по 1300 кадров.

Разработка архитектуры модели, обучение и инференс модели

Модель написана на языке Python в качестве библиотеки машинного обучения была использована Tensorflow-Keras. Для работы с векторами многомерными матрицами была использована библиотека NumPy. Для визуализации полученных данных использовалась библиотека Matplotlib.

Архитектура модели представляет собой нейронную сеть, сочетающую сверточные и рекуррентные слои с долгой краткосрочной памятью (LSTM). Применение LSTM-сетей обосновано тем, что они хорошо справляются с темпоральной обработкой данных [O'Donncha et al., 2022], [Pham, 2021].

Формально темпоральное объединение кадров можно определить следующим образом, пусть существует:

- $\{F_0, F_1, \dots, F_n\}, F_t \in \{0, 1\}^{H \times W}$ – последовательность бинаризованных кадров от 0 до n , и каждый элемент этой последовательности, кадр F_t это бинарная матрица размером $H \times W$, т.е. кадр из чёрно-белых пикселей.
- $\{M_0, M_1, \dots, M_n\}, M_t \in \{0, 1\}^{H \times W}$ – последовательность бинаризованных масок от 0 до n , и каждый элемент этой последовательности, кадр M_t это бинарная матрица размером $H \times W$, т.е. маска из чёрно-белых пикселей.

Тогда:

$$F_{\text{объединенное}}(x, y) = \max_{t \in \{0, \dots, n\}} \{F_t(x, y) | M_t(x, y) = 0\},$$

где $x \in [0, W)$, $y \in [0, H)$ – координаты пикселя в кадре.

Для обработки пространственных данных активно применяются сверточные нейронные сети, поскольку они хорошо справляются с этим классом задач [Li et al., 2020], [Melnikov et al., 2017]. Алгоритм объединения имеет особенность в виде двухступенчатого способа объединения кадров, то есть, на первом этапе кадры объединяются на уровне сэмплов, а на втором этапе на уровне сцен видеолекции, архитектура модели отражена на рис. 3. Алгоритм двухступенчатого объединения также был выбран для избежания ошибки out of memory.

Основой архитектуры выступает слой ConvLSTM2D с 16 фильтрами и ядром размером (3, 3), данный слой объединяет возможности сверточных нейронных сетей и LSTM-сетей. Для введения нелинейности, применяется функция активации relu.

За этим слоем следует Dropout с коэффициентом 0.3, который, случайным образом, обнуляет часть выходных данных во время обучения, для предотвращения переобучения, и улучшения обобщающей способности модели.

Финальный этап обработки осуществляется с помощью слоя Conv2D, содержащего один фильтр с ядром (3, 3). Этот слой выполняет заключительную свертку для предсказания маски, сохраняя размерность благодаря параметру padding='same'. Активационная функция sigmoid нормализу-

ет выходные значения в диапазон от 0 до 1. Выходной слой использует тип данных float32. Модель компилируется (обучается) с оптимизатором Adam (скорость обучения 0.0001, clipnorm=1.0 для стабилизации градиентов), функцией потерь binary_crossentropy и метриками accuracy.

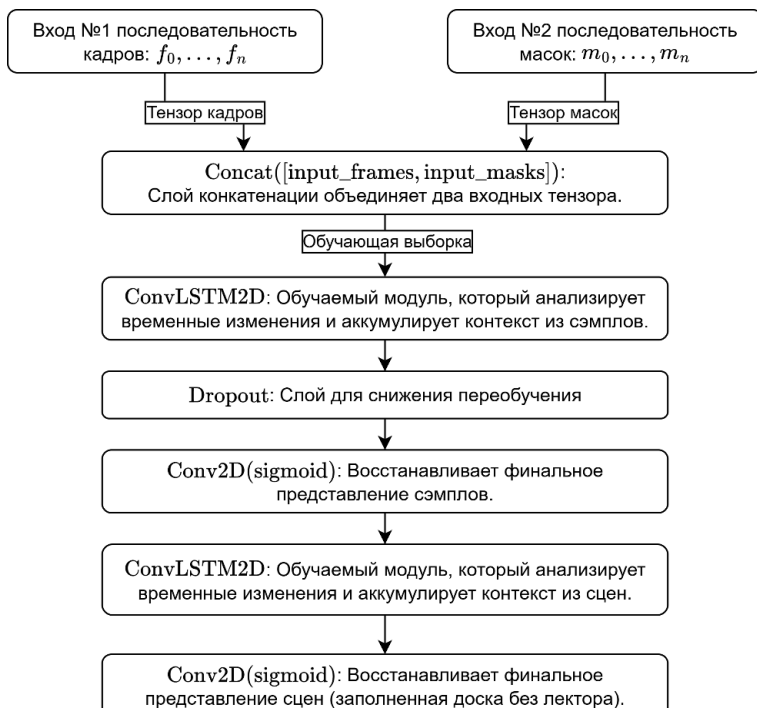


Рис. 3. Архитектура нейронной сети

Для предотвращения переобучения, применяется функция EarlyStopping, которая, отслеживает валидационную потерю, и восстанавливает лучшие веса после 10 эпох без улучшения. Количество эпох обучения 50. Для ускорения инференса модели, использовалась библиотека ONNXRuntime, так как, эта библиотека имеет встроенный инструментарий оптимизации [Someki et al., 2022], [ONNX Runtime, 2025].

В качестве среды разработки была выбрана среда Kaggle Notebook основанная на Jupiter Notebook, модель обучалась на двух графических ускорителях серии NVidia Tesla T4 с 16 гигабайт GDDR6 видеопамати для каждого ускорителя, обучение производилось с использованием стратегии MirroredStrategy, для использования двух графических ускорителей.

По результатам обучения имеем следующие показатели по метрикам loss и ассигасу, лучшее значение loss 0.0773, лучшее значение ассигасу 0.7711 для тестовой выборки, лучшее значение loss 0.0787, лучшее значение ассигасу 0.773 для тестовой выборки. Более подробно с процессом обучения можно ознакомиться на рис. 4 и 5, отражающие графики обучения.

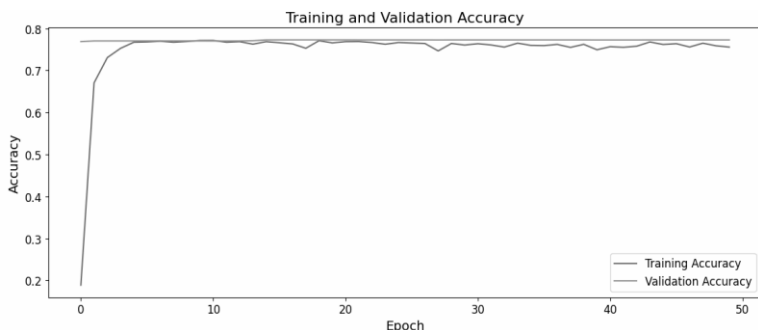


Рис. 4. График Ассигасу обучения модели



Рис. 5. График Loss обучения модели

Качественная метрика восстановления изображения PSNR

Для оценки качества реконструкции изображения, принято использовать меру пикового отношения сигнала к шуму (англ. peak signal-to-noise ratio или PSNR). Данная мера характеризует соотношение между максимумом возможного значения сигнала и мощностью шума, искажающего значения сигнала [Shen et al., 2024], [Keleş et al., 2021]. Измеряется в децибелах, и характеризует разницу оригинального изображения и изображения полученного в ходе генерации при инференсе модели.

Определяется по следующей формуле:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right),$$

где MAX_I – максимальное значение пикселя (например, 255 для 8-битных изображений);

MSE (*Mean Squared Error*) – среднеквадратичная ошибка;

MSE определяется по следующей формуле:

$$MSE = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n [I(i, j) - K(i, j)]^2,$$

где I – оригинальное изображение;

K – сгенерированное изображение;

m и n – размерность изображения.

Хорошим результатом считается если мера PSNR полученного изображения достигла диапазона от 32 до 40 децибел, если выше, то считается, что изображение не отличимо от оригинала [Huynh-Thu et al., 2008], [Lim et al., 2017]. В ходе эксперимента, мера PSNR достигла 35,011 децибел, что является хорошим показателем восстановления изображений. В контексте бинаризированных изображений, мера PSNR равная 35,011 характеризует практически полное совпадение положения пикселей [Корешев и др., 2020].

В качестве Baseline взята модель из статьи [Park et al., 2020], в ходе исследования, модель BVDNet достигла значения по мере PSNR 34,7055, что также является хорошим результатом. В исследовании стояла задача восстановления цветных кадров видео.

Если сравнивать изображения визуально (рис. 6), то можно сделать следующие выводы. При создании Ground Truth изображений, были пропущены некоторые части математических выражений (выделено сплошной рамкой), но модель восстановила не только пропущенные части, но и то, что изначально не было указано в эталонах. Также, на восстановленных изображениях, присутствуют артефакты оставшиеся от процесса вычитания маски, с той лишь разницей, что эти артефакты на оригинальных изображениях носят четкий характер с резкими границами, а в восстановленном изображении, эти артефакты размытые (выделено пунктирной рамкой). Также, из особенностей реконструкции, можно выделить многократное наложение водяного знака онлайн школы (выделено штриховой рамкой), и утолщение контуров надписей.

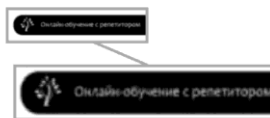
Изображение эталон

$$\textcircled{1} \lim_{x \rightarrow 2} \frac{x^3 - 2x^2 + x - 2}{x^3 + x^2 - 8x + 4} = \left[\frac{0}{0} \right] = \lim_{x \rightarrow 2} \frac{(x+1)(x-2)}{(x-2)(x^2 + 4x + 4)} =$$

$$\left[\frac{\infty}{\infty} \right], \left[0 \infty \right] \rightarrow x \cdot (x+1) - 2(x^2 + 4) =$$

$$\left[\infty - \infty \right], \left[1 \infty \right] \rightarrow x^2 + x - 8x + 4 =$$

$$\left[0^+ \right], \left[\infty^+ \right] \rightarrow \frac{x^2 - 7x + 4}{x^2 + 4x + 4} = \frac{x^2 - 7x + 4}{(x+2)^2}$$



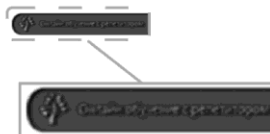
Реконструированное изображение

$$\textcircled{1} \lim_{x \rightarrow 2} \frac{x^3 - 2x^2 + x - 2}{x^3 + x^2 - 8x + 4} = \left[\frac{0}{0} \right] = \lim_{x \rightarrow 2} \frac{(x+1)(x-2)}{(x-2)(x^2 + 4x + 4)} =$$

$$\left[\frac{\infty}{\infty} \right], \left[0 \infty \right] \rightarrow x \cdot (x+1) - 2(x^2 + 4) =$$

$$\left[\infty - \infty \right], \left[1 \infty \right] \rightarrow x^2 + x - 8x + 4 =$$

$$\left[0^+ \right], \left[\infty^+ \right] \rightarrow \frac{x^2 - 7x + 4}{x^2 + 4x + 4} = \frac{x^2 - 7x + 4}{(x+2)^2}$$



Источник: видео TutorOnline (<https://youtu.be/19TSR9rwxQ>), авторские права принадлежат TutorOnline

Рис. 6. Результаты реконструкции изображения

Заключение

В данном исследовании, была разработана, и реализована модель темпорального объединения кадров видеолекций с целью реконструкции содержимого доски, скрываемого лектором. Модель является частью пайплайна агента, который восстанавливает полный образ доски в видеолекции, с последующей обработкой агентом оптического распознавания символов. Метрики обучения показали хорошую сходимость. Лучшее значение loss – 0.0773 (тренировочная выборка), 0.0787 (тестовая выборка). Лучшее значение ассигасу – 0.7711 (тренировочная выборка), 0.773 (тестовая выборка). Качество восстановления (PSNR = 35.011 дБ) ненамного превысило baseline-модель BVDNet (34.7055 дБ), что подтверждает эффективность предложенного подхода. Визуальный анализ показал, что модель не только восстанавливает пропущенные в Ground Truth элементы, но и корректно заполняет области, изначально скрытые лектором, несмотря на наличие незначительных артефактов. Для дальнейшего повышения метрик ассигасу, loss, планируется увеличение обучающей выборки, и применение аугментации данных. Для повышения меры PSNR, планируется улучшение алгоритма бинаризации и других алгоритмов предобработки данных.

Список литературы

- [Исмагулов, 2024] Исмагулов М.Е. Абстрагирование данных в контексте машинного обучения: методы, применения и перспективы // Информационные технологии и математическое моделирование: Труды XXIV Международной конференции ITMM-2024 (Томск, 2024). – Томск: Изд-во Томского гос. ун-та, 2024. – С. 605-607. – URL: https://www.researchgate.net/publication/391833448_1_Conference_proceedings_with_your_article_Ismagulov_M_E_Methods_and_Algorithms_for_Multimodal_Conversion_of_Video_Lectures (дата обращения: 17.05.2025).
- [Корешев и др., 2020] Корешев С.Н., Старовойтов С.О., Смородинов Д.С., Фролова М.А. Методы оценки качества изображений бинарных объектов, восстановленных с помощью синтезированных голограмм-проекторов // Науч.-техн. вестн. инф.-технол., мех. и оптики. – 2020. – Т. 20, № 3. – С. 327-334.
- [Feichtenhofer et al., 2018] Feichtenhofer C., Fan H., Malik J., He K. SlowFast Networks for Video Recognition [Электронный ресурс] // arXiv. – 2018. – arXiv:1812.03982. – URL: <https://arxiv.org/abs/1812.03982> (дата обращения: 18.05.2025). – DOI: 10.48550/arXiv.1812.03982.
- [Huynh-Thu Q., et al., 2008] Huynh-Thu Q., Ghanbari M. Scope of validity of PSNR in image/video quality assessment // Electronics Letters. – 2008. – Vol. 44, No. 13. – P. 800-801.
- [Keleş et al., 2021] Keleş O., Yılmaz M.A., Tekalp A.M., Korkmaz C., Dogan Z. On the Computation of PSNR for a Set of Images or Video [Электронный ресурс] // arXiv. – 2021. – arXiv:2104.14868. – URL: <https://arxiv.org/abs/2104.14868> (дата обращения: 19.05.2025). – DOI: 10.48550/arXiv.2104.14868.
- [Kumar, P., et al., 2020] Kumar, P., Ambati, R., & Raj, L. An Efficient Scene Content-Based Indexing and Retrieval on Video Lectures. // Advances in Intelligent Systems and Computing, vol 1171. – 2020. – P. 521-534.
- [Li, Z., et al., 2020] Li Z., Liu F., Yang W., Peng S., & Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects // IEEE Transactions on Neural Networks and Learning Systems. – 2020. – 33. – P. 6999-7019.
- [Lim B., et al., 2017] Lim B., Son S., Kim H., Nah S., Mu Lee K. Enhanced Deep Residual Networks for Single Image Super-Resolution // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). – 2017. – P. 136-144.
- [Melnikov et al., 2017] Melnikov A., Sochenkova A., Sochenkov I., Makovetskii A., Vokhmintsev A. Convolutional neural networks and face recognition task [Электронный ресурс] // Proc. SPIE Applications of Digital Image Processing XL. – 2017. – Vol. 10396. – P. 103962L. – URL: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10396/103962L/Convolutional-neural-networks-and-face-recognition-task/10.1117/12.2273624.short> (дата обращения: 18.05.2025).
- [ONNX Runtime, 2025] ONNX Runtime Documentation [Электронный ресурс]. – URL: <https://onnxruntime.ai/docs/> (дата обращения: 19.05.2025).
- [O'Donncha et al., 2022] O'Donncha F., Hu Y., Palmes P., Burke M., Filgueira R., Grant J. A spatio-temporal LSTM model to forecast across multiple temporal and spatial scales // Ecological Informatics. – 2022. – Vol. 69. – P. 101687. – ISSN 1574-9541. – URL: <https://www.sciencedirect.com/science/article/pii/S1574954122001376> (дата обращения: 19.05.2025). – DOI: 10.1016/j.ecoinf.2022.101687.

- [**Park et al., 2020**] Recurrent Temporal Aggregation Framework for Deep Video Inpainting // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2020. – Vol. 42, No. 10. – P. 2547-2561. – URL: https://joonyoung-cv.github.io/assets/paper/20_tpami_recurrent_temporal.pdf (дата обращения: 18.05.2025).
- [**Pham, 2021**] Pham T.D. Time–frequency time–space LSTM for robust classification of physiological signals // Scientific Reports. – 2021. – Vol. 11. – P. 6936. – URL: <https://www.nature.com/articles/s41598-021-86432-7> (дата обращения: 19.05.2025). – DOI: 10.1038/s41598-021-86432-7.
- [**Shen, W., et al., 2024**] Shen W., Tian X., Zeng D., & Zhang Y. Multi-scale image compression and reconstruction algorithm for structural health monitoring system // Engineering Structures. – 2024.
- [**Someki et al., 2022**] Someki M., Higuchi Y., Hayashi T., & Watanabe S. ESPnet-ONNX: Bridging a Gap Between Research and Production // Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). – 2022. – P. 420-427.
- [**Urala Kota et al., 2018**] Urala Kota B., Davila K., Stone A., Setlur S., Govindaraju V. Automated Detection of Handwritten Whiteboard Content in Lecture Videos for Summarization // Proceedings of the 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). – 2018. – P. 19-24. – DOI: 10.1109/ICFHR-2018.2018.00013.
- [**Urala Kota et al., 2019**] Urala Kota B., Davila K., Stone A., Setlur S., Govindaraju V. Generalized framework for summarization of fixed-camera lecture videos by detecting and binarizing handwritten content // International Journal on Document Analysis and Recognition (IJDAR). – 2019. – Vol. 22. – P. 221-233. – DOI: 10.1007/s10032-019-00327-y.
- [**Wang et al., 2017**] Wang X., Girshick R., Gupta A., He K. Non-local Neural Networks [Электронный ресурс] // arXiv. – 2017. – arXiv:1711.07971. – URL: <https://arxiv.org/abs/1711.07971> (дата обращения: 18.05.2025). – DOI: 10.48550/arXiv.1711.07971.
- [**Wang et al., 2022**] Wang T., He M., Shen K., Liu W., Tian C. Learned regularization for image reconstruction in sparse-view photoacoustic tomography // Biomedical Optics Express. – 2022. – Vol. 13, No. 11. – P. 5721-5737. – DOI: 10.1364/BOE.469460.
- [**Zhang et al., 2024**] Zhang, T., He, X., Teng, Q., Cheng, J., & Ren, C. Spatio-Temporal Adaptive Weighted Fusion Network for Compressed Video Quality Enhancement. // IEEE Transactions on Circuits and Systems II: Express Briefs. – 2024. – 71. – P. 5064-5068.

УДК 004.89

doi: 10.15622/rcai.2025.071

АВТОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ТЕЗАУРУСА: ОЦЕНКА МЕТОДОВ КЛАСТЕРИЗАЦИИ И ВЕКТОРНОГО ПРЕДСТАВЛЕНИЯ СЛОВ

И.А. Коломойцева (*bolatiger@mail.ru*)

О.И. Федяев (*olegfedyayev@mail.ru*)

Донецкий национальный технический университет, Донецк

В статье исследуются методы автоматизированного построения тезауруса для построения системы интеллектуального поиска информации «CodeLex», с акцентом на выявление синонимии между терминами в области программирования. Рассматривается применение алгоритмов кластеризации (K-means, DBSCAN) в сочетании с моделями векторного представления слов (Word2Vec, BERT) для определения отношения синонимии. Эксперименты проводились на фрагментах текстов из книги по языку программирования PHP. Результаты показали, что связка Word2Vec+K-means наиболее перспективна для определения синонимии между именами функций и их описаниями. BERT продемонстрировал низкую эффективность в данной задаче. Полученные результаты могут быть использованы для улучшения работы системы CodeLex.

Ключевые слова: информационный поиск, тезаурус, кластеризация, DBSCAN, K-means, Word2Vec, BERT.

Введение

Поиск учебных и научных материалов – актуальная задача для любого учебного заведения. Первое, что используется с этой целью как источник информации, – это библиотека, в которой в современных условиях хранятся не только бумажные источники, но и электронные. Ещё одним таким источником материалов может быть структурное подразделение учебного заведения, например, кафедра вуза. В кафедральных хранилищах могут быть представлены книги, монографии, учебные пособия, методические указания и другие материалы, созданные сотрудниками кафедры. Также в кафедральных хранилищах есть большое количество служебных документов. Искать информацию в этих источниках могут препода-

даватели, которым нужно подготовить лекцию, обновить некоторые факты и т.п. К этим же ресурсам могут обращаться и студенты, которым нужно найти ответы на вопросы, возникающие в процессе обучения. Поиск в таких локальных хранилищах с помощью разработанной в рамках проекта «Виртуальная кафедра» системы интеллектуального поиска информации в области программирования «CodeLex» является важной задачей.

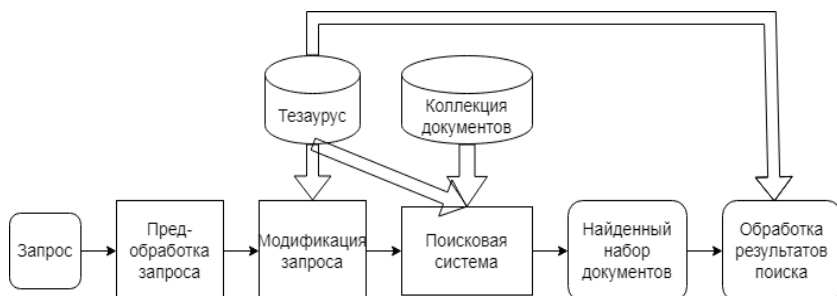


Рис. 1. Схема системы интеллектуального поиска информации «CodeLex»

В системе «CodeLex» интеллектуального поиска в документах образовательной организации (рис. 1), кроме типовых элементов, есть принципиально важный дополнительный – тезаурус, ориентированный на предметную область «Программирование».

Тезаурус – это словарь, в котором слова и словосочетания с близкими значениями сгруппированы в единицы, называемые понятиями, и в котором явно (в виде отношений, иерархии) указываются отношения между этими понятиями [Лукащевич, 2011].

Тезаурус участвует в трёх аспектах функционирования системы «CodeLex». Во-первых, с его помощью модифицируется (расширяется или уточняется) запрос пользователя. Это позволяет увеличить полноту, а в случае уточнения запроса, и точность поиска. Тезаурус также участвует в построении инвертированного индекса – важной составляющей «движка» поисковой системы. Элементами такого индекса выступают термины из тезауруса, а не полный набор терминов из коллекции документов, с которой работает поисковая система. Подобный подход ускорит поиск и сократит место в памяти для хранения индекса. Кроме того, тезаурус используется в алгоритме ранжирования документов. Благодаря тезаурусу, более высокий ранг имеют документы, включающие термины, связанные отношениями с терминами из запроса.

Сложность работы с тезаурусом в предметной области программирования заключается в том, что его надо сформировать, а затем постоянно поддерживать в актуальном состоянии.

Существуют два основных подхода к созданию тезаурусов: ручной и автоматизированный. Ручной подход – создание тезауруса экспертами-лингвистами или специалистами предметной области. Он трудоёмкий и дорогостоящий. Автоматизированный подход – использование методов обработки естественного языка (NLP), машинного обучения и алгоритмов кластеризации. Этот метод позволяет быстро обрабатывать большие объёмы данных, но нуждается в последующей ручной проверке и коррекции.

Целью данной статьи является исследование методов автоматизированного построения тезаурусов, в частности, использование алгоритмов кластеризации для определения отношения синонимии между терминами.

1. Анализ подходов к автоматизированному построению тезаурусов

Автоматический подход к созданию тезаурусов использует методы машинного обучения и обработки естественного языка (NLP) для автоматизации различных этапов создания. И так как размеченных текстов в области программирования недостаточно, то предпочтительными являются методы обучения без учителя.

Для автоматизированного построения тезауруса чаще всего используются методы:

- кластеризации;
- снижения размерности;
- векторного представления слов;
- тематического моделирования;
- ассоциативных правил.

Рассмотрим особенности применения этих методов для задачи формирования тезаурусов.

Кластеризация выполняет группировку терминов по семантической близости. Например, кластеры могут соответствовать категориям тезауруса (напр., «ООП», «Сетевое программирование2»). Кластеры также позволяют выявить группы связанных терминов. Иерархическая кластеризация дополнительно строит древовидную структуру, что соответствует иерархиям в тезаурусе (родо-видовые отношения).

Среди методов кластеризации для решения поставленной задачи подходящими методами представляются следующие:

- K-means;
- иерархическая кластеризация;
- DBSCAN (Density-Based Spatial Clustering of Applications with Noise).

K-means разбивает данные на K кластеров, где каждый элемент относится к кластеру с ближайшим средним (центроидом).

Иерархическая кластеризация строит иерархию кластеров, начиная с каждого элемента как отдельного кластера и постепенно объединяя их, или наоборот, начиная со всего набора данных как одного кластера и постепенно разделяя его.

DBSCAN находит кластеры, основанные на плотности точек данных, выделяя "шум" – точки, не принадлежащие ни к одному кластеру.

Перед кластеризацией или анализом, чтобы упростить данные или улучшить их визуализацию, применяют следующие методы снижения размерности: PCA, t-SNE, UMAP.

PCA может быть использован для предварительной обработки данных, чтобы снизить размерность и улучшить производительность алгоритмов кластеризации. Его основным недостатком является то, что он не гарантирует сохранения локальной структуры данных.

t-SNE подходит для визуализации тезауруса в 2D или 3D пространстве, позволяя оценить качество кластеризации. Его недостаток – может искажать глобальную структуру данных.

UMAP обеспечивает лучшее сохранение как локальной, так и глобальной структуры данных, что делает его хорошим выбором для снижения размерности перед кластеризацией.

Векторные представления слов используются для вычисления семантической близости терминов (обнаружения синонимов, ассоциации). При обучении без учителя используются модели Word2Vec, FastText, BERT.

Тематическое моделирование – это статистический метод, позволяющий извлечь скрытые темы из большой коллекции текстовых документов. Он основан на предположении, что каждый документ является смесью нескольких тем, и каждая тема состоит из распределения слов. Вместо ручного анализа текста, тематическое моделирование автоматически идентифицирует эти темы и определяет, как каждая тема представлена в каждом документе. Самыми популярными методами тематического моделирования являются методы: латентное размещение Дирихле (LDA), и латентный семантический анализ (LSA). LSA позволяет выявить в текстах синонимию и полисемию, и поэтому его важно исследовать в контексте построения тезауруса.

Ассоциативные правила могут выявлять отношения между словами на основе их совместной встречаемости в текстах. Эти правила могут обнаружить такие типы отношений, как родо-видовые, часть-целое, а также связанные понятия.

Методы, использующие ассоциативные правила, которые можно использовать в контексте построения тезауруса: Apriori, Eclat.

2. Методы кластеризации и векторного представления слов

В данной статье исследуется эффективность применения моделей построения векторного представления слов – Word2Vec и BERT и методов кластеризации – K-means и DBSCAN [Жилов, 2023].

Word2Vec для формирования векторного представления слов использует нейронные сети с малым числом слоёв. На вход нейросети подается большой текстовый корпус, в котором каждому слову сопоставляется вектор. После создания словаря вычисляется векторное представление слов, основанное на семантической близости. Мерой близости является косинусное сходство между двумя не нулевыми векторами [Попова, 2023].

Bidirectional Encoder Representations from Transformers (BERT) – нейронная сеть, основанная на энкодере трансформера. BERT используется для решения задач обработки естественно-языковых текстов, в том числе, и для построения векторного представления слов. Архитектура модели BERT организована в виде цепочки блоков. Первому блоку на вход подаётся цепочка эмбеддингов, а на выходе получается последовательность векторов той же длины [Новикова, 2024].

Метод K-means [Дюличева, 2021] является алгоритмом кластеризации, который разбивает данные на k кластеров, минимизируя расстояние внутри кластера. Он в теории не предназначен для поиска синонимов, но его можно использовать для определения близких по смыслу слов. Чтобы его использовать для поиска синонимов, необходимо создать векторное представление слов с помощью методов Word2Vec или FastText или получить контекстные эмбеддинги с помощью моделей трансформеров, например, BERT. Далее необходимо применить алгоритм K-means к полученным векторным представлениям слов. Каждый кластер, сформированный K-means, должен представлять собой группу слов, семантически близких друг к другу.

DBSCAN – это метод кластеризации, основанный на плотности данных [Фомичев, 2023]. При использовании не надо предварительно указывать количество кластеров. Метод DBSCAN может обнаруживать кластеры произвольной формы. Как и метод K-means, DBSCAN по умолчанию не предназначен для поиска синонимов, но при определенных условиях его можно использовать для этой цели. Как и для K-means, для DBSCAN потребуются векторные представления слов (или контекстные эмбеддинги), на основе которых DBSCAN и выдаст кластеры близких по смыслу слов. В отличие от K-means, для которого требуется задать количество кластеров, DBSCAN нужно задать радиус окрестности (eps), в котором будут искаться точки для формирования кластеров, и минимальное количество точек (min_samples), необходимое для формирования "плотного" кластера, чтобы снизить размерность и улучшить производительность алгоритмов.

3. Оценка эффективности Word2Vec и BERT в задачах кластеризации синонимов идентификаторов и их описаний

Структура тезауруса с области программирования, используемого в CodeLex, приведена в [Коломойцева, 2024]. Одной из важных особенностей этого тезауруса является то, что в качестве синонимов могут выступать имена функций (методов класса), классов, структур. Возможность автоматически расширить запрос за счёт добавления имени идентификатора увеличит точность поиска.

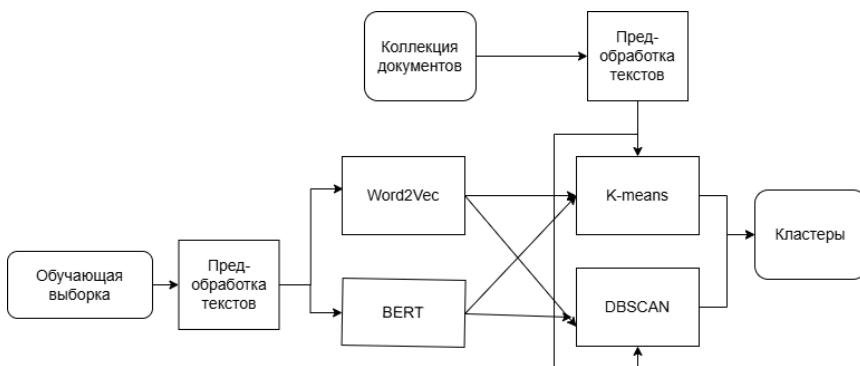


Рис. 2. Схема экспериментальной оценки эффективности задачи кластеризации синонимов

Синонимическую связь между названием идентификатора и термином, определяющим его назначение можно установить с помощью алгоритмов кластеризации. В качестве таких алгоритмов для исследования выбраны K-means и DBSCAN. Так как эти алгоритмы работают с числовым представлением информации, то потребовалось векторное представление слов. Для реализации этой задачи решено использовать модели Word2Vec и BERT.

Задачей исследования является определение эффективности использования связок моделей и методов Word2Vec+K-means, Word2Vec+DBSCAN, BERT+K-means, BERT+DBSCAN (рис. 2).

Эффективность измеряется по формуле

$$\Xi = \frac{K_H}{K_C},$$

где K_H – количество имён идентификаторов и их описаний, попавших в один кластер, т.е. являющихся синонимами, K_C – количество имён идентификаторов и их описаний, содержащихся в коллекции документов.

Эксперименты проводились с помощью программы на языке Python. Для предобработки текста использовалась библиотека NLTK, для лемматизации терминов – `rumorphy2`. Реализация методов K-means и DBSCAN взята из библиотеки `scikit-learn`. Работа с моделью Word2Vec выполняется с помощью `gensim`. С BERT работа выполняется с помощью `transformers`, и в качестве наиболее подходящей модели выбрана мультязычная модель `bert-base-multilingual-cased`, позволяющая работать с текстами на русском языке.

Выборкой, на которой обучались модели Word2Vec и BERT, были фрагменты текстов из книги по программированию на языке PHP, содержащие описание работы со строками и массивами.

Перед обучением моделей тексты прошли процесс предобработки, который включает следующие этапы:

- разбиение текста на токены;
- удаление неинформативных токенов;
- приведение слов на русском языке к нижнему регистру;
- лемматизация токенов.

К неинформативным токенам относятся стоп-слова (из набора стоп-слов для русского языка из библиотеки NLTK), числа (в том числе и числа с дополнительными символами, например, #7), аргументы функций, имена переменных. Также удалены слова общей лексики, которые не могут быть терминами из области программирования, например, «также», «существовать», «избегать». Такие «лишние» слова определены экспертным путём и содержатся в дополнительном словаре.

К нижнему регистру приводятся слова для обеспечения унификации. Применяется это только к словам на русском языке из-за того, что слова, написанные латиницей, могут быть именами идентификаторов, для которых регистр имеет смысл.

Эффективность методов проверялась на двух наборах текстов, каждый из которых содержал примерно по 3800 слов. Набор1 содержал описание 19 функций работы со строками, а набор2 – описание 12 функций работы с массивами.

Эксперимент выявил, что использование модели BERT для векторного представления слов, привело к тому, что все слова, записанные латиницей, оказались в одном кластере. И ни одно русскоязычное описание не попало в этот кластер. При этом оказалось неважно, каким методом выполнялась кластеризация – K-means или DBSCAN.

При использовании для векторного представления слов модели Word2Vec DBSCAN разбил тексты на кластеры таким образом, что для набора1 и набора 2 нашлось по 3 верных сопоставления. В тоже время алгоритм K-means для набора1 составил кластеры таким образом, что

совпали имена функций и их описания в 13 случаях, а для набора2 – в 10. Расчёт эффективности для различных комбинаций моделей векторного представления слов и алгоритмов кластеризации представлен в табл. 1.

Во время эксперимента использовались такие параметры модели Word2Vec:

- размерность вектора слов – 100 (варьировались от 100 до 300);
- размер окна контекста – 4 (варьировались от 3 до 7);
- минимальная частота слова – 1;
- количество потоков – 4;
- алгоритм обучения CBOW (Continuous Bag of Words), т.е. модель пытается восстановить центральное слово по окружающим словам.

Остальные параметры использовали значение по умолчанию, предоставляемое моделью Word2Vec из библиотеки `gensim`.

Размерность выходных векторов для модели `bert-base-multilingual-cased` 768. Модель BERT переводится в режим оценки с помощью `model.eval()`. Это необходимо, чтобы отключить слои, которые используются только во время обучения. При токенизации слов с использованием BERT tokenizer добавлены параметры `truncation=True` и `padding=True`. Это необходимо для обработки случаев, когда входные слова длиннее максимальной длины последовательности, поддерживаемой BERT, и для обеспечения одинаковой длины всех входных последовательностей.

Число кластеров для K-means установлено опытным путём равным 15. Все остальные параметры использовали значение по умолчанию, установленные для этого метода библиотекой `scikit-learn`.

Для модели DBSCAN:

- максимальное расстояние между двумя точками, при котором они считаются соседними друг другу, `eps=0.065`;
- минимальное количество соседей, необходимое точке, чтобы стать центром кластера, `min_samples=4`.

Таблица 1

	Word2Vec+ K-means	Word2Vec+ DBSCAN	BERT+ K-means	BERT+ DBSCAN
Набор1	0,68	0,16	0	0
Набор2	0,83	0,25	0	0

Данный эксперимент показал, что для определения синонимии между именем функции и её описанием наиболее перспективной оказалась связка Word2Vec+K-means, обеспечивающая наилучший баланс между точностью определения синонимических связей и вычислительной эффективностью, что делает её подходящим кандидатом для использования в системе CodeLex. Однако, полученные результаты не являются окончательными и требуют дальнейшей проверки на большем объеме данных и с применением различных стратегий предобработки текста.

Заключение

Проведенное исследование подтвердило перспективность автоматического построения тезауруса для предметной области программирования, в частности, для автоматического выявления синонимических связей между идентификаторами их текстовыми описаниями. Экспериментально установлено, что комбинация методов Word2Vec и K-means показала наилучшие результаты в сравнении с другими исследованными подходами (Word2Vec+DBSCAN, BERT+K-means, BERT+DBSCAN). Однако, несмотря на относительно высокую эффективность связки Word2Vec+K-means (достигающей 0.68 и 0.83 для разных наборов данных), полученные результаты нельзя считать окончательными, и они требуют дальнейшего улучшения и углубленного анализа.

Низкая эффективность BERT в данной задаче требует особого внимания. Вероятной причиной неудовлетворительной работы модели является отсутствие адаптации (fine-tuning) на специализированном корпусе данных, включающем примеры кода и документации по программированию. Предварительно обученные модели BERT, как правило, оптимизированы для работы с общими текстами и могут испытывать трудности с обработкой специфического вокабуляра и синтаксиса языков программирования. Также возможно, что мультиязычная модель BERT недостаточно хорошо адаптирована к сочетанию русского языка (в описаниях) и латиницы (в именах идентификаторов), что привело к обособлению латинских терминов в отдельные кластеры.

Для повышения качества автоматического построения тезауруса следует предпринять ряд шагов.

Во-первых, сравнить использование Word2Vec и BERT с моделью векторного представления слов FastText. FastText, в отличие от Word2Vec, использует информацию о морфологии слов. Это может быть полезно для обработки словоформ в описаниях функций и для более адекватной обработки имён идентификаторов.

Во-вторых, выполнить fine-tuning BERT на специализированных данных, т.е. адаптировать модель BERT к предметной области программирования путем fine-tuning на корпусе, состоящем из исходного кода, документации, комментариев и других текстовых материалов, связанных с языком PHP и другими популярными языками программирования. После fine-tuning необходимо повторно провести эксперименты с BERT+K-means и BERT+DBSCAN, чтобы оценить влияние адаптации модели на качество кластеризации.

Также актуальной задачей является расширение обучающей выборки и анализ ошибок. Вместе с тем важно оптимизировать параметры кластеризации. В дальнейшем, кроме эмпирического подбора параметров, следует применить методы автоматической оптимизации, например, метод локтя (Elbow method) для определения количества кластеров для K-means.

Для расширения тезауруса необходимо исследовать возможность использования тематического моделирования (например, LDA или NMF) для выявления скрытых тем и семантических связей между терминами в области программирования и применения методов, основанных на ассоциативных правилах,

Таким образом, проведенное исследование можно использовать в качестве основы для построения тезауруса в области программирования системы интеллектуального поиска информации «CodeLex».

Список литературы

- [Дюличева, 2021] Дюличева Ю.Ю. Учебная аналитика MOOK как инструмент анализа математической тревожности // Вопросы образования. – 2021. – № 4. – С. 243-265. – doi 10.17323/1814-9545-2021-4-243-265.
- [Жилов, 2023] Жилов Р.А. Интеллектуальные методы кластеризации данных // вестия Кабардино-Балкарского научного центра РАН. – 2023. – № 6(116). – С. 152-159. – DOI 10.35330/1991-6639-2023-6-116-152-159.
- [Коломойцева, 2024] Коломойцева И.А. Особенности структуры тезауруса для повышения качества поиска документов в области программирования // Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС-2024): Сборник материалов и докладов V Международной научно-практической конференции, Донецк, 27–28 ноября 2024 года. – Донецк: Донецкий национальный технический университет, 2024. – С. 245-250.
- [Лукашевич, 2011] Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Московского университета, 2011. – 512 с.
- [Новикова, 2024] Новикова О.А., Ермолов А.Е. Сравнительный анализ эффективности работы алгоритмов кластеризации текстов // Перспективы науки. – 2024. – № 5(176). – С. 24-31.
- [Попова, 2023] Попова О.А. Анализ методов векторизации текстовых документов // Вестник Рязанского государственного радиотехнического университета. – 2023. – № 85. – С. 96-102. – doi: 10.21667/1995-4565-2023-85-96-102.
- [Фомичев, 2023] Фомичев, Д. А. Кластеризация вакансий по их описанию с использованием машинного обучения и методов анализа текста // Международная конференция по мягким вычислениям и измерениям. – 2023. – Т. 1. – С. 201-204.

УДК 004.855.5

doi: 10.15622/rcai.2025.072

ГЛУБОКИЙ ЛЕС ДЛЯ АНАЛИЗА ВЫЖИВАЕМОСТИ В РАМКАХ МНОГОВАРИАНТНОГО ОБУЧЕНИЯ¹

А.В. Константинов (*konstantinov_av@spbstu.ru*)

Л.В. Уткин (*utkin_lv@spbstu.ru*)

Санкт-Петербургский политехнический университет
Петра Великого, Санкт-Петербург

В работе рассматривается задача на стыке машинного обучения и анализа выживаемости в условиях многовариантного обучения с табличными данными. Проблема возникает, когда каждый объект характеризуется не одним, а набором векторов признаков, однако событие происходит для объекта в целом. Для решения данной задачи предложен новый метод на основе глубокого леса. Ключевой особенностью метода является построение модели, обрабатывающей множество векторов признаков в качестве входа, элементами которой являются классические ансамбли деревьев решения. Численные эксперименты с использованием реальных данных демонстрируют применимость предложенного метода.

Ключевые слова: анализ выживаемости, многовариантное обучение, деревья решений, ансамбли, глубокий лес.

Введение

Анализ выживаемости является важной задачей в различных областях науки, включая медицину, инженерию и экономику, позволяющей оценивать вероятность события (например, смерти пациента, отказа оборудования или оттока клиента) в зависимости от времени. В отличие от классификационных моделей, методы анализа выживаемости позволяют оценивать функцию распределения или функцию выживаемости для каждого отдельного объекта, характеризующегося вектором признаков. В последние годы наблюдается растущий интерес к применению методов машинного обучения для решения задач анализа выживаемости, особенно в условиях, когда число примеров мало, а данные представлены в табличном виде с большим

¹ Работа выполнена при финансовой поддержке РФФИ (проект № 25-21-00103).

количеством признаков. Традиционные методы анализа выживаемости, такие как модель Кокса, случайный лес выживаемости (Random Survival Forest, RSF), успешно применяются на практике, однако не дают возможности использовать в качестве входа множества векторов признаков.

Вместе с тем возникает специфическая проблема при работе с данными, где каждый объект характеризуется не одним набором признаков, а множеством экземпляров, каждый из которых представляется в виде отдельного вектора признаков. Эти векторы могут представлять собой характеристики однородных элементов системы, результаты различных анализов одного пациента в разных условиях или параметры взаимодействующих объектов. При этом событие (например, отказ) относится к объекту в целом, а не к конкретному вектору признаков из множества. Этот сценарий соответствует парадигме многовариантного обучения (MBO) [Carbonneau et al., 2018], которая традиционно применялась для задач классификации изображений и распознавания объектов, анализа гистологических изображений высокого разрешения, прогнозирования активности препаратов, и так далее. Интеграция методов анализа выживаемости с подходом MBO представляет собой актуальную и сложную задачу, требующую разработки новых алгоритмов, способных эффективно обрабатывать данные в виде множеств векторов признаков в условиях цензурированности.

В настоящей работе предлагается новый метод решения задачи анализа выживаемости в условиях MBO на основе композиции случайных лесов выживаемости в формате глубокого леса [Zhou, 2019]. Предложенный подход заключается в построении модели, которая принимает на вход множество векторов признаков и использует ансамбль деревьев решений для оценки функции риска. Ключевой особенностью разработанного метода является его способность эффективно агрегировать информацию из различных векторов признаков, представляющих один объект, для прогнозирования времени до события. Результаты численных экспериментов на реальных данных демонстрируют перспективность предложенного подхода и его применимость в задачах анализа выживаемости с многовариантными данными.

1. Анализ выживаемости и многовариантное обучение

В данном разделе приведены постановки задачи выживаемости, MBO и рассматриваются способы определения задачи анализа выживаемости в рамках MBO.

1.1. Анализ выживаемости

Задача анализа выживаемости заключается в моделировании условного распределения времени до наступления определенного события (например, смерти, отказа оборудования или рецидива заболевания) при за-

данном векторе признаков объекта \mathbf{x} . А именно, требуется оценить функцию выживаемости $S(t | \mathbf{x}) = P(T > t | \mathbf{x})$. В отличие от традиционных задач классификации и регрессии, анализ выживаемости учитывает возможность цензурирования данных – ситуации, когда событие не было зафиксировано в течение периода наблюдения за объектом. Поэтому набор обучающих данных, по которому требуется построить оценку, состоит из троек $(\mathbf{x}_i, y_i, \delta_i)$, где помимо вектора признаков \mathbf{x}_i и времени y_i содержится метка цензурирования δ_i , принимающая либо значение 1, обозначающее что событие в данный момент произошло, либо 0, если событие не произошло до этого момента.

1.2. Многовариантное обучение

В парадигме МВО набор данных состоит из множеств или групп экземпляров, каждый из которых содержит отдельные экземпляры-векторы признаков. Стандартной задачей в рамках МВО является бинарная классификация, где класс группы считается положительным, если хотя бы один экземпляр в этой группе является ее представителем, то есть соответствует определенному условию. Задача заключается в построении модели, которая, принимая на вход множество экземпляров, определяет класс для всей группы. В отличие от традиционных задач обучения с учителем, где каждый объект имеет однозначную метку, в многовариантном обучении метка доступна только для группы [Carbonneau et al., 2018]. Поскольку целью настоящей работы является построение модели выживаемости, рассматривается более общая постановка, где метка группы может соответствовать не одному конкретному экземпляру группы, а части или всем экземплярам в совокупности. Такая общая постановка рассмотрена в работе [Yao et al., 2019], где для анализа выживаемости в рамках многовариантного обучения проводился анализ гистологических образцов высокого разрешения, фрагменты которых представляли собой отдельные экземпляры, а времена событий были доступны только для целых изображений. Таким образом, требуется построить не зависимость от одного экземпляра и определить активный экземпляр в группе, а функцию множества экземпляров. Альтернативный подход к определению МВО для анализа выживаемости, заключающийся во включении явной агрегации функций выживаемости в постановку задачи, требует разработки специализированных правил для каждой конкретной прикладной области и поэтому не рассматривается в данной работе.

1.3. Модели на основе деревьев решений

Наиболее часто применяемыми для обработки табличных данных являются модели на основе ансамблей деревьев решений. Для задач выживаемости таким ансамблем является случайный лес выживаемости [Ishwaran, 2008]. Он существенно отличается от классических лесов де-

реьев решений. В листьях каждого дерева такого леса строится безусловная модель Каплана-Мейера, а правила расщепления в узлах определяются таким образом, чтобы максимизировать логранговый критерий. Стоит отметить, что несмотря на высокий потенциал деревьев решений для обработки табличных данных и успешное использование таких моделей для МВО [Leistner et al., 2010], модели выживаемости на основе деревьев в рамках МВО ранее не применялись.

Для реализации модели на основе деревьев решений для анализа выживаемости в рамках МВО, в настоящей работе предлагается строить более сильную, композиционную модель, включающую множество этапов обработки данных, то есть модель, представляющую собой глубокий лес (Deer Forest) [Zhou, 2019]. Такой лес, в отличие от классического алгоритма случайного леса, где каждое дерево строится независимо, строится послойно, используя выходные данные предыдущих деревьев в качестве входных признаков для последующих. Это позволяет модели реализовывать обучение представлениям, учитывая более сложные зависимости между признаками и повышая за счет этого обобщающую способность. Глубокий лес демонстрирует высокую точность и эффективность на различных задачах машинного обучения, включая задачи классификации, регрессии, а также анализа выживаемости [Utkin et al., 2020], однако ранее глубокий лес для анализа выживаемости в рамках МВО разработан не был.

2. Метод построения глубокого леса выживаемости в условиях МВО

Ансамбли деревьев решений позволяют строить наиболее точные модели для табличных данных, в том числе в рамках анализа выживаемости. Поэтому в данной работе предлагается реализовать подход на основе таких ансамблей, опирающийся на их преимущества, и вместе с тем позволяющий обрабатывать данные в формате МВО.

Для применения ансамбля деревьев решений к группам примеров можно рассмотреть два подхода. Первый подход предполагает выбор репрезентативных объектов (представителей) из каждой группы для формирования набора данных в формате “объект – время – метка события”. Однако данный подход оказывается неприменим в случае, если у группы отсутствует один представитель, т.е. если более одного экземпляра в группе влияют на функцию выживаемости. Поэтому предлагается второй подход, состоящий в создании псевдо-разметки путем повторения метки группы для каждого экземпляра и последующему применению ансамблевого метода ко всем экземплярам каждой группы. На рис. 1 представлена схема предлагаемого метода, где указано как для всех экземпляров первой группы $x_{1,1}$, $x_{1,2}$, $x_{1,3}$ повторяются метки y_1 , δ_1 . Сформированный набор данных, где каждому экземпляру сопоставлены метки времени и цензурирования, используется для обучения случайного леса выживаемости.

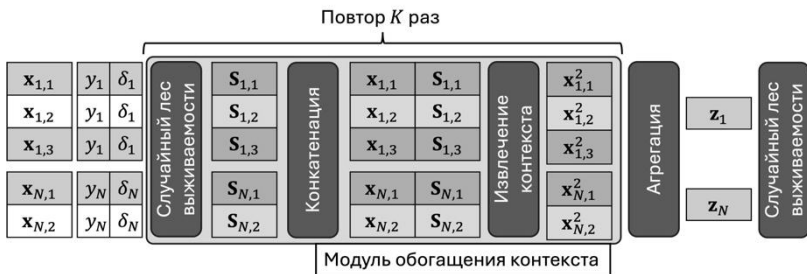


Рис. 1. Общая схема предлагаемого глубокого леса

Далее с помощью построенного леса выживаемости для каждого экземпляра строится оценка функции выживаемости: $S_{1,1}, \dots, S_{N,2}$. Поскольку число уникальных времен событий E в обучающем наборе данных конечно, функции выживаемости представляются в виде векторов, размерность которых совпадает с E , а каждое значение задает вероятность, что событие произойдет после соответствующего временного интервала. Полученные оценки можно рассматривать как дополнение представления исходных векторов экземпляров. С целью снижения объёма обрабатываемых данных производится понижение разрешения временной сетки, путем усреднения значений выживаемости в соседних интервалах. Для получения нового представления исходные векторы конкатенируются к соответствующим оценкам выживаемости, таким образом формируются новые векторы $(\mathbf{x}_{1,1}, S_{1,1}), \dots, (\mathbf{x}_{N,2}, S_{N,2})$.

Полученные новые представления экземпляров комбинируются в блоке извлечения контекста, для расширения представления экземпляра информации о группе, в которую он включён. К каждому вектору признаков конкатенируется среднее значение представлений всех остальных экземпляров в группе. С целью снижения вычислительной сложности данной процедуры, которая в простейшей реализации является квадратичной по числу экземпляров в группе, для каждой группы размера g сперва рассчитывается сумма всех представлений, затем для каждого экземпляра из суммы вычитается его вектор-представление, и полученный вектор умножается на $1/(g - 1)$. Сложность при такой реализации становится линейной. Блок извлечения контекста позволяет учесть взаимодействия отдельных экземпляров и их влияние в совокупности на распределение времени события.

После применения вышеописанной процедуры, составляющей модуль обогащения контекста, для каждого экземпляра формируется новое представление, содержащее, помимо информации о данном экземпляре, контекст, то есть необходимую информацию о других экземплярах группы,

или ее часть. Данный модуль применяется K раз (с различными случайными лесами). Затем для построения модели на уровне групп применяется агрегация за счет усреднения представлений в каждой группе. В результате формируется обучающий набор данных, где вместо экземпляров содержатся векторы-представления групп $\mathbf{z}_1, \dots, \mathbf{z}_N$, которым соответствуют исходные времена y_1, \dots, y_N . С помощью данного обучающего набора строится финальный случайный лес выживаемости, который позволяет строить функцию выживаемости для групп, а не экземпляров в отдельности.

3. Численные эксперименты

Предложенный метод был реализован на языке Python с использованием фреймворка для разработки глубоких лесов Bosk и случайных лесов выживаемости из пакета scikit-survival. Для проведения численных экспериментов с реальными данными были использованы четыре широко применяемых для анализа выживаемости набора данных: Breast Cancer (BC), German Breast Cancer Study Group 2 (GBSG2), Veterans и Worcester Heart Attack Study (WHAS500).

Для исследования предложенного метода в рамках MBO исходные наборы данных были приведены к формату группа – время – метка события с помощью следующего преобразования. Для каждого примера вектор признаков разбивался на множество новых векторов признаков, описывающее объект (группу) обучающей выборки в рамках MBO. Каждый такой вектор признаков соответствует отдельной компоненте исходного вектора, и состоит из значения и номера данной компоненты. Таким образом, размер множества (группы) равен числу признаков в исходной задаче, а сами векторы признаков в новой задаче принадлежат дизъюнктному объединению областей определения признаков исходной задачи.

Тестирование проводилось с помощью перекрёстной проверки (кросс-валидации) с 5 разбиениями, сбалансированными по доле цензурированных примеров, повторяющейся с различными разбиениями 20 раз. Для оценки точности применялся индекс конкордации (С-индекс), позволяющий оценить корректность ранжирования ожидаемого времени события с помощью модели, и интегрированная оценка Брайера (IBS), позволяющая оценить соответствие предсказанного распределения тестовой выборке. Далее представлены результаты сравнения предложенного глубокого леса (ГЛ-MBO), решающего поставленную проблему в рамках MBO, с выступающим в качестве базового метода классическим случайным лесом выживаемости (RSF) и RSF со рандомизированным построением деревьев решений (RSF-Extra), применёнными к исходной выборке. Результаты для С-индекса приведены в табл. 1, где каждая строка соответствует отдельной тестируемой модели. Чем выше значение С-индекса, тем ниже ошибка ранжирования ожидаемого времени моделями.

Как видно из табл. 1, предложенная модель не уступает по точности базовому RSF в трех из четырех случаев, а для набора данных Veteran позволяет существенно превзойти его по точности, в то время как задача в многовариантной постановке является более сложной, чем исходная.

Таблица 1

Модель	BC	GBSG2	Veterans	WHAS500
RSF	0,650	0,689	0,675	0,767
RSF-Extra	0,689	0,673	0,653	0,745
ГЛ-MBO	0,677	0,687	0,702	0,760

Для интегрированной оценки Брайера результаты приведены в табл. 2. Чем меньше значение оценки, тем лучше прогнозы модели соотносятся тестовой выборкой.

Таблица 2

Модель	BC	GBSG2	Veterans	WHAS500
RSF	0,162	0,173	0,129	0,167
RSF-Extra	0,154	0,179	0,133	0,177
ГЛ-MBO	0,155	0,173	0,127	0,175

Аналогично, в соответствии с табл. 2, предложенная модель превосходит базовый RSF на всех наборах данных, кроме WHAS500, а базовый RSF-Extra на всех, кроме BC.

Таким образом, можно сделать вывод, что предложенный метод позволяет решать задачи выживаемости в условиях многовариантного обучения с достаточной точностью. С вычислительной точки зрения, затраты для построения глубокого леса превосходят таковые у базового метода (RSF) приблизительно в $(K + 1)$ раз, что не является существенным ограничением для применимости данного метода в условиях MBO.

Заключение

В настоящей работе предложен новый метод решения задачи анализа выживаемости в условиях MBO с табличными данными на основе глубокого леса. Разработанный подход позволяет эффективно обрабатывать в качестве входа множество векторов признаков, представляющих один объект, и агрегировать информацию из них для прогнозирования распределения времени до события. Ключевой особенностью метода является построение многослойной модели, состоящей из ансамблей деревьев решений и блоков извлечения контекста, способной учитывать сложные зависимости между признаками и обеспечивающей высокую точность оценки функции выживаемости.

Численные эксперименты с использованием реальных данных продемонстрировали перспективность предложенного подхода и его применимость в задачах анализа выживаемости с многовариантными данными. Полученные результаты показали, что разработанный метод демонстрирует сопоставимую точность по сравнению с ансамблями, примененными к классической задаче анализа выживаемости на исходных данных, либо превосходит базовые методы по показателям индекса конкордации и интегрированной оценки Брайера.

Несмотря на достигнутые результаты, следует отметить некоторые ограничения предложенного метода. В частности, эффективность работы алгоритма может зависеть от качества и структуры входных данных, а также от выбора оптимальных гиперпараметров модели. Кроме того, дальнейшие исследования могут быть направлены на разработку методов автоматической настройки гиперпараметров для повышения устойчивости и обобщающей способности предложенного подхода на каждом слое глубокого леса. В перспективе представляется интересным изучение возможности применения разработанного метода к другим типам данных, таким как изображения, разбитые на фрагменты – патчи, или текстовые данные, а также расширение его функциональности для обеспечения интерпретируемости модели в целом с целью выявления экземпляров и их признаков, вносящих наиболее значимый вклад в предсказания.

Список литературы

- [Carboneau et al., 2018] Carboneau M.A., Cheplygina V., Granger E., Gagnon G. Multiple instance learning: A survey of problem characteristics and applications // *Pattern recognition*. – 2018. – Vol. 77. – P. 329-353. – doi: 10.1016/j.patcog.2017.10.009.
- [Ishwaran, 2008] Ishwaran H., Kogalur U.B., Blackstone E.H., Lauer M.S. Random Survival Forests // *The Annals of Applied Statistics*. – 2008. – P. 841-860. – doi: 10.1214/08-AOAS169.
- [Leistner et al., 2010] Leistner C., Saffari A., Bischof H. MIForests: Multiple-instance learning with randomized trees // *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI 11*. – Springer Berlin Heidelberg, 2010. – P. 29-42. – doi: 10.1007/978-3-642-15567-3_3.
- [Utkin et al., 2020] Utkin L.V., Konstantinov A.V., Lukashin A.A., Muliukha V.A. An adaptive weighted deep survival forest // *2020 XXIII International Conference on Soft Computing and Measurements (SCM)*. – IEEE, 2020. – P. 198-201. – doi: 10.1109/SCM50615.2020.9198755.
- [Yao et al., 2019] Yao J., Zhu X., Huang J. Deep multi-instance learning for survival prediction from whole slide images // *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*. – Springer International Publishing, 2019. – P. 496-504. – doi: 10.1007/978-3-030-32239-7_55.
- [Zhou, 2019] Zhou Z.H., Feng J. Deep forest // *National science review*. – 2019. – Vol. 6, No. 1. – P. 74-86. – doi: 10.1093/nsr/nwy108.

УДК 004.838.2

doi: 10.15622/rcai.2025.073

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ КОМБИНИРОВАНИЯ МЕТОДОВ ЗАЩИТЫ СИСТЕМ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ ОТ ГРАДИЕНТНЫХ СОСТЯЗАТЕЛЬНЫХ АТАК¹

И.В. Котенко (*ivkote@comsec.spb.ru*)

И.Б. Саенко (*ibsaen@comsec.spb.ru*)

В.Е. Садовников (*bladimir1998@mail.ru*)

Санкт-Петербургский Федеральный исследовательский центр РАН,
Санкт-Петербург

В статье рассматривается подход к защите систем распознавания изображений от градиентных состязательных атак, основанный на комбинировании различных методов защиты, включающих зашумление, сжатие и нейросетевую очистку изображений. Комбинирование методов защиты предполагает нахождение оптимальных параметров, характеризующих эти методы, при которых точность распознавания становится максимальной. Оценка эффективности рассматриваемых методов защиты производится на датасете STL-10. Выбор этого датасета обусловлен его широким применением в задачах классификации изображений. Результаты экспериментов показывают, что комбинирование указанных методов защиты позволяет достигнуть максимальной точности распознавания изображений в условиях воздействия на них градиентных состязательных атак.

Ключевые слова: защита от атак, искусственный интеллект, состязательная атака, распознавание изображений, машинное обучение, зашумление, кибербезопасность.

Введение

Современные системы распознавания изображений на базе глубоких нейронных сетей демонстрируют высокую точность в задачах классификации, сегментации, детектирования объектов и других областях компью-

¹ Работа выполнена при частичной финансовой поддержке бюджетной темы FFZF-2025-0016.

терного зрения. Эти достижения широко применяются в таких критически важных сферах, как медицинская диагностика, системы видеонаблюдения, автономное вождение, биометрическая идентификация и интеллектуальные промышленные системы [Понкин и др., 2024]. Однако несмотря на достигнутый прогресс нейросетевые модели остаются уязвимыми к состязательным атакам – специально сконструированным входным данным (примерам), способным вызвать ошибочное поведение модели при визуальной неотличимости искажений для человека [Легашев и др., 2024].

Исследования показали, что даже минимальные возмущения, добавленные к исходному изображению, могут привести к резкому снижению точности классификации и полному нарушению работы модели [Баев и др., 2024]. Наиболее известным примером таких атак является метод Fast Gradient Sign Method (FGSM). При этом методе для создания возмущения, искажающего классификацию, используется однократное вычисление градиента функции потерь. Развитие этого подхода привело к появлению метода Iterative FGSM (IFGSM), в котором атака осуществляется через серию последовательных шагов, повышая эффективность вмешательства. В условиях ограниченного доступа к внутренним параметрам модели применяются методы «черного ящика», такие как Zeroth Order Optimization (ZOO), основанные на численной оценке градиента без использования внутренней информации о сети. Эти методы демонстрируют высокую эффективность при обходе защиты, даже при минимальной информации о целевой модели. Все градиентные атаки наглядно демонстрируют уязвимость глубоких моделей и подчеркивают необходимость разработки надежных средств защиты, особенно в контексте практического применения вне контролируемой среды.

На фоне растущей актуальности проблемы в последние годы активно исследуются подходы к повышению устойчивости моделей к подобным воздействиям. Среди них – модификации архитектур, обогащение обучающих выборок, использование робастных функций потерь, а также методы предобработки входных данных, направленные на нейтрализацию состязательных искажений. К числу последних относятся такие подходы, как JPEG-сжатие, добавление случайных шумов, а также нейросетевая очистка изображений с помощью автоэнкодеров.

Вклад и новизна статьи заключаются в следующем: (1) предложен новый подход к защите от градиентных состязательных атак, основанный на комбинировании методов защиты; (2) проведено экспериментальное исследование эффективности защиты в одиночном, попарном и комбинированном режимах применения методов; (3) определены оптимальные параметры методов защиты, при которых точность обнаружения атак становится максимальной. Эксперименты подтвердили, что комбинирование методов защиты с оптимальными параметрами дает наивысшую точность обнаружения градиентных атак.

1. Анализ текущего состояния исследований

Одним из наиболее эффективных направлений обеспечения устойчивости систем компьютерного зрения являются гибридные методы, сочетающие зашумление, сжатие и нейронную очистку входных изображений, позволяющие снижать влияние атак без большого ущерба для эффективности распознавания [Фомичева и др., 2023].

В [Li et al., 2023] описан метод ComDefend, основанный на адаптивном сжатии и восстановлении структуры изображения без изменения архитектуры классификатора. Он демонстрирует высокую устойчивость к атакам типа FGSM. Метод Compress and Restore [Naveen et al., 2023] использует итеративное JPEG-сжатие с переменным качеством и генеративное восстановление (AR-GAN). В работе [Gardella et al., 2022] подчеркивается эффективность сочетания шума и JPEG-сжатия для судебно-медицинского анализа изображений и восстановления изображений.

Гибридный метод HAT (Hybrid Adversarial Training) [Ali et al., 2022] дополняет обучение состязательными примерами, созданными путем объединения атак DeepFool и FGSM, тем самым увеличивая устойчивость моделей глубокого обучения в течение установленного периода времени против различных атак.

В [Jain, 2024] исследуется влияние различных типов атак, включая атаку FGSM, на точность алгоритма обнаружения и дорожных знаков YOLOv5. Результаты показывают, что алгоритм подвержен этим атакам, причем показатели неправильной классификации увеличиваются по мере увеличения величины возмущений.

В [Kotlyarov et al., 2023] рассмотрен алгоритм создания нейросети, ориентированной на распознавание образов, и рассмотрены несколько видов состязательных атак, включая FGSM. Получено подтверждение гипотезы о снижении точности распознавания нейросети при реализации состязательной атаки злоумышленником.

В [Tian et al., 2024] предложен подход к построению моделей глубокого обучения, устойчивых к воздействию различных состязательных атак, основанный на преобразовании изображений для удаления состязательных шумов. Экспериментальные результаты подтвердили потенциал методов преобразования изображений как надежной защиты от состязательных атак в системах классификации изображений на основе глубокого обучения, особенно в сочетании с передовыми архитектурами нейронных сетей, такими как ResNet50 и DenseNet121.

В [Khamaiseh et al., 2022] рассматриваются методы состязательных атак, включая FGSM, IFGSM и ZOO, с акцентом на объяснение математических концепций и терминологии, а также механизмы защиты с обсуж-

дением их эффективность в защите глубоких нейронных сетей от состязательных атак. Показано, что идея использовать нейронную очистку с автоэнкодерами, зашумление и сжатие изображений являются перспективными методами защиты от современных состязательных атак.

Авторами настоящей работы также изучался данный вопрос. Так, в [Kotenko et al., 2024] была предложена схема JPEG-сжатия и нейронной фильтрации, показавшая высокую эффективность на двух наборах данных. В [Котенко и др., 2025] был успешно протестирован подход на основе JPEG и Neural Cleanse, позволивший почти полностью восстановить точность классификации на атакованных изображениях, особенно при использовании ансамблевых моделей.

Таким образом, анализ современных исследований в области защиты от градиентных состязательных атак показывает, что гибридные методы являются здесь одним из наиболее перспективных направлений. Настоящая работа посвящена исследованию комбинирования трех методов защиты.

2. Градиентные состязательные атаки

Метод FGSM стал одной из первых и наиболее широко используемых реализаций градиентных состязательных атак на нейросетевые классификаторы [Goodfellow et al., 2015]. Основная идея FGSM заключается в том, чтобы модифицировать входное изображение таким образом, чтобы оно вызывало ошибочную классификацию модели, при этом изменение оставалось незаметным для человеческого глаза.

Атака основывается на градиенте функции потерь по отношению к входным данным. Состязательное возмущение вычисляется как

$$\eta = \varepsilon \cdot \text{sign}(\nabla_x J(\theta, x, y)), \quad (2.1)$$

где η – добавляемое к изображению возмущение; ε – малая константа, определяющая интенсивность искажений; $J(\theta, x, y)$ – функция потерь модели; $\nabla_x J(\theta, x, y)$ – градиент функции потерь по отношению к входу. Метод предполагает однократное вычисление градиента функции потерь и направляет изменение пикселей входного изображения в сторону, в которой ошибка модели возрастает максимально быстро. Это делает FGSM не только эффективным, но и вычислительно дешевым методом, так как он не требует итеративного процесса оптимизации.

Метод IGSM представляет собой итеративное расширение метода FGSM, обеспечивающее более точную настройку состязательных возмущений [Jiakai, 2018]. Идея IGSM состоит в итеративном применении атаки FGSM к изображению с небольшим шагом, позволяя более точно направлять изображение в сторону неверной классификации, сохраняя при этом низкую заметность искажений.

Состязательное возмущение вычисляется по итеративной формуле

$$x^{(t+1)} = \text{Clip}_{x,\varepsilon} \left(x^{(t)} + \varepsilon \cdot \text{sign} \left(\nabla_x J(\theta, x, y) \right) \right), x^{(0)} = x, \quad (2.2)$$

где $x^{(t)}$ – изображение на итерации t ; ε – малый шаг изменения пикселей на каждой итерации; $J(\theta, x^{(t)}, y)$ – функция потерь; $\nabla_x J(\theta, x^{(t)}, y)$ – градиент функции потерь, $\text{Clip}_{x,\varepsilon}$ – операция отсечения, гарантирующая, что возмущенное изображение x останется в заданной области ε допустимых значений пикселей. За счет итеративного процесса IGSM демонстрирует более высокую эффективность по сравнению с одношаговым FGSM, однако требует больших вычислительных затрат.

Атака ZOO представляет собой метод «черного ящика», не требующий доступа к внутренним параметрам модели или ее градиентам [Chen et al., 2017]. В отличие от традиционных атак «белого ящика», она позволяет выполнять состязательные воздействия на нейросеть, имея доступ только к значениям выходов модели. Основная идея атаки состоит в численной аппроксимации градиентов функции потерь относительно входного изображения. Так как градиенты напрямую недоступны, ZOO использует метод конечных разностей, чтобы оценить производную по каждой координате (пикселю) входа:

$$\nabla_{x,i} f(x) \approx \frac{f(x + \varepsilon e_i) - f(x - \varepsilon e_i)}{2\varepsilon}, \quad (2.3)$$

где $\hat{\nabla}_{x,i} f(x)$ – аппроксимированная производная функция потерь; x – входное изображение; e_i – единичный вектор с единицей на i -й позиции и нулями в остальных координатах; $f(x)$ – функция потерь, зависящая от выхода модели и целевого класса. Полученные градиенты используются для итеративного обновления входного изображения. Вместо одновременного обновления всех координат, ZOO применяет стохастический координатный спуск – на каждой итерации обновляется одна или несколько координат, что позволяет существенно сократить количество вызовов модели. ZOO показал эффективность, сравнимую с «белыми» атаками, особенно на задачах классификации изображений, при этом оставаясь применимым в условиях ограниченного доступа к модели.

3. Методы защиты

Метод защиты на основе JPEG-сжатия основан на идее удаления из изображения высокочастотных компонентов, не воспринимаемых человеком [Das et al., 2017]. В JPEG-сжатии изображение делится на блоки (обычно 8×8) и преобразуется с помощью дискретного косинусного преобразования с последующим квантованием коэффициентов, что ослабляет высокочастотные составляющие. Так как состязательный шум чаще всего проявляется именно в виде незначимых высокочастотных колебаний,

JPEG-сжатие действует как фильтр, ослабляющий такие возмущения. Экспериментально подтверждена его эффективность: увеличение степени сжатия приводит к «восстановлению» точности на атакованных изображениях [Aydemir et al., 2024]. Однако сильное сжатие ухудшает качество изображений (блоковые артефакты, искажение цветности), что может негативно сказываться на распознавании.

Наложение шума на изображение также является эффективным методом защиты нейросетевых моделей от состязательных атак [Hossain et al., 2022]. Идея метода заключается в следующем. Так как состязательные искажения представляют собой малые возмущения, незаметные для человеческого зрения, наложение более интенсивного случайного шума способно «подавить» их влияние, снижая вероятность некорректного распознавания. Метод зашумления может быть реализован как на стадии предобработки входных изображений, так и в процессе обучения модели в рамках стратегии расширения обучающей выборки. Однако чрезмерное усиление шумового компонента способно привести к искажению семантически значимой информации. Таким образом, данный метод требует выбора оптимального уровня шума.

Нейросетевая очистка входных данных является методом защиты от состязательных атак, в котором используются специальные нейронные архитектур, способные устранять нежелательные искажения до подачи изображения в основную модель классификации. В качестве таких нейросетей обычно применяются автоэнкодеры, которые обучены восстанавливать исходное изображение на основе его искаженной версии.

Наиболее популярными являются следующие типы автоэнкодеров: автоэнкодер с шумоподавлением (Denoising Autoencoder, DAE); глубокий автоэнкодер на базе сверточных и транспонированных сверточных слоев (Convolution Autoencoder, CAE); вариационный автоэнкодер (Variational Autoencoder, VAE).

Автоэнкодеры DAE обучаются на парах (x_{noisy}, x) , где x_{noisy} – искаженный вариант изображения, а x – его «чистый» аналог. Задача автоэнкодера DAE сводится к восстановлению исходного изображения по его искаженной версии [Chen et al., 2024]. Это приводит к обучению DAE с целью минимизации потерь восстановления.

Автоэнкодеры CAE благодаря своей глубокой структуре и обучению на больших выборках могут выявлять локальные и глобальные закономерности, что позволяет повысить точность классификации после атаки на 10–20 % по сравнению с необработанными входами [Ashraf et al., 2024].

Автоэнкодеры VAE моделируют вероятностное распределение в латентном пространстве [Xie et al., 2021]. Вместо отображения входа x в фиксированный вектор признаков z , энкодер формирует два параметра: среднее $\mu(x)$ и дисперсию $\sigma^2(x)$, задающие нормальное распределение. Из этого распределения далее выбирается латентный вектор. Благодаря

регуляризации латентного пространства, VAE обеспечивает устойчивость к мелким искажениям входных данных. Он не только восстанавливает семантику изображения, но и ограничивает возможные вариации в латентной области, что делает затруднительным прохождение состязательных примеров через декодер без потерь. Таким образом, VAE может рассматриваться как робастный автоэнкодер, обладающий потенциалом к устранению состязательных возмущений.

4. Реализация

Предложенный подход был реализован в среде PyCharm. Были установлены следующие библиотеки: `import torch; import torch.nn as nn; import torch.optim as optim; import matplotlib.pyplot as plt; import numpy as np.`

В качестве набора данных изображений использовался датасет STL-10 (Stanford TensorLab). Он применяется для задач классификации цветных изображений и содержит 30000 снимков размером 96 x 96 пикселей, распределенных по 10 классам. Изображения 96 x 96 находят практическое применение в системах распознавания объектов на краевых устройствах, таких как камеры видеонаблюдения или смартфоны, где важны скорость обработки и низкое энергопотребление. Этот размер часто используется в промышленной автоматизации для классификации дефектов на конвейере – например, различия между деталями, браком или упаковкой – без необходимости в высоком разрешении. Также 96×96 активно применяется в медиасервисах для предварительной сортировки изображений, например, в фильтрации контента или тегировании фото в облачных хранилищах, где обрабатываются тысячи изображений в реальном времени.

Для распознавания изображений использовалась предобученная модель ResNet-18, адаптированная для задачи классификации на наборе данных STL-10. В первом сверточном слое conv1 проведена замена оригинальных параметров `kernel_size=7`, `stride=2`, `padding=3` на `kernel_size=3`, `stride=1`, `padding=1`. Это позволяет лучше обрабатывать изображения меньшего размера (96×96 пикселей), характерные для STL-10, сохраняя при этом пространственную информацию и уменьшая агрессивное уменьшение размерности на первом шаге. Последний полносвязный слой fc заменен с 1000 выходов на слой с 10 выходами, соответствующими количеству классов в STL-10. После этих модификаций модель дообучается на обучающей выборке STL-10, используя оптимизатор Adam и функцию потерь – кросс-энтропию. Обучающая выборка состояла из 5000 изображений, а тестовая – из 8000.

Автоэнкодеры имеют схожие архитектуры: начинаются с входного слоя, который принимает изображения размером 96x96x3. В энкодере используется архитектура, включающая два сверточных слоя. Первый сверточный слой имеет 32 фильтра с размером ядра 3x3 и функцией акти-

вации ReLU. За ним следует слой MaxPooling с размером пулинга 2x2. Второй сверточный слой содержит 64 фильтра, также с размером ядра 3x3 и активацией ReLU, после чего применяется еще один слой MaxPooling с размером пулинга 2x2.

Данный подход, объединяющий методы защиты на основе автоэнкодеров, наложения шума и JPEG-сжатия, не рассматривался в контексте систем обработки изображений в реальном времени (real-time). Автоэнкодеры, и JPEG-сжатие – являются вычислительно затратными операциями, особенно при их последовательном применении. Это приводит к увеличению общей задержки на кадр, что делает сложно достижимым соблюдение временных ограничений, характерных для real-time систем, таких как видеонаблюдение, автономные транспортные средства или интерактивные приложения.

5. Проведение экспериментов

Общий план проведения экспериментов заключался в следующем: применение каждого метода защиты в отдельности; комбинация двух методов; комбинация лучших значений всех трех методов. Определялась точность распознавания изображений из датасета STL-10 после воздействия атак FGSM, IFGSM и ZOO по следующей формуле:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (5.1)$$

где TP – число правильных решений о наличии объекта; FP – число ошибок второго рода; FN – число ошибок первого рода; TN – число правильных решений об отсутствии объекта.

В табл. 1 показаны значения $Accuracy$ при применения каждого метода защиты в отдельности. Значение ϵ для каждой атаки было равно 0,04 или 0,06. Значение $Accuracy$ без атаки было равно 0,75.

Таблица 1

Метод защиты	FGSM		IFGSM		ZOO	
	0,04	0,06	0,04	0,06	0,04	0,06
DAE	0,26	0,19	0,39	0,34	0,66	0,62
CAE	0,35	0,29	0,40	0,36	0,49	0,51
VAE	0,21	0,12	0,23	0,13	0,64	0,64
JPEG ($q = 5$)	0,37	0,32	0,41	0,38	0,52	0,51
JPEG ($q = 10$)	0,36	0,25	0,41	0,36	0,67	0,62
JPEG ($q = 15$)	0,30	0,20	0,37	0,29	0,64	0,69
Гайсс ($\sigma = 0,2$)	0,18	0,16	0,19	0,18	0,16	0,16
Гайсс ($\sigma = 0,3$)	0,13	0,12	0,12	0,12	0,13	0,13
Гайсс ($\sigma = 0,4$)	0,11	0,11	0,11	0,11	0,09	0,09

Из рис. 1 видно, что для атак FGSM и IFGSM наилучшую точность распознавания обеспечивал автоэнкодер CAE. Для атаки ZOO при $\varepsilon = 0,04$ лучшим был DAE, а при $\varepsilon = 0,06$ – VAE. Наилучшим параметром сжатия JPEG для атак FGSM и IFGSM был $q = 5$, а для ZOO – $q = 10$ при $\varepsilon = 0,04$ и $q = 15$ при $\varepsilon = 0,06$. Наилучшим параметром шума Гаусса для всех атак – $\sigma = 0,2$, хотя влияние шума на защиту от атак было наименее сильным.

При применении попарной комбинации методов оптимальные параметры методов для различных атак могут становиться другими. В табл. 2 показаны значения *Accuracy* с учетом этих оптимальных параметров.

Таблица 2

Комбинация методов защиты	FGSM		IFGSM		ZOO	
	0,04	0,06	0,04	0,06	0,04	0,06
DAE + JPEG ($q = 5$)	0,42	0,37	-	-	-	-
CAE + JPEG ($q = 10$)	0,36	0,32	0,48	0,44	0,57	0,56
CAE + JPEG ($q = 15$)	0,36	0,32	0,46	0,42	0,62	0,64
DAE + Гаусс ($\sigma = 0,2$)	0,30	0,23	0,38	0,35	0,47	0,49
JPEG ($q = 10$) + Гаусс ($\sigma = 0,2$)	0,37	0,36	0,38	0,35	0,49	0,31
JPEG ($q = 15$) + Гаусс ($\sigma = 0,2$)	0,36	0,30	0,35	0,30	0,48	0,46

Из сравнения табл. 1 и 2 видно, что значения *Accuracy* при попарной комбинации (AE + JPEG) выше, чем при одиночном использовании этих методов при тех же параметрах. Если в попарной комбинации используется метод зашумления, то значения *Accuracy* будут выше, чем при одиночном использовании метода зашумления, но ниже, чем при одиночном использовании метода нейронной очистки или JPEG-сжатия.

Последним этапом проведения экспериментов являлось полное комбинирование всех трех методов защиты (табл. 3).

Таблица 3

Комбинация методов защиты	FGSM		IFGSM		ZOO	
	0,04	0,06	0,04	0,06	0,04	0,06
DAE+JPEG($q=5$)+Гаусс($\sigma=0,2$)	0,36	0,30	0,39	0,35	0,39	0,45
DAE+JPEG ($q=10$)+Гаусс($\sigma=0,2$)	0,31	0,26	0,35	0,33	0,39	0,41
CAE+JPEG($q=5$)+Гаусс ($\sigma=0,2$)	0,23	0,21	0,27	0,25	0,32	0,29
CAE+JPEG ($q=10$)+Гаусс($\sigma=0,2$)	0,30	0,23	0,27	0,26	0,29	0,27

Как видно из табл. 3, для всех атак при различных значениях ε наибольшее значение *Accuracy* достигается в случае комбинирования методов DAE, JPEG-сжатия с $q = 5$ и добавления шума Гаусса с $\sigma = 0,2$. При этом, сравнивая табл. 3 и 2, можно заметить, что при полном комбинировании методов защиты значения *Accuracy* меньше, чем при парном применении автоэнкодеров и JPEG-сжатия, меньше, чем при парном приме-

нении автоэнкодеров и зашумления и примерно равны при парном применении JPEG-сжатия и зашумления. Это объясняется особенностями применения метода зашумления. По-видимому, этот метод, помимо противодействия состязательным атакам, все-таки оказывает негативное воздействие и на другие методы защиты. Таким образом, проведенные исследования показывают, что наибольшей эффективностью противодействия градиентным состязательным атакам обладает комбинация методов нейронной очистки и JPEG-сжатия.

Заключение

В работе исследован подход к защите систем распознавания изображений от градиентных состязательных атак, основанный на комбинировании нескольких методов защиты, в качестве которых рассматривались методы нейронной очистки с помощью автоэнкодеров, JPEG-сжатия и наложения шумов Гаусса. При этом были найдены оптимальные значения параметров методов защиты, при которых достигается наивысшая точность распознавания изображений.

Исследование показало, что наиболее эффективным является способ попарного применения нейронной очистки и JPEG-сжатия. Способ полного комбинирования методов защиты уступает по эффективности из-за негативного эффекта метода зашумления, который влияет не только на состязательные атаки, но и на другие методы защиты.

Предложенный в работе подход обладает потенциалом для дальнейшего развития. Планируется рассмотреть другие комбинации защитных механизмов, применять различные архитектуры нейронных сетей для обработки изображений, протестировать подход на дополнительных датасетах и более сложных моделях классификации, а также провести сравнение предложенного метода с гибридными методами.

Список литературы

- [Баев и др., 2024] Баев А.В., Самонов А.В., Сафонов В.М., Краснов С.В., Малышев С.Р. Методы защиты моделей нейронных сетей от состязательных атак уклонения и отравления // Автоматизация процессов управления. – 2024. – № 4 (78). – С. 39-48. – doi: 10.35752/1991-2927_2024_4_78_39.
- [Котенко и др., 2025] Котенко И.В., Сасенко И.Б., Лаута О.С., Васильев Н.А., Садовников В.Е. Метод противодействия состязательным атакам на системы классификации изображений // Вопросы кибербезопасности. – 2025. – № 2(66). – С. 114-123. – doi: 10.21681/2311-3456-2025-2-114-123.
- [Легашев и др., 2024] Легашев Л.В., Жигалов А.Ю. Исследование состязательных атак на регрессионные модели машинного обучения в беспроводных сетях 5G // Вопросы кибербезопасности. – 2024. – № 3(61). – С. 61-67. – doi: 10.21681/2311-3456-2024-3-61-67.

- [Понкин и др., 2024] Понкин И.В., Куприяновский В.П., Морева С.Л., Лаптева А.И. Компьютерное зрение: концепт, функционально-целевое назначение, структура, регуляторика // International Journal of Open Information Technologies. – 2024. – Т. 12, № 5. – С. 57-66.
- [Фомичева и др., 2023] Фомичева С.Г., Беззатеев С.В. Механизмы защиты моделей машинного обучения от состязательных атак // T-Comm: Телекоммуникации и транспорт. – 2023. – Т. 17, № 10. – С. 28-42. – doi: 10.36724/2072-8735-2023-17-10-28-42.
- [Ali et al., 2022] Ali Y., Wani M.A. HAT: Hybrid Adversarial Training to Make Robust Deep Learning Classifiers // In: Proc. 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom). – 2022. – P. 433-436. – doi: 10.23919/INDIACom54597.2022.9763284.
- [Ashraf et al., 2024] Ashraf S.N., Siddiqi R., Farooq H. Auto encoder-based defense mechanism against popular adversarial attacks in deep learning // PLoS ONE. – 2024. – Vol. 19(10). – P. e0307363. doi: 10.1371/journal.pone.0307363.
- [Aydemir et al., 2024] Aydemir A.E., Temizel A., Temizel T.T. The effects of JPEG and JPEG2000 compression on attacks using adversarial examples. 2024. arXiv:1803.10418.
- [Chen et al., 2017] Chen P.Y., Zhang H., Sharma Y., Yi J., Hsieh Ch.-J. Zoo: Zeroth order optimization based black-box attacks to deep neural networks without training substitute models // In: Proc. 10th ACM workshop on artificial intelligence and security (AISec '17). – 2017. – P. 15-26. – doi: 10.1145/3128572.3140448.
- [Chen et al., 2024] Chen Z., Chen Q., Zhou H., Zhang H. DAE-Net: Deforming auto-encoder for fine-grained shape co-segmentation // In: Proc. ACM SIGGRAPH 2024 Conference Papers (SIGGRAPH '24), Denver, CO USA, 2024. Article No. 82. – P. 1-11. – doi: 10.1145/3641519.3657528.
- [Das et al., 2017] Das N., Shanbhogue Мюб., Chen Sh.-T., Hohman F., Chen L., Kounavis M.E., Chau D.H. Keeping the bad guys out: Protecting and vaccinating deep learning with JPEG compression. – 2017. – arXiv:1705.02900.
- [Gardella et al., 2022] Gardella M., Nikoukhah T., Li Y., Bammey Q. The Impact of JPEG Compression on Prior Image Noise // In: Proc. ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Singapore, Singapore, 2022. – P. 2689-2693. – doi: 10.1109/ICASSP43922.2022.9746060.
- [Goodfellow et al., 2015] Goodfellow I.J., Shlens J., Szegedy C. Explaining and harnessing adversarial examples. – 2015. – arXiv:1412.6572.
- [Hossain et al., 2022] Hossain M.T., Badsha S., La H., Islam S., Khalil I. Exploiting Gaussian noise variance for dynamic differential poisoning in federated learning // IEEE Transactions on Artificial Intelligence. – 2022. – Vol. 1, No. 01. – P. 1-17. – doi: 10.1109/TAI.2025.3540030.
- [Jain, 2024] Jain S. Adversarial attack on Yolov5 for traffic and road sign detection // In: Proc. 2024 4th International Conference on Applied Artificial Intelligence (ICAPAI). – 2024. – P. 1-5. – doi: 10.1109/ICAPAI61893.2024.10541282.
- [Jiakai, 2018] Jiakai W. Adversarial examples in the physical world. Artificial intelligence safety and security // In: Thirtieth International Joint Conference on Artificial Intelligence (JCAI-21). – 2018. – P. 4925-4926. – doi: 10.24963/ijcai.2021/694.

- [**Khamaiseh et al., 2022**] Khamaiseh S.Y., Bagagem D., Al-Alaj A., Mancino M., Alomari H.W. Adversarial deep learning: A survey on adversarial attacks and defense mechanisms on image classification // IEEE Access. – 2022. – Vol. 10. – P. 102266-102291. – doi: 10.1109/ACCESS.2022.3208131.
- [**Kotenko et al., 2024**] Kotenko I., Saenko I., Laut O., Vasiliev N., Sadovnikov V. An approach to countering adversarial attacks on image recognition based on JPEG-compression and Neural-Cleanse // In: Proc. 2024 IEEE Ural-Siberian Conference on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT). – 2024. – P. 76-79. – doi: 0.1109/USBEREIT61901.2024.10584049.
- [**Kotlyarov et al., 2023**] Kotlyarov D.V., Dyudun G.D., Rzhetskaya N.V., Lapina M.A., Babenko M.G. Investigation of adversarial attacks on pattern recognition neural networks // Proceedings of the Institute for System Programming of the RAS. – 2023. – Vol. 35, No. 2. – P. 35-48. – doi: 10.15514/ISPRAS-2023-35(2)-3.
- [**Li et al., 2023**] Li B., Wu S., Yang Y., Zhang G. Analysis and research of neural network adversarial samples for power grid security // In: Proc. 2023 8th International Conference on Data Science in Cyberspace (DSC). – 2023. – P. 520-525, – doi: 10.1109/DSC59305.2023.00081.
- [**Naveen et al., 2023**] Naveen I.G., Inchara, Meghana H., Preethi N., Neha B. A combined approach for efficient compression and restoration of multispectral satellite images // In: Proc. 2023 International Conference on Applied Intelligence and Sustainable Computing (ICAISC). – 2023. – P. 1-5. – doi: 10.1109/ICAISC58445.2023.10200770.
- [**Tian et al., 2024**] Tian P., Poreddy S., Danda C., Gowrineni C., Wu Y., Liao W. Evaluating impact of image transformations on adversarial examples // IEEE Access. – 2024. – Vol. 12. – P. 186217-186228. – doi: 10.1109/ACCESS.2024.3487479.
- [**Xie et al., 2021**] Xie Z., Liu C., Zhang Y., Lu H., Wang D., Ding, Y. Adversarial and contrastive variational autoencoder for sequential recommendation // In: Proc. Web Conference 2021 (WWW'21). – 2021. – P. 449-459. – doi: 10.1145/3442381.3449873.

УДК 004.056.5

doi: 10.15622/rcai.2025.074

ПРИМЕНЕНИЕ МЕТОДОВ КЛАССИФИКАЦИИ И КЛАСТЕРИЗАЦИИ В ПРОЦЕССЕ РАССЛЕДОВАНИЯ ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

А.В. Мешков (*meshkov.av@itmo.ru*)

Национальный исследовательский университет ИТМО,
Санкт-Петербург

В данной статье рассматриваются способы борьбы с растущей сложностью киберугроз путем разработки методов и алгоритмов для расследования инцидентов безопасности в операционных системах Windows и Linux. Используя фреймворк MITRE ATT&CK, предлагаемый подход облегчает сопоставление событий безопасности с соответствующими тактиками, техниками и процедурами атакующих, позволяя проводить структурированную классификацию и кластеризацию. Для оптимизации анализа событий применяются передовые методы предварительной обработки и алгоритмы машинного обучения, что значительно сокращает время обработки и сохраняет точность анализа. Экспериментальные оценки подтверждают эффективность использования фреймворка в повышении возможностей обнаружения и реагирования на инциденты информационной безопасности (ИБ). Будущие направления исследований включают совершенствование производительности алгоритмов и адаптацию фреймворка к возникающим проблемам кибербезопасности.

Ключевые слова: инциденты ИБ, MITRE ATT&CK, машинное обучение.

Введение

Возрастающая сложность киберугроз обуславливает необходимость разработки современных методов расследования инцидентов информационной безопасности. Традиционные подходы к обеспечению защиты информационных систем оказываются недостаточно эффективными в условиях постоянно эволюционирующих тактик злоумышленников, эксплуатирующих уязвимости как в сетевой инфраструктуре, так и в операцион-

ных системах. Это актуализирует потребность в применении передовых технологий, обеспечивающих более высокую точность и оперативность обнаружения и анализа событий информационной безопасности.

В современных условиях особую значимость приобретает применение методов машинного обучения, обеспечивающих интеллектуальную поддержку процессов анализа и интерпретации событий информационной безопасности. Алгоритмы классификации позволяют автоматически отнести каждое событие к определённой категории, соответствующей конкретной технике или стадии атаки, что существенно ускоряет первичную фильтрацию и приоритизацию инцидентов. В то же время кластеризация предоставляет возможность выявления структурных связей между событиями, обнаружения аномалий и ранее неизвестных сценариев атак. Комбинированное применение этих методов позволяет реализовать более гибкий, адаптивный и масштабируемый подход к расследованию инцидентов, особенно в условиях кроссплатформенной среды и ограниченных временных ресурсов на реагирование.

В исследовании рассматриваются методы и алгоритмы, направленные на автоматизацию процессов расследования инцидентов информационной безопасности в информационных системах. Основное внимание уделено снижению времени анализа событий при сохранении глубины и строгости оценки. В качестве концептуальной основы используется фреймворк MITRE ATT&CK, позволяющий соотносить регистрируемые события с конкретными тактиками и техниками атак, что обеспечивает формализованный и воспроизводимый подход к расследованию.

1. Методы

На начальном этапе осуществляется систематизированный сбор данных о событиях информационной безопасности, в результате которого формируется выборка, включающая разнородные события, зафиксированные в операционных системах Windows и Linux. Данные агрегируются из нескольких источников, что позволяет обеспечить репрезентативность и разнообразие выборки – от штатной активности до потенциально вредоносного поведения. Целью данного этапа является формирование достоверного и полноценно размеченного набора данных, пригодного для разработки, обучения и тестирования алгоритмов обнаружения атак, при этом обеспечивается охват широкого спектра событий, типичных для различных операционных сред.

Формально такой набор можно описать следующим образом:

$$D = \{E_i | i = 1, \dots, N\}, \quad (1.1)$$

где E_i – отдельное событие с множеством связанных атрибутов, а N – общее количество событий в наборе данных.

На следующем этапе производится сопоставление каждого события в D с соответствующими тактиками и техниками из матрицы MITRE ATT&CK. Это позволяет осуществить формализованную классификацию событий в соответствии с известными ТТР (Tactics, Techniques, and Procedures), обеспечивая содержательную разметку выборки с привязкой к моделям поведения злоумышленников.

Пусть T – множество всех тактик и техник из MITRE ATT&CK, где каждый элемент $t \in T$ представляет конкретную технику или тактику. Тогда для каждого события E_i определяется функция $f: E_i \rightarrow T \in T$ соответствующая E_i и паре "тактика-техника" t_{ij} , что в результате:

$$f(E_i) = t_{ij}, t_{ij} \in T, \quad (1.2)$$

где t_{ij} – пара "тактика-техника", соответствующая конкретному событию E_i . Такое отображение позволяет установить семантическую связь между событиями и известными векторами атак, что является основой для последующего обучения классификационных и кластеризационных моделей.

После аннотирования набора данных метками, соответствующими техникам и тактикам MITRE ATT&CK, проводится этап предварительной обработки, направленный на подготовку данных к применению алгоритмов машинного обучения. Предобработка включает нормализацию, фильтрацию и преобразование признаков, что позволяет повысить пригодность данных для задач классификации и кластеризации. Для каждого события E_i формируется вектор признаков x_i , который проходит преобразование, обеспечивающее нулевое среднее значение и единичную дисперсию признаков:

$$x'_i = \frac{x_i - \mu}{\sigma}. \quad (1.3)$$

где μ – среднее значение, а σ – стандартное отклонение соответствующего вектора признаков. Проведение такой нормализации критически важно для корректного функционирования алгоритмов, чувствительных к масштабам данных, например, опирающихся на евклидовы расстояния.

После завершения предобработки набор данных подвергается анализу с использованием методов машинного обучения для решения задач классификации и кластеризации.

В рамках классификации используется следующее обучение: каждому событию E_i , представленному вектором признаков x_i , сопоставляется метка y_i , отражающая соответствующую технику атаки. Целью классификатора h является минимизация функции потерь L , измеряющей отклонение между предсказанным значением $h(x_i)$ и фактической меткой y_i :

$$\min_h \sum_{i=1}^N L(h(x_i), y_i). \quad (1.4)$$

В качестве классификаторов могут использоваться алгоритмы, такие как деревья решений, машины опорных векторов (SVM), нейронные сети и другие модели обучения.

Для кластеризации применяются различные методы обучения, например, алгоритм k-средних (k-means) или иерархическая кластеризация. Целью является разбиение множества событий на кластеры C_1, C_2, \dots, C_k таким образом, чтобы суммарное внутрикластерное расстояние до центров кластеров было минимизировано:

$$\sum_{j=1}^k \sum_{x \in C_j} \|x - \mu_j\|^2, \quad (1.5)$$

где μ_j – центр (среднее значение) кластера C_j . Такое разбиение позволяет выделять группы схожих событий и обнаруживать аномальные кластеры, потенциально указывающие на скрытые угрозы.

Совокупное применение методов классификации и кластеризации формирует основу предлагаемой концепции автоматизированного анализа событий информационной безопасности, обеспечивая как сопоставление с известными шаблонами атак, так и выявление новых векторов угроз в ИТ-инфраструктурах.

2. Результаты

Сформированный набор данных включает $N = 10000$ событий информационной безопасности, собранных из операционных систем Windows и Linux. Он охватывает широкий спектр как нормальной, так и аномальной активности. Каждое событие E_i было аннотировано соответствующими тактиками и техниками на основе матрицы MITRE ATT&CK, что обеспечило содержательную разметку для последующего анализа. В табл. 1 представлено распределение событий по категориям.

Таблица 1

Категория	Количество событий	Доля, %
Нормальная активность	6000	60
Подозрительная активность	2500	25
Подтверждённые угрозы	1500	15
Итого	10000	100

На этапе предварительной обработки были нормализованы все векторы признаков, что обеспечило единообразие представления данных и корректность их интерпретации алгоритмами. Среднее время предобработки одного события составило 0,01 секунды, что подтверждает масштабируемость решения для больших объёмов данных.

К полученному набору применялись как методы контролируемого обучения (для классификации), так и неконтролируемого обучения (для кластеризации). В табл. 2 приведены сравнительные характеристики эф-

фektivности трёх моделей классификации по метрикам: точность (accuracy), полнота (recall), прецизионность (precision) и среднее время обработки одного события.

Таблица 2

Модель	Точность, %	Прецизионность (Precision), %	Полнота, %	Время обработки (мс)
Дерево решений	92,4	91,2	90,8	12,5
Машина опорных векторов	95,6	94,7	94,2	22,3
Нейронная сеть	97,3	96,8	96,5	35,6

Наилучшие результаты по точности классификации продемонстрировала нейронная сеть, достигшая значения 97,3%, однако при этом характеризующаяся более высоким временем обработки одного события по сравнению с другими моделями.

Для выявления скрытых закономерностей и групп схожих событий в наборе данных была проведена кластеризация без учителя. В качестве основного метода использовался алгоритм k-средних (k-means). Результаты кластерного анализа представлены на рис. 1, где отчётливо прослеживается разделение аномальных событий от рутинной активности, что существенно упрощает идентификацию потенциальных угроз информационной безопасности и способствует более целенаправленному анализу инцидентов.

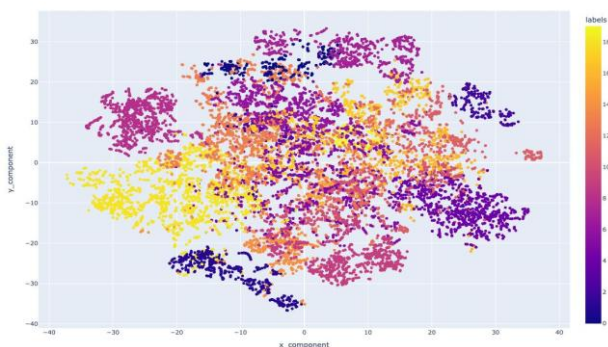


Рис. 1. Результаты кластеризации событий информационной безопасности с использованием алгоритма K-Means

Предложенные в работе методы обеспечили существенное повышение эффективности обработки событий информационной безопасности. Применение алгоритмов машинного обучения, особенно нейросетевых моде-

лей, позволило достичь высокой точности классификации и надёжного выявления атакующих действий. В свою очередь, методы кластеризации продемонстрировали способность эффективно группировать схожие события, что способствует идентификации потенциальных шаблонов атак и аномального поведения.

В совокупности полученные результаты подтверждают результативность предложенного фреймворка в задаче оптимизации процессов расследования инцидентов в информационных системах.

Заключение

Предложенные методы и алгоритмы значительно повысили эффективность обработки событий информационной безопасности. За счёт применения технологий машинного обучения среднее время анализа одного события было снижено приблизительно на 60 % по сравнению с традиционными методами ручного анализа. Этот прирост производительности обусловлен автоматизированной классификацией и кластеризацией, которые существенно упрощают идентификацию критических инцидентов и снижают когнитивную нагрузку на аналитиков.

Дополнительный вклад в точность и стабильность результатов вносит этап предварительной обработки, включающий нормализацию и извлечение признаков. Это обеспечило согласованность и надёжность входных данных, что позволило алгоритмам машинного обучения функционировать с высокой точностью. Структурированный подход к формированию выборки также способствовал беспрепятственной интеграции событий, поступающих из различных источников – как из систем Windows, так и Linux – продемонстрировав гибкость и адаптивность фреймворка в условиях гетерогенной ИТ-среды.

Интеграция алгоритмов классификации и кластеризации в единый аналитический контур позволила значительно оптимизировать процесс расследования инцидентов. Модели классификации обеспечили точное определение типа события, что, в свою очередь, способствовало оперативной приоритизации инцидентов с высоким уровнем риска. Так, нейросетевая модель достигла точности 97,3 %, продемонстрировав высокую надёжность при анализе сложных атакующих сценариев.

Кластеризационные алгоритмы дополнили классификационный анализ, обеспечив выявление скрытых закономерностей и аномальных паттернов поведения в больших объёмах данных. Полученные аналитические сведения способствовали идентификации потенциальных векторов атак, снижая вероятность пропуска критичных угроз. Совместное использование методов контролируемого и неконтролируемого обучения расширило аналитические возможности системы, одновременно повысив её общую эффективность за счёт сокращения объёмов дублирующего анализа и более целенаправленного реагирования.

Разработанный фреймворк продемонстрировал высокую результативность в обнаружении продвинутых киберугроз, включая сложные техники и тактики, формализованные в MITRE ATT&CK. Сопоставление событий с конкретными элементами этой матрицы позволило получить детализированное представление о потенциальных траекториях атаки, что повысило точность и полноту детектирования.

Оценочные метрики, такие как точность, прецизионность и полнота, подтвердили способность решений эффективно различать легитимную и вредоносную активность. В частности, алгоритм машин опорных векторов (SVM) продемонстрировал уровень полноты 94,2 %, что указывает на высокую чувствительность к настоящим позитивным срабатываниям. Алгоритмы кластеризации выявили аномальные кластеры, коррелирующие с известными векторами атак, тем самым подтвердив способность модели выявлять новые и возникающие угрозы.

В целом, предложенные методы и алгоритмы не только повысили скорость и точность обработки событий, но и усилили аналитические и диагностические возможности систем реагирования на инциденты. Полученные результаты позволяют рассматривать данный фреймворк как масштабируемое, адаптируемое и перспективное решение для противодействия растущей сложности киберугроз в условиях современной ИТ-инфраструктуры.

Список литературы

- [**Stouffer, 2011**] Stouffer K., Falco J., Scarfone K. Guide to Industrial Control Systems (ICS) Security // NIST Special Publication. – 2011. – 800-82. – P. 800-882.
- [**Soares, 2023**] Soares L. The evolution of cyber threats and its future landscape. – 2023.
- [**Shrestha, 2019**] Shrestha A., Mahmood A. Review of Deep Learning Algorithms and Architectures // IEEE Access. – 2019. – P. 1-1. – doi: 10.1109/ACCESS.2019.2912200.
- [**Scarfone, 2007**] Scarfone K., Mell P. Guide to Intrusion Detection and Prevention Systems (IDPS) // NIST Special Publication. – 2007. – 800-94.
- [**Abdelwahab, 2024**] Abdelwahab I., Hefny H., Darwish N. Enhancing cybersecurity defenses: a multicriteria decision-making approach to MITRE ATT&CK mitigation strategy // arXiv. 2024. – doi: 10.48550/arXiv.2407.19222.

УДК 004.8

doi: 10.15622/rcai.2025.075

ПРИМЕНЕНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В МНОГОПОЛЬЗОВАТЕЛЬСКИХ СТРАТЕГИЧЕСКИХ ИГРАХ

А.С. Мисник (*misnik.as@phystech.edu*)

Московский физико-технический институт, Москва

В работе рассматриваются современные методы, используемые для многоагентного обучения с подкреплением на примере многопользовательских стратегических игр. Описывается модель игры, раскрывающей тонкости кооперативного многоагентного взаимодействия. Приведены результаты экспериментов в модели описанной игры, особенности каждого из рассмотренных подходов, а также предлагаемые оптимизации, способствующие ускорению и повышению качества обучения.

Ключевые слова: обучение с подкреплением, многоагентное обучение, нейросетевые методы, графовые нейронные сети, декомпозиция вознаграждений.

Введение

Обучение с подкреплением представляет собой методику в машинном обучении, в которой агент обучается взаимодействовать с окружающей средой, принимая решения для достижения определенных целей. Агенты получают сигналы вознаграждения или наказания в зависимости от результатов своих действий, и их цель заключается в максимизации итогового вознаграждения.

Многоагентное обучение с подкреплением – это раздел обучения с подкреплением, в которой несколько агентов взаимодействуют в одной среде, обучаясь принимать решения для достижения индивидуальных или коллективных целей [Stefano, 2024]. Этот раздел представляет особый интерес в практических приложениях, поскольку многие реальные задачи сводятся к совместной работе нескольких агентов в конкурентных средах.

Однако многоагентное обучение с подкреплением также связано с рядом технических сложностей – возрастает вычислительная сложность из-за экспоненциального роста пространства состояний и действий с увели-

чением числа агентов. Важными вызовами также являются координация агентов, распределение вознаграждения в командных задачах и сбалансированное сотрудничество [Foerster, 2016]. В работе рассмотрены методы и их оптимизации, которые позволяют эффективно решать проблемы координации, масштабируемости и распределения вознаграждений в много-агентных системах, а также их применение в реальных сценариях.

1. Постановка задачи

Задачу обучения с подкреплением традиционно формализуют как задачу Марковского процесса принятия решений, задаваемого кортежем (S, A, T, R, γ) , где:

S – множество всех возможных состояний среды,

A – множество допустимых действий агента,

$T : S \times A \times S \rightarrow [0, 1]$ – модель переходов, определяющая вероятность перехода в состояние $s' \in S$ из состояния $s \in S$ при выборе действия $a \in A$: $T(s, a, s') = P(s_{t+1} = s' | s_t = s, a_t = a)$,

$R : S \times A \rightarrow \mathbb{R}$ – функция вознаграждения, определяющая ожидаемое мгновенное вознаграждение, получаемое при выборе действия a в состоянии s ,

$\gamma \in [0, 1]$ – коэффициент дисконтирования, отражающий относительную значимость будущих вознаграждений.

Цель агента: Найти оптимальную стратегию $\pi^* : S \rightarrow A$, которая максимизирует математическое ожидание суммы дисконтированных вознаграждений [Саттон, 2014]:

$$\pi^* = \arg \max_{\pi} \mathbb{E}_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(s_t)) \right]$$

В многоагентном обучении с подкреплением постановка задачи усложняется за счёт взаимодействия нескольких агентов, каждый из которых стремится максимизировать свою собственную или общую награду. Многоагентная система может быть описана кортежем

$$(N, S, \{A_i\}_{i=1}^N, T, \{R_i\}_{i=1}^N, \gamma):$$

N – количество агентов в системе,

A_i – множество допустимых действий i -го агента,

$T : S \times A_1 \times \dots \times A_N \times S \rightarrow [0, 1]$ – функция переходов, определяющая вероятность перехода в состояние s' при совместных действиях агентов,

$R_i : S \times A_1 \times \dots \times A_N \times S \rightarrow \mathbb{R}$ – функция вознаграждения i -го агента, зависящая от состояний и действий всех агентов.

Цель каждого агента i заключается в поиске стратегии π_i^* , максимизирующей его ожидаемую дисконтированную награду:

$$\pi_i^* = \arg \max_{\pi_i} \mathbb{E}_{\pi_1, \dots, \pi_N} \left[\sum_{t=0}^{\infty} \gamma^t R_i(s_t, a_{1,t}, \dots, a_{N,t}) \right],$$

где $a_{j,t} \sim \pi_j(\cdot | s_t)$ – действие агента j в момент времени t .

2. Модель рассматриваемой игры

В работе для проведения экспериментов и сравнения результатов используется модель игры из публичного соревнования, приуроченного к конференции AI Journey 2024 по теме Multiagent AI.

Среда представляет из себя клеточное поле, разделенное на 9 секторов в виде квадрата 3×3 с возможными случайными смещениями. Центральный сектор недоступен для агентов, а на остальных восьми расположен агент, его пункт переработки отходов и его пункт выдачи награды. Игра происходит по шагам, на каждом из которых в среде с некоторой вероятностью появляется ресурс. Для получения награды агенту нужно дойти до ресурса, взять его и перенести в пункт выдачи награды. После этого на секторе агента формируется отход, для устранения которого агенту нужно взять его и донести до пункта переработки.

Вероятность выпадения новых ресурсов на секторе агента зависит от “экологического показателя” – количества отходов на секторе агента и на соседних по стороне секторах. Эта особенность игровой модели лежит в основе кооперативной концепции – действия каждого агента влияют на общий успех.

Метрика, которую нужно максимизировать – средняя награда среди всех пользовательских агентов за заранее известное количество шагов игры.

Важным замечанием здесь является то, что перед началом игры случайным образом определяется, какие агенты будут управляться пользовательскими стратегиями, а какие – недоступными для пользователя стратегиями (например, системой). Это является серьезной проблемой для стратегий, которые полагаются на кооперативные взаимодействия.

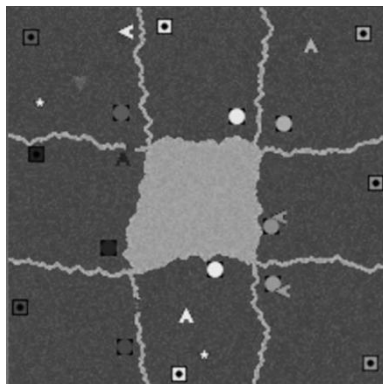


Рис. 1. Визуализация модели игры. Размер карты 210 * 210.
Стрелки – агенты, которые собирают ресурсы (звезды).
Квадраты – пункты переработки и выдачи награды

3. Нейросетевые подходы и эксперименты

В рамках исследования реализовано несколько подходов и оптимизаций к ним. При зафиксированной конфигурации среды (количество шагов, вероятность выпадений ресурсов на шагах, размер карты и прочее) предложенные методы сравнивались по следующим параметрам:

1. Награда, полученная при управлении всеми 8 агентами обученной для каждого из них стратегией.
2. Награда, полученная при управлении случайно выбранными 4 агентами, взаимодействующими с агентами, управляемыми какой-то из других стратегий (для проверки того, насколько обученные агенты зависят от кооперации).
3. Время обучения до получения стабильного уровня наград.
4. Эффективность относительно взаимодействия со средой (sample efficiency) – этот параметр не всегда коррелирует со временем обучения, поскольку эффективность использования данных может отличаться.

3.1. Метод Value Decomposition Network (VDN)

VDN [Sunehag, 2017] – алгоритм кооперативного обучения с подкреплением, основанный на разложении общей награды в сумму индивидуальных наград. Он отличается простотой реализации и может служить хорошим базовым уровнем в многих задачах многоагентного обучения. Его преимуществом относительно других многоагентных алгоритмов служит то, что стратегии агентов можно обучать многопоточно, позволяя ускорить процесс обучения на многоядерных архитектурах. Но тем не

менее, классическая реализация VDN без дополнительных оптимизаций не является наилучшим решением в поставленной задаче по нескольким причинам:

1. Поскольку агенты обучаются в предположении, что общая награда разбивается в композицию наград отдельных агентов, обученная стратегия сугубо кооперативна и, как следствие, значительно теряет качество при взаимодействии с агентами, обученными по другим стратегиям.
2. Эффективность взаимодействия со средой не оптимальна из-за параллельного обучения.

3.2. Hindsight Experience Replay (HER) при обучении VDN

Hindsight Experience Replay [Andrychowicz, 2017] – способ целеполагания в обучении с подкреплением, способствующих обучению в средах с разреженными наградами. Поскольку награда в модели рассматриваемой игры достаточно разрежена, интеграция этого метода в процесс обучения VDN способствует ускорению обучения – требуется меньше шагов симуляции, чтобы агенты обучились базовым паттернам поведения, которые смогут развивать в дальнейшем. Подобная идея была выдвинута при обучении DDPG [Zhou et al., 2023], а в данной работе она развита и адаптирована для параллельного обучения VDN.

3.3. Метод Multi-Agent Proximal Policy Optimization (MAPPO)

MAPPO – адаптация алгоритма Proximal Policy Optimization для многоагентного обучения. В сравнении с VDN, MAPPO требует больше вычислительных ресурсов для обучения - из-за отсутствия возможности параллельного обучения агентов, предоставляемой в подходе VDN. Еще одной особенностью этого подхода является высокая зависимость качества обучения от правильно подобранных гиперпараметров – например, размера примеров в батче. Но несмотря на описанные сложности, этот подход добивается наилучших результатов в сравнении с описанными ранее методами.

3.4. Оптимизация MAPPO с использованием GNN

Внедрение графовых нейронных сетей (GNN) в алгоритм MAPPO позволяет значительно улучшить координацию между агентами за счет явного моделирования их взаимодействий. В отличие от классической реализации MAPPO, где агенты обмениваются информацией через общие скрытые слои или глобальное состояние среды, GNN явно кодирует структуру взаимодействий в виде графа, где узлы соответствуют агентам, а ребра – их связям.

Кроме этого, представление взаимодействий между агентами в виде графа позволяет оценивать, насколько действия соседних агентов влияют друг на друга и насколько их наблюдения пересекаются. Эти оценки используются в процессе обучения для того, чтобы дополнительно штрафовать стратегии за чрезмерную кооперативность.

4. Сравнение подходов

Для экспериментов были зафиксированы параметры среды:

1. Поле размера $210 * 210$ разделено на сектора $70 * 70$.
2. Игровой эпизод состоит из 1000 тактов, в каждый из которых агенты совершают движение размером 7 клеток.
3. Базовая вероятность (при отсутствии мусора) появления ресурса на каждом из тактов = 0.1.

В табл. 1 приведены результаты сравнительного анализа рассмотренных методов. Время обучения приведено в условных единицах, поскольку зависит от ресурсов, использованных для обучения. Эффективность взаимодействия со средой тоже не представляется возможным оценить в абсолютных величинах, поэтому приведены относительные оценки, основанные на собранных метриках (4 – наиболее эффективно, 1 – наименее эффективно):

Таблица 1

Метод	Награда ¹ (модель управляет 8 агентами)	Награда ¹ (модель управляет 4 случайными агентами)	Ресурсоемкость обучения (в условных единицах)	Эффектив- ность взаи- модействия со средой
VDN	46.58	41.72	2.3 T	1
VDN + HER	47.30	42.29	1 T	3
MAPPO	52.43	50.89	7.8 T	2
MAPPO + GNN	56.14	54.11	8.3 T	4

Выводы, которые можно сделать из полученных результатов:

1. Методы с использованием MAPPO добиваются наилучших сравнительных результатов, обладают большей эффективностью взаимодействия со средой, но требуют больше ресурсов для обучения.
2. Приведенные оптимизации положительно влияют на процесс обучения – например, оптимизация VDN с помощью переопределения целей сильно улучшает ресурсоемкость.

Заключение

В данной работе рассмотрены современные подходы к многоагентному обучению с подкреплением и их применение в кооперативной среде, где успех агентов зависит как от их индивидуальных действий, так и от взаимодействия с другими участниками. Проведено сравнение нескольких методов: Value Decomposition Network (VDN), его оптимизированную

¹ Среднее по агентам количество ресурсов, утилизированных за эпизод.

версию с Hindsight Experience Replay (HER), а также Multi-Agent Proximal Policy Optimization (MAPPO) и его улучшенный вариант с использованием графовых нейронных сетей (GNN).

Реализованные оптимизации ранее не освещались в научных исследованиях и представляют альтернативный взгляд на известные методы многоагентного обучения.

В общем случае, выбор оптимального метода зависит от конкретной задачи, ограничений на вычислительные ресурсы и сложности взаимодействия со средой. Например, метод HER, приведенный в работе, подходит для задач с разреженной наградой. Метод MAPPO с приведенными оптимизациями способен достичь приемлемых результатов в задачах, где взаимодействие со средой не тратит много вычислительных ресурсов.

Применяющиеся на практике подходы не ограничиваются приведенными: зачастую наилучший результат показывают ансамбли моделей, сочетающие преимущества разных подходов.

Одной важной концептуальной проблемой предложенных методов является необъяснимость приведенных стратегий – поскольку нейронные сети хранят только переработанную информацию о состояниях среды и могут обмениваться данными только с помощью дистилляции, не способной передать структурную информацию о среде. Это отдельное направление для дальнейших исследований – применимость структурного обучения в многоагентных средах.

Список литературы

- [Саттон и др., 2020] Ричард С. Саттон, Эндрю Дж. Барто // Обучение с подкреплением (второе издание). – MIT Press, 2020. – 553 с.
- [Andrychowicz et al., 2017] Andrychowicz M. et al. Hindsight experience replay // Advances in NeurIPS. – 2017. – doi: 10.48550/arXiv.1707.01495.
- [Foerster et al., 2016] Foerster J. et al. // Learning to communicate with deep multi-agent reinforcement learning // Advances in NeurIPS. – 2016. – doi: 10.48550/arXiv.1605.06676.
- [Jiang et al., 2018] Jiang J. et al. // Graph convolutional reinforcement learning // arXiv preprint. – 2018. – doi: 10.48550/arXiv.1810.09202.
- [Munikoti et al., 2023] Munikoti S. et al. // Challenges and opportunities in deep reinforcement learning with graph neural networks: A comprehensive review of algorithms and applications // IEEE Transactions on Neural Networks and Learning Systems. – 2023. – doi: 10.1109/TNNLS.2023.3275593.
- [Stefano et al., 2024] Stefano V. Albrecht, Filippos Christianos, Lukas Schäfer // Multi-Agent Reinforcement Learning: Foundations and Modern Approaches – MIT Press, 2024. – 395 с.
- [Sunehag et al., 2017] Sunehag P. et al. Value-decomposition networks for cooperative multi-agent learning. – 2017. – doi: 10.48550/arXiv.1706.05296.
- [Yu et al., 2022] Yu C. et al. // The surprising effectiveness of PPO in cooperative multi-agent games // Advances in NeurIPS. – 2022. – doi: 10.48550/arXiv.2203.02155
- [Zhou et al., 2023] Zhou Y. et al. // Cooperative multi-agent target searching: a deep reinforcement learning approach based on parallel hindsight experience replay // Complex Intell. Syst. – 2023.

УДК 004.896

doi: 10.15622/rcai.2025.076

ВЫЯВЛЕНИЕ АНОМАЛИЙ В ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДАХ НА ОСНОВЕ МНОГОМАСШТАБНОГО АВТОКОДИРОВЩИКА¹

А.Ю. Пучков (*putchkov63@mail.ru*)^А

В.С. Минин (*Mininvssm@yahoo.com*)^В

^А Филиал Национального исследовательского университета «МЭИ»

в г. Смоленске, Смоленск

^В ООО «ВИСОМ», Смоленск

Предложен метод выявления аномалий в финансовых временных рядах, данные в которых отображены с различными интервалами дискретизации по времени. Новизна метода заключается в применении многомасштабного автокодировщика, на вход которого подается расширенный набор данных, включающий как значения самого временного ряда с различной временной дискретизацией, так и их первые и вторые разности. Проведенные модельные эксперименты показали, что такое расширение увеличивает ошибку автокодировщика при восстановлении входных данных с аномалиями, по сравнению с ошибкой автокодировщика без такого расширения.

Ключевые слова: выявление аномалий в данных, финансовые временные ряды, искусственные нейронные сети, автокодировщик.

Введение

В трейдинге важно вовремя реагировать не только на краткосрочные изменения котировок валют, акций, объемов торгов, разнообразных биржевых и внебиржевых инструментов, но и на более сложные и долговременные тенденции развития фондовых рынков. Перечисленные характеристики торгов (с привязкой ко времени их фиксации) отображаются в виде финансовых временных рядов. Они имеют свою специфику, не позволяющую найти единственно верный и работающий метод анализа и прогнозирования [Зиненко, 2023]. Однако результаты такого анализа представляет интерес для инвесторов-практиков, особенно в условиях развития цифровых финансов и нестабильности мировых рынков [Verma, 2024].

¹ Работа выполнена при финансовой поддержке РФФ (проект № 22-61-00096).

Алгоритмическая биржевая торговля сейчас представляет собой основной продуктом финансового рынка, так как подавляющее количество сделок полностью автоматизированы. Поэтому в алгоритмическое обеспечение торговых роботов (специальных программ, автоматизирующих биржевую торговлю) включается широкий спектр методов – начиная от классических статистических и заканчивая методами искусственного интеллекта [Pricope, 2021], [Chahuan-Jimenez, 2024], [Dorfeshan, 2022].

Одна из задач анализа финансовых временных рядов – обнаружение аномалий в данных (Anomaly detection, AD). Методы AD хорошо себя зарекомендовали при обнаружения подозрительной активности в финансовых транзакциях, потенциально сигнализирующих об отмывании денег [Gadimov, 2025].

В связи с этим разработка новых методов и алгоритмов AD в финансовых временных рядах представляет собой актуальную исследовательскую задачу, решение которой позволяет инвесторам обоснованно подходить к выбору стратегий торговли – от высокочастотного трейдинга (High Frequency Trading, HFT) до долгосрочных инвестиций в сбалансированные портфели биржевых инструментов, а торговым роботам повышать эффективность принимаемых решений.

К наиболее востребованным задачам анализа финансовых временных рядов относятся задачи прогнозирования и выявления аномалий в данных временного ряда, причем решение последней представляет особый интерес, так как позволяет своевременно выявлять недобросовестное поведение участников торгов, проявление неконтролируемых тенденций, событий.

Целью представляемого исследования являлась разработка метода выявления аномалий в финансовых временных рядах на основе применения искусственных нейронных сетей для анализа котировок биржевых инструментов, представленных с различной временной дискретизацией.

В процессе исследования был проведен обзор существующих методов выявления аномалий в финансовых временных рядах с указанием их ограничений, разработаны архитектура нейронной сети и метод ее применения для выявления аномалий в финансовых данных, проведены численные эксперименты для апробирования предложенного метода.

1. Методы выявления аномалий во временных рядах на фондовом рынке

После представления биржевых котировок на фондовом рынке в виде временного ряда задачи его анализа могут быть формализованы (абстрагированы от предметной области) и решаться с применением представления ряда как во временной, так и частотной областях методами машинного обучения [Прогнозирование, 2016a], [Sakib, 2025].

Под аномалией в данных понимается отклонение значений временного ряда от характерных для него. В зависимости от вида искомым аномалий (точечные, коллективные, аномалии контекста) выбираются методы их выявления, которые можно разделить на статистические, контролируемые, полуконтролируемые и неконтролируемые.

Статистические методы позволяют отслеживать лишь такие типовые характеристик временных рядов, как период, сезонность, циклы, тренд, корреляцию, поэтому они сейчас не входят в популярные инструменты новых методов выявления аномалий. Более глубокие закономерности в рядах способны выявлять модели нейросетевого неконтролируемого обнаружение аномалий, что важно во многих приложениях, включая финансовую аналитику, обнаружение кибервторжений, медицинскую диагностику [Sefati, 2025].

Аномалии временного ряда, в частности, выбросы, приводят к расширению возможного диапазона его значений, что используется в подходах к выявлению аномалий на основе методов интервального анализа. Так, в [Калмыков, 1986] представлены интервальные аналоги интерполяционных формул Лагранжа, Ньютона, Эрмита, применение которых для выявления аномалий в данных основывается на возрастании ошибки восстановления значений временного ряда по сравнению с нормальными данными, для описания которых формировались интервальные оценки. В случае различных интервалов дискретизации компонент временного ряда возможно применение обобщенных интервальных оценок [Стернин, 2005]. Однако интервальные подходы требуют затрат на формирование исходного описания данных в терминах интервального анализа, что повышает их трудоемкость. Поэтому методы на основе машинного обучения, автоматизирующие процесс получения моделей, сейчас находят наибольшее применение в прикладных задачах выявления аномалий в данных. Анализ эффективности различных моделей машинного обучения при выявлении аномалий в финансовых данных показывает, что модели обеспечивают дополнительную ежемесячную доходность около двух процентов [Azevedo, 2023].

Неконтролируемое обнаружение аномалий с успехом реализуется с применением архитектуры автокодировщика (Autoencoder, AE) – специального вида нейронных сетей, принцип работы которых состоит в получении на выходном слое данных, наиболее близких к входному. Это обстоятельство положено в основу обнаружения аномалий – если AE «видит» на входе знакомые данные (близкие к тем, на которых он обучался), то ошибка воспроизведения этих данных на выходе AE будет небольшая, в противном случае, если данные аномальные, ошибка будет больше, чем обычно и можно говорить об аномальности входных данных. Для исключения тривиального решения в архитектуре AE один из промежуточных

слоев имеет меньшую размерность, чем входной и выходной слою, что заставляет АЕ выявлять наиболее значимые паттерны в данных для генерации выходных данных.

Применение АЕ для выявления аномалий в финансовых данных представлено широко. Так, в [Long, 2025a] представлен АЕ на основе сети LSTM (Long Short-Term Memory), который учитывает структуру финансовых рынков и демонстрирует способность выявлять финансовые кризисы. Для этого применяется визуализация матрицы весов LSTM на тепловой карте – большее преобладание ярких цветов по сравнению с темными цветами, по мнению авторов свидетельствует о кризисных явлениях.

Идентификация аномальных колебаний цен на сельскохозяйственную продукцию с использованием сверточной нейронной сети (Convolutional Neural Network, CNN) описана в [Time, 2023a]. В этой работе изображения временных рядов (Time Series Images, TSI) классифицируются CNN. Для преобразования разреженного одномерного временного ряда цен на сельскохозяйственную продукцию в TSI применяется поле перехода Маркова (Markov Transition Field, MTF). CNN автоматически извлекает признаки из TSI, разделенные на две категории: «нормальные» и «аномальные». По заверениям авторов такая структура выявления аномалий обеспечивает точность, в среднем на 20% выше по сравнению с другими методами оценки аномальных колебаний.

Более обширный обзор архитектур и перспектив применения АЕ представлен в [Mienye, 2025].

Анализ моделей выявления аномалий в финансовых временных рядах показал, что в своем большинстве они используют данные с одним и тем же масштабом времени (временное окно одного размера), а интервал временной дискретизации постоянен. Однако, если смотреть статистику котировок, то она обычно доступна с различной временной дискретностью. Например, на сайте ProFinance котировки биржевых инструментов можно получить с дискретностью 1, 3, 5, 15 минут, 1, 2, 4 часа, один день, одна неделя. Соответственно, рассматривая краткосрочное окно данных невозможно определить аномалию, присущую более длительным временным интервалам, которая, в конечном итоге, будет сказываться и на более краткосрочных временных промежутках.

Аналогичный вывод делается в работе [A Multi-scale, 2024], где отмечается, что окно данных фиксированной длины затрудняет одновременное обнаружение как точечных аномалий, так и крупномасштабных контекстных. Там же предлагается модель с двумя АЕ, использующими механизмы внимания, для изучения закономерностей временных рядов в разных масштабах и отмечается, что она превосходит большинство существующих моделей.

Еще один вариант модели многомасштабного неконтролируемого обнаружения аномалий в многомерных временных рядах представлен в [Probabilistic, 2023a], где применяется вероятностный АЕ. Отмечается, что эксперименты с моделью на реальных наборах данных показали метрику $F1$, превосходящую современные модели выявления аномалий в данных.

Однако и в этих исследованиях анализируются только сами значения временного ряда, в то время как дополнительные характеристики, такие как первые и вторые разности, отнесенные к соответствующим интервалам дискретизации (аналоги скорости и ускорения в кинематике), как несложный результат процедуры конструирования признаков (Feature Engineering, FE), могут дать дополнительные возможности для повышения качества модели выявления аномалий в финансовых временных рядах.

2. Постановка задачи выявления аномалий в данных

Пусть доступны данные о котировках различных биржевых инструментов, представленные с различной дискретизацией по времени. Эти данные сгруппированы в матрицы D_i , $i=1, 2, \dots, N$, где N – количество биржевых инструментов. Вся совокупность данных сгруппирована в матрицу $D=\{D_i\}$. Предполагается, что данные в D являются «нормальными», то есть не аномальными.

Каждая из матриц D_i содержит K строк, где K – количество различных значений интервалов дискретизации по времени, представленные в данных о котировках. Другими словами, k -я строка матрицы D_i – это значения котировок i -го биржевого инструмента для интервала дискретизации Δt_k . В результате матрица имеет структуру: $D_i = \{D_{i,1}; D_{i,2}; \dots, D_{i,k}; \dots, D_{i,K}\}$, где «;» обозначает переход на новую строку. Предполагается, что $K = const$. Таким образом, длина k -й строки в D_i равна количеству дискретных котировок J_k для k -го интервала дискретизации, а общее количество строк равно K .

Чтобы не увеличивать размерность матриц $D_{i,k}$ за счет добавления временной оси, фиксирующей моменты получения котировок, формируется вектор интервалов дискретизации $\Delta T=\{\Delta t_k\}$. На его основе, при заданном порядковом номере котировки j и Δt_k , всегда можно определить дискретный момент времени $t_j=j\Delta t_k$, где $j=0,1, 2, \dots, J_k$, где J_k – количество дискретных котировок для k -го интервала дискретизации.

Ставится задача разработки метода выявления коллективных (групповых) аномалий в новых данных, то есть, не представленных на временных интервалах, для которых сформирована матрица D . При этом отдельно не оговариваются особенности аномалий, которые могли бы дать о них дополнительную информацию. Конкретизация задачи по-поводу коллективных аномалий снимает необходимость выявления выбросов, для которых разработано множество методов, учитывающих особенности их проявления [Pehlivan, 2024].

3. Метод решения

Включение дополнительных признаков, извлекаемых из имеющихся временных рядов на этапе препроцессинга, может повышать точность модели AD [de Jesus, 2025]. В предлагаемом методе решения поставленной задачи для выявления аномалий в финансовых рядах процедура FE заключается в формировании на основе матрицы D расширенного набора данных – матрицы DM , содержащей дискретные аналоги первой и второй производной котировок по времени.

Для i -го биржевого инструмента и для интервала дискретизации Δt_k одна строка $D_{i,k}$ матрицы D_i модифицируется в три строки:

$$DM_{i,k} = \begin{pmatrix} (D_{i,k})_1 & (D_{i,k})_2 & \dots & (D_{i,k})_{J_k} \\ (\Delta D_{i,k} / \Delta t_k)_1 & (\Delta D_{i,k} / \Delta t_k)_2 & \dots & (\Delta D_{i,k} / \Delta t_k)_{J_k} \\ (\Delta^2 D_{i,k} / \Delta t_k^2)_1 & (\Delta^2 D_{i,k} / \Delta t_k^2)_2 & \dots & (\Delta^2 D_{i,k} / \Delta t_k^2)_{J_k} \end{pmatrix}, \quad (3.1)$$

где $\Delta D_{i,k} / \Delta t_k$ – «скорость» изменения котировок i -го биржевого инструмента; $\Delta^2 D_{i,k} / \Delta t_k^2$ – «ускорение».

В результате применения (3.1) ко всем строкам D будет получена расширенная матрица $DM = \{DM_{i,k}\}$, содержащая $3NK$ строк.

Учитывая, что если $\Delta t_{k+1} > \Delta t_k$ то, J_{k+1} не совпадает с J_k , и в этом случае количество элементов в разных строках D_i будет разным. Это будет затруднять подготовку обучающих наборов данных для модели машинного обучения. Для исключения этой ситуации в процессе FE также проводится выравнивание количества элементов в строках с помощью тривиальной процедуры – заполнения $\Delta t_{k+1} / \Delta t_k$ одинаковыми значениями $D_{i,k+1}$ диапазона Δt_{k+1} . После такого «выравнивания» количество столбцов в матрице DM будет равно $J = \max(J_k | k=1, 2, \dots, K)$.

Наборы обучающих данных формируются из матрицы DM , при этом один набор X_j для момента времени t_j содержат значения котировок за промежуток времени $[t_{j-h}; t_j]$: $X_j = [DM_{j-h} \Delta t_{j-h}, \dots, DM_j \Delta t_j]$, $j=h, h+1, \dots, J$, где h – историческая глубина анализируемых данных. Величина h может подбираться экспериментально, но для ее выбора рекомендуется рассчитывать время корреляции t_r для котировок с самым большим интервалом дискретизации $\Delta t_{max} = \max(\Delta T)$ из матрицы D и брать h с учетом условия $h \Delta t_{max} > t_r$.

Проведенный выше обзор показал, что для выявления аномалий в данных, представленных в форме временных рядов, широко применяется АЕ, что обусловлено особенностями их работы. Поэтому для выявления аномалий в имеющейся постановке задачи также предложена архитектура АЕ, в основе которой лежит рекуррентная нейронная сеть LSTM.

В приведенных выше обозначениях работу АЕ можно описать зависимостью $Y_j = f(g(X_j))$, где g – алгоритм кодировщика, отображающего X_j в скрытое состояние S , f – алгоритм декодера, разворачивающая S в выходные данные $Y_j \approx X_j$.

Архитектура разрабатывалась с применением инструмента Deep Network Designer среды MatLAB и представлена на рис. 1, где так же отражено, что соединение слоев выполнено корректно (об этом свидетельствует сообщение «errors 0»).

Параметры слоев на рис. 1 указаны для входных данных при $K=4$ и $N=2$, поэтому для других входных данных эти параметры нуждаются в адаптации, однако общая архитектура слоев остается неизменной. Для снижения вероятности переобучения в архитектуре используются слои прореживания (dropout).

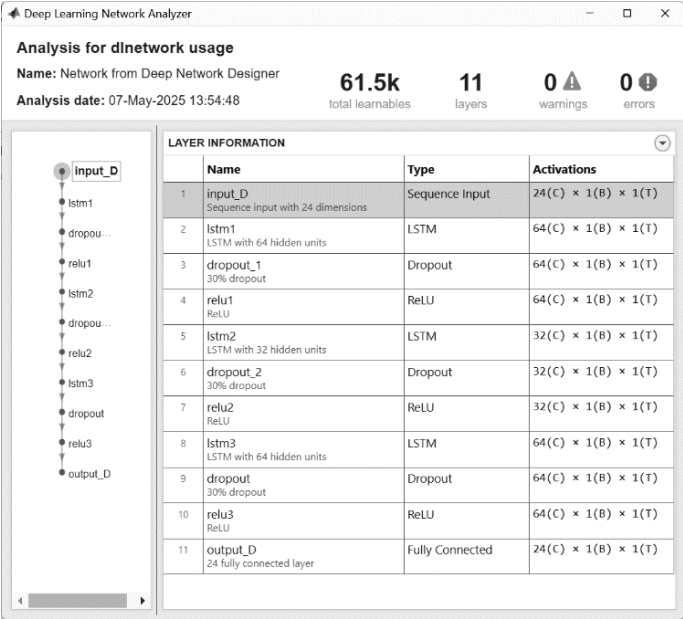


Рис. 1. Архитектура и параметры слоев АЕ

Количество входных признаков в первом слое (Sequence input) определялось исходя из формулы $3NK=3 \cdot 2 \cdot 4=24$. Общее количество обучаемых параметров АЕ в данном случае составляет 61500, но очевидно, что при других K и N количество параметров будет меняться.

Учитывая сказанное, сам метод укрупненно можно представить в виде последовательности стандартных (для применения нейронных сетей) шагов: получение из входных данных матрицы (3.1); формирование обучающих наборов; адаптация архитектуры АЕ к форме входных данных; загрузка рабочих наборов данных и применение к ним АЕ; анализ результатов.

4. Результаты и обсуждение

Для апробирования предложенного метода были проведены модельные эксперименты с использованием реальных данных о курсах американского доллара и евро с 10 января 2025 по 1 мая 2025, представленные на сайте инвестиционной компании «ФИНАМ» (<https://www.finam.ru/>). Эти данные принимались за «нормальные». Данные были взяты из столбца «close» с дискретностью 30 минут, один час, 10 часов и один день – их фрагмент для курса евро (после выравнивания количества значений), представлен на рис. 2. Из них, в соответствии с (2.1), была сформирована матрица DM , а затем обучающие наборы X_j для АЕ. Стандартизация значений элементов X_j проводится автоматически в слое Sequence input с использованием метода $zscore(X_j)$. Всего было получено 3624 значений курса валют для $\Delta t_j=30$ минут. Из этого количества 80% было взято для обучающего набора данных, а 20% – для валидационного.

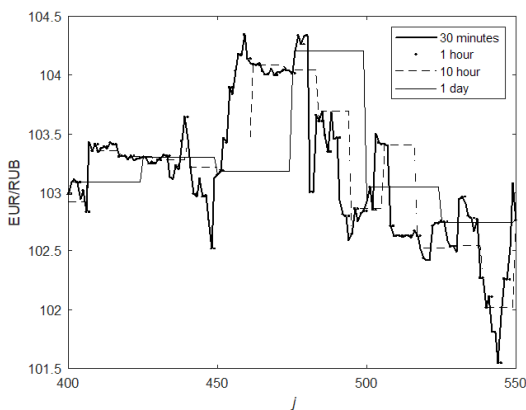


Рис. 2. Фрагмент строк с номерами 1, 4, 7 и 10 матрицы DM

Обучение модели выполнялось на персональном компьютере с процессором Intel i5-12400 (2.50 GHz) и ОЗУ 16 Gb, время обучения составило 4 минуты 45 секунд. Использовался оптимизатор adam, метрики обучения RMSE и loss, количество эпох было принято относительно небольшим – 50, что рассматривалось как еще один инструмент (помимо слоев

dropout) для борьбы с переобучением. На последней эпохе обучения RMSE достигла значений 0.098 и 0.162 на обучающем и валидационном наборах, соответственно.

На сайте «ФИНАМ» отмечается, что «возможны сбои в поставке данных и отклонения, особенно в нерабочее время», но в эксперименте, для проверки детектирующей способности АЕ, в рабочий набор исходных данных с курсами валют за период с 2 по 30 декабря 2024 года аномалии были добавлены искусственно (на рис. 3 область с коллективной аномалией обведена овалом).

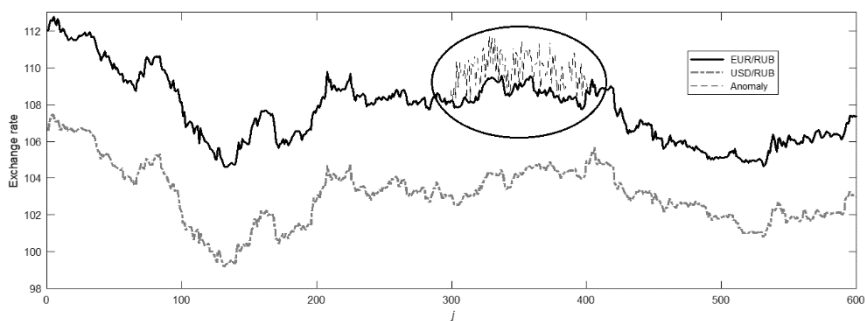


Рис. 3. Визуализация аномалии в данных

Внедренная аномалия трудно определяема статистическими методами, так как моменты различных порядков для «нормального» и «аномального» наборов будут практически одинаковыми. Применение АЕ, учитывающего «скорость» и «ускорение» изменение котировок, а также взаимосвязь с другими биржевыми инструментами (в данном случае курсом доллара) способно выявить подобную аномалию, так как в области аномалии RMSE имеет явное увеличение по сравнению с «нормальными» данными. Были проведены эксперименты с модификациями АЕ, которые обозначались цифрами: 1 – для $N = 2$, $K=4$ и с учетом первых и вторых разностей данных (далее – разностей); 2 – для $N = 2$, $K=4$, но без учета разностей; 3 – для $N = 2$, $K=1$ и с учетом разностей; 4 – для $N = 2$, $K=1$ и без учета разностей. Результаты эксперимента приведены на рис. 4, где RMSE для модификаций АЕ обозначены соответствующими подстрочными индексами.

На рис. 4 область значений j , которая соответствует аномалии в исходных данных, выделена прямоугольником. Визуальный анализ области позволяет сделать вывод, что применение АЕ на основе LSTM к многомасштабному признаковому пространству (значениям временного ряда, взятых с различными интервалами дискретизации) приводит к выраженному увеличению ошибки восстановления данных АЕ, что может быть использовано для классификации данных на аномальные и нормальные.

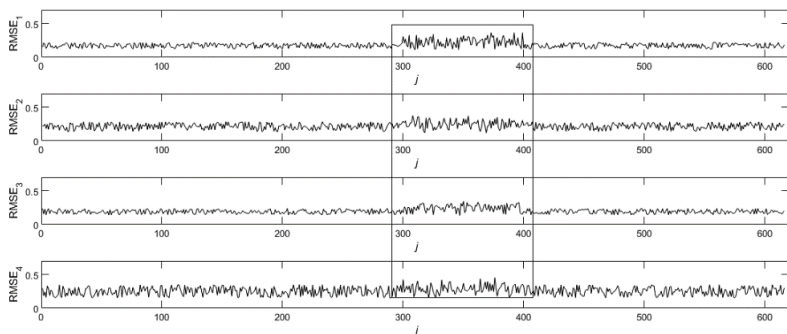


Рис. 4. RMSE на рабочем наборе данных

Для расчета метрики F1, характеризующей качество классификации данных на аномальные и нормальные, временная ось разбивалась на окна одинаковой длины. Если окно содержит аномальный паттерн (с увеличенной RMSE), все окно помечалось как аномальное. В результате, для представленных наборов данных и модификаций АЕ были получены следующие значения: для первой модификации $F1=0.84$, для второй $F1=0.79$, для третьей $F1=0.80$, для четвертой $F1=0.61$. Эти значения метрик уступают полученным в [A Multi-scale, 2024] и [Probabilistic, 2023a], где для различных наборов данных достигаются значения F1 от 0.80 и выше, однако подбор параметров N и K может сильно влиять на F1, что требует дальнейших исследований.

Заключение

В основе разработанного метода выявления аномалий в финансовых временных рядах, представленных с различной временной дискретизацией, лежит рекуррентная нейронная сеть LSTM, выполненная в архитектуре АЕ. Отличительной чертой метода выступает расширение признакового пространства АЕ за счет первых и вторых разностей многомасштабных исходных данных. Проведенные модельные эксперименты показали, что такое расширенное признаковое пространство позволяет добиться акцентированного увеличения ошибки восстановления входных данных при наличии коллективной аномалии в них.

Полученные результаты сейчас находят применение в учебном процессе по дисциплине «Интеллектуальные информационные системы» при освоении компетенций в области моделей машинного обучения. Также, полученные результаты могут быть использованы в алгоритмическом обеспечении систем автоматизации биржевой торговли.

Список литературы

- [A Multi-scale, 2024] Bao J., Gao H., Zhang C., Jia W., Gao J., Yang T. A Multi-scale Parallel Unsupervised Model for Multivariate Time Series Anomaly Detection // Artificial Intelligence Applications and Innovations. AIAI 2024. IFIP Advances in Information and Communication Technology. – 2024. – Vol 714. – doi.org/10.1007/978-3-031-63223-5_18.
- [Azevedo, 2023] Azevedo V., Hoegner C. Enhancing stock market anomalies with machine learning // Rev Quant Finan Acc. – 2023. – No. 60. – P. 195-230. – doi.org/10.1007/s11156-022-01099-z.
- [Chahuan-Jimenez, 2024] Chahuan-Jimenez K. Neural Network-Based Predictive Models for Stock Market Index Forecasting [Электронный ресурс] // J. Risk Financial Manag. – 2024. – No. 17(242). – URL: <https://doi.org/10.3390/jrfm17060242> (дата обращения: 25.04.2025).
- [de Jesus, 2025] de Jesus L.C., Fernandez-Navarro F., Carbonero-Ruz M. Enhancing financial time series forecasting through topological data analysis // Neural Comput & Applic. – 2025. – Vol. 37. – P. 6527-6545. – doi.org/10.1007/s00521-024-10787-x.
- [Dorfeshan, 2022] Dorfeshan Y., Taleizadeh A., Toloo M. Assessment of risk-sharing ratio with considering budget constraint and disruption risk under a triangular Pythagorean fuzzy environment in public-private partnership projects [Электронный ресурс] // Expert Systems with Applications. – 2022. – No. 203. – URL: <https://doi.org/10.1016/j.eswa.2022.117245> (дата обращения 20.04.2025).
- [Gadimov, 2025] Gadimov E., Birihanu E. Real-time suspicious detection framework for financial data streams // Int. j. inf. Tecnol. – 2025. – doi.org/10.1007/s41870-025-02529-6.
- [Long, 2025a] Tuhin K.H., Nobi A., Rakib M.H., Lee J.W. Long short-term memory autoencoder based network of financial indices // Humanit Soc Sci Commun. – 2025. – No. 12, 100. – doi.org/10.1057/s41599-025-04412-y.
- [Mienye, 2025] Mienye I.D., Swart T.G. Deep Autoencoder Neural Networks: A Comprehensive Review and New Perspectives // Arch Computat Methods Eng. – 2025. – doi.org/10.1007/s11831-025-10260-5.
- [Pehlivan, 2024] Pehlivan H. A novel outlier detection method based on Bayesian change point analysis and Hampel identifier for GNSS coordinate time series // EURASIP J. Adv. Signal Process. – 2024. – Vol. 2024, article number 44. – doi.org/10.1186/s13634-023-01097-w.
- [Pricope, 2021] Pricope T.-V. Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review [Электронный ресурс] // Computer Science. – 2021. – URL: [arXiv:2106.00123 \[cs.LG\]](https://arxiv.org/abs/2106.00123) (дата обращения 20.04.2025).
- [Probabilistic, 2023a] Zhang G., Gao X., Wang L., Xue B., Fu S., Yu J., Huang Z., Huang Xu. Probabilistic autoencoder with multi-scale feature extraction for multivariate time series anomaly detection // Appl Intell. – 2023. – Vol. 53. – P. 15855-15872. – doi.org/10.1007/s10489-022-04324-3.
- [Sakib, 2025] Sakib M. Ensemble deep learning techniques for time series analysis: a comprehensive review, applications, open issues, challenges, and future directions // Cluster Comput. – 2025. – No. 28(73). – doi.org/10.1007/s10586-024-04684-0.
- [Sefati, 2025] Sefati S.T., Razavi, S.N., Salehpour, P. Enhancing autoencoder models for multivariate time series anomaly detection: the role of noise and data amount // J Supercomput. – 2025. – 81, 559. – doi.org/10.1007/s11227-025-07044-w.

- [Time, 2023a] Jiang W., Zhang D., Ling L., Cai G., Zeng L. Time series to imaging-based deep learning model for detecting abnormal fluctuation in agriculture product price // Soft Comput. – 2023. – No. 27. – P. 14673-14688. – doi.org/10.1007/s00500-023-09121-9.
- [Verma, 2024] Verma A., Giri A.K., Debata B. ICT Diffusion, Financial Instability, and Shadow Economy: Panel Evidence from SAARC Economies [Электронный ресурс] // J Knowl Econ. – 2024. – URL: <https://doi.org/10.1007/s13132-024-02420-y>.
- [Зиненко, 2023] Зиненко А.В. Прогнозирование финансовых временных рядов с использованием сингулярного спектрального анализа // Бизнес-информатика. – 2023. – Т. 17, № 3. – С. 87-100. –DOI: 10.17323/2587-814X.2023.3.87.100.
- [Калмыков, 1986] Калмыков С.А., Шокин Ю.И., Юлдашев З.Х. Методы интервального анализа. – Новосибирск: Наука, 1986. – 222 с.
- [Прогнозирование, 2016a] Ярушев С.А., Федотова А.В., Тарасов В.Б., Аверкин А.Н. Прогнозирование временных рядов на основе гибридных нейронных сетей // Наука и образование: научное издание МГТУ им. Н.Э. Баумана. – 2016. – № 12. – С. 233-246. – doi: 10.7463/1216.0852597.
- [Стернин, 2005] Стернин М.Ю., Чугунов Н.В., Шепелев Г.И. Обобщенные интервальные оценки в моделях предметных областей систем поддержки экспертных решений // Труды Института системного анализа Российской академии наук. – 2005. – Т. 12. – С. 95-113.

УДК 004.852

doi: 10.15622/rcai.2025.077

МЕТОД УПРАВЛЕНИЯ ПОКРЫТИЕМ В ЗАДАЧЕ ОБУЧЕНИЯ С ДЕЛЕГИРОВАНИЕМ¹

А.В. Пономарев (*ponomarev@iias.spb.su*)

Санкт-Петербургский Федеральный исследовательский центр РАН,
Санкт-Петербург

Во многих случаях модель машинного обучения используется не автономно, а как часть более сложной системы, которая может включать экспертов-людей. Методы обучения с делегированием (learning to defer) позволяют создавать модели, учитывающие вероятность ошибок как модели машинного обучения, так и эксперта-человека, и соответствующим образом распределять образцы для максимизации общей точности системы. Однако многие такие методы не позволяют ограничивать долю образцов, назначаемых эксперту, что оказывается важным в практических приложениях, поскольку количество экспертов и их пропускная способность обычно ограничены. В данной статье предлагается простой, но эффективный эвристический метод, позволяющий накладывать ограничения на долю образцов, перенаправляемых эксперту, тем самым помогая сбалансировать метрики точности предсказаний и покрытия. Предложенный метод может использоваться совместно со многими существующими методами обучения с делегированием и обучения с отказом (rejection learning).

Ключевые слова: обучение с отказом, обучение с делегированием, человеко-машинная система, принятие решений, совместная работа человека и ИИ.

Введение

Во многих практических приложениях модели машинного обучения не используются автономно, они являются частью процесса принятия решений, включающего и эксперта-человека. При этом сама организация процесса принятия решений, объединяющая действия, совершаемые моделью и экспертом, может очень сильно разниться [Пономарев и др., 2025].

¹ Работа выполнена при финансовой поддержке РФН (проект № 24-21-00337).

В ряде работ показано [Madras et al., 2018], [Wilder et al., 2020] что организация процесса обучения с учетом всех элементов системы, в рамках которой будет применяться модель, как правило, способствует улучшению результатов, получаемых этой системой. Таким образом, модели машинного обучения, учитывающие возможность взаимодействия с человеком (например, просто воздерживающиеся от классификации образцов, в отношении которых они не уверены), особенно востребованы в ответственных и критически важных приложениях. Наиболее широко используемый метод построения таких моделей заключается в обучении модели (например, для классификации) с последующей оценкой неопределенности предсказания на этапе вывода и перенаправлении к эксперту-человеку только тех образцов, для которых модель не уверена в результатах. Существует несколько подходов к оценке неопределенности [Cordelia et al., 1995], [Gal et al., 2016], [Lakshminarayanan et al., 2017], и потенциально может быть использован любой из них. Основным недостатком данного подхода заключается в том, что он игнорирует ограниченность знаний человека (и соответствующую вероятность ошибки), рассматривая его как оракула.

Данный недостаток был скорректирован в работах, посвященных обучению с делегированием (learning to defer) [Madras et al., 2018], в которых эксперт-человек рассматривается как часть оптимизируемой системы. Это означает, что такие алгоритмы оптимизируют стратегию делегирования, учитывая не только качество предсказания модели машинного обучения в различных областях пространства признаков, но и точность эксперта-человека, которая также может различаться в разных областях этого пространства.

Большинство подходов к обучению с перенаправлением опираются на специально разработанные функции потерь, балансирующие компоненты, отвечающие за автоматический классификатор и эксперта-человека [Mozannar et al., 2020], [Verma et al., 2022], [Wilder et al., 2020]. Однако существенным ограничением многих методов обучения с делегированием является то, что они игнорируют ограниченную пропускную способность эксперта [Leitão et al., 2022].

Было также предложено несколько методов для учёта ограниченной пропускной способности эксперта, обычно через формулировку задачи обучения с перенаправлением как задачи смешанного программирования (MILP, Mixed-Integer Linear Programming), например [Alves et al., 2024], [De et al., 2020], [De et al., 2021], [Mozannar et al., 2023]. В работах [De et al., 2020], [De et al., 2021] MILP решается для обучающего набора, после чего его решение аппроксимируется дополнительной моделью (предложенная MILP-формулировка ограничена определёнными типами моделей). В [Alves et al., 2024], [Mozannar et al., 2023] MILP строится на этапе вывода, поэтому эти подходы могут лишь распределять фиксиро-

ванный набор образцов, но плохо подходят для ситуаций, когда образцы должны распределяться между моделью и экспертом по мере их поступления. Кроме того, решение MILP может быть вычислительно затратным, особенно для больших наборов данных.

В статье рассматривается классическая постановка задачи обучения с делегированием, когда имеется набор данных, содержащий для образов помимо эталонных меток (истинных классов) также экспертные метки, что в неявной форме задает модель ошибок эксперта и характеризует области его компетенций. Необходимо найти функцию, для заданного образца возвращающую либо класс, к которому относится образец, либо осуществляющую перенаправление образца эксперту. При этом в статье предлагается эффективный эвристический метод, позволяющий накладывать ограничения на долю образцов, направляемых эксперту, тем самым способствуя балансированию метрик точности и покрытия. Данный метод может использоваться совместно со многими существующими методами обучения с делегированием и обучения с отказом (rejection learning). Проведена оценка эффективности метода с использованием трёх распространённых методов обучения с делегированием и двух наборов данных – синтетического и реального, собранного с помощью краудсорсинга.

Постановка задачи

Имеется набор данных $\{(X_i, m_i, y_i)\}_{i=1}^n \sim \mathcal{D}$, где $X_i \sim \mathcal{X}$ – признаки, описывающие объекты, $y_i \in \mathcal{Y}$ – истинные метки объектов, а $m_i \in \mathcal{Y}$ – метки, присвоенные объектам экспертом (не обязательно соответствующие истинным меткам). Необходимо найти две функции – классификатор $h: \mathcal{X} \rightarrow \mathcal{Y}$ и функцию делегирования $r: \mathcal{X} \rightarrow \{0, 1\}$. Предсказание для определенного образца получается с помощью этой пары функций следующим образом:

$$\hat{y}_i = \begin{cases} h(x_i), & \text{если } r(x_i) = 1 \\ m_i, & \text{если } r(x_i) = 0 \end{cases}.$$

Поскольку доступ к экспертным оценкам может быть ограничен во время вывода, эти функции должны максимизировать качество классификации при соблюдении ограничения на количество обращений к эксперту (или долю образцов из общего множества, для которых такое обращение производится). Формально:

$$(h^*, r^*) = \arg \max_{(h, r)} \mathbb{E}_{(x, m, y) \sim \mathcal{D}} [\mathbb{I}\{r(x)h(x) + (1 - r(x))m = y\}],$$

$$\mathbb{E}[r(x)] \leq C_r$$

На практике математические ожидания в уравнении выше обычно оцениваются с помощью эмпирических метрик, вычисляемых на основе тестового набора данных, взятого из того же распределения \mathcal{D} . Математическое ожидание числа правильных ответов соответствует точности, а математическое ожидание выборов, назначенных модели h , соответствует покрытию (доля образцов, классифицированных моделью без обращения к эксперту).

Необходимость учёта ошибок как модели машинного обучения, так и эксперта приводит к естественной функции потерь, используемой для обучения h и r [Madras et al., 2018]:

$$L_{nat}(x_i, m_i, y_i, h, r) = r(x_i)\ell_m(h(x_i), y_i) + (1 - r(x_i))\ell_{exp}(m_i, y_i),$$

где ℓ_m – функция потерь модели, а ℓ_{exp} – функция потерь эксперта. Например, если это задача бинарной классификации, то ℓ_m может быть бинарной кросс-энтропией (для ℓ_{exp} ситуация несколько сложнее, см. ниже).

Однако прямое использование естественной функции потерь имеет два основных недостатка:

- В обучающих данных m представляет класс, предоставленный конечным пользователем, поэтому, если эксперт ошибается, бинарная кросс-энтропия не имеет конечного значения. На практике можно либо использовать какое-то достаточно большое значение, либо использовать другую функцию потерь (например, L_1). В любом случае, конкретное значение, соответствующее ошибке пользователя, должно быть каким-то образом масштабировано до диапазона первого члена (ℓ_m).

- Функция потерь оптимизирует модели только с точки зрения точности, не обращая внимания на то, что доступ к пулу экспертов может быть ограничен.

В литературе предложены некоторые суррогатные функции потерь (например, [Mozannar et al., 2020]), смягчающие первую проблему, но они по-прежнему оптимизируют только с точки зрения точности.

Предлагаемый метод

Предлагаемый метод основан на возможности получения оценки, отражающей относительную уверенность классификации образца моделью по отношению к классификации того же образца экспертом. Обозначим эту оценку как $score_{(h,r)}(x)$. Абсолютные значения этой оценки не имеют значения; вместо этого, оценка устанавливает порядок: если $score_{(h,r)}(x_i) > score_{(h,r)}(x_j)$, то назначение x_i модели (а не эксперту) приведёт к меньшей вероятности ошибки, чем назначение x_j модели. То есть, данная функция устанавливает предпочтительность назначения объекта модели. Конкретные примеры построения $score_{(h,r)}(x)$ для нескольких существующих методов обучения с делегированием приведены в этом разделе.

Метод состоит из трёх этапов:

- Обучение моделей h и r с использованием существующего алгоритма обучения с делегированием. Полученная пара может выполнять делегирование, обычно достигая хорошей (или оптимальной в некотором смысле) точности, но не удовлетворяя ограничениям по покрытию.

- С использованием отдельного набора данных $\mathcal{V}^c = \{(X_i^c, m_i^c, y_i^c)\}_{i=1}^n \sim \mathcal{D}$ и Алгоритма 1 (рис. 1) найти пороговое значение оценки θ , соответствующее требуемому покрытию C_r .

- Применить Алгоритм 2 (рис. 2) для классификации любых поступающих экземпляров, взятых из выборки \mathcal{D} .

Алгоритм обучения (Алгоритм 1) оценивает значения оценок для всех образцов \mathcal{V}^c , а затем рассматривает каждое значение оценки как возможное пороговое значение для назначения всех образцов с большими (или равными) значениями оценки модели, а остальных – эксперту (что выполняется с помощью функции f_{s_i} , определённой как $h(x_j)$, если $score_{(h,r)}(x_j) \geq s_i$, и m_j в противном случае). Затем он оценивает точность и покрытие каждого такого разбиения и выбирает значение оценки такое, что: по меньшей мере C_r (требуемое покрытие) образцов \mathcal{V}^c имеют большие значения оценок, а точность максимальна (среди всех значений, удовлетворяющих ограничению по требуемому покрытию).

```

Алгоритм 1 – Обучение
Входные данные:  $C_r \in [0; 1], \mathcal{V}^c, h, r$ 
Выходные данные:  $\theta$ 
for  $i \in |\mathcal{X}^c|$  do
     $s_i := score_{(h,r)}(x_i)$ 
end for
 $a := 0$ 
 $\theta := None$ 
for all  $i \in |\mathcal{X}^c|$  do
     $\tilde{a} := Accuracy(f_{s_i}(\mathcal{X}^c, m^c), y^c)$ 
     $\tilde{c} := |\{j | j \in \{1, \dots, |\mathcal{X}^c|\}, s_j \geq s_i\}| / |\mathcal{X}^c|$ 
    if  $\tilde{c} \geq C_r$  and  $\tilde{a} > a$  then
         $a := \tilde{a}$ 
         $\theta := s_i$ 
    end for
end for

```

Рис. 1. Алгоритм обучения

```

Алгоритм 2 – Вывод
Входные данные:  $x, \theta$ 
Выходные данные: значение (класс) или запрос эксперту
if  $score_{(h,r)}(x) \geq \theta$  then
    return  $h(x)$ 
else
    return  $ASK$ 
end if

```

Рис. 2. Алгоритм вывода

Алгоритм вывода (Алгоритм 2) оценивает значение функции $score_{(h,r)}(\cdot)$ заданного образца, сравнивает его с пороговым значением θ , найденным в процессе обучения, и перенаправляет его соответствующим образом. Следует отметить, что алгоритм использует только значение параметра θ , оцененное с помощью Алгоритма 1, и образец x , следовательно, он может применяться к выборкам по мере их поступления, без необходимости обработки больших объемов данных (в отличие, например, от [Alves et al., 2024]).

Рассмотрим функции оценки для нескольких алгоритмов обучения с делегированием. Простейший алгоритм основан на пороговом значении уверенности (будем обозначать его как `Threshold`). Он не имеет отдельной модели отклонения, но перенаправляет экземпляры, для которых максимальный выходной сигнал многопеременной логистической функции (softmax) [Cordelia et al., 1995] ниже определенного порога (порог обычно устанавливается для максимизации точности), к эксперту-человеку. Данный алгоритм нечувствителен к различиям компетентности эксперта в пространстве входных признаков, поэтому функция $score_{(h,r)}(x)$ может быть определена как любая мера уверенности классификатора, например, максимальное значение многопеременной логистической функции (softmax).

Другой подход заключается в прямом использовании «естественной» функции потерь для одновременного обучения двух моделей – классификатора h и функции делегирования r (`NatLoss`). В этом случае имеется отдельная функция делегирования, выход которой (сигмоидальная функция до бинаризации) задает значение, соответствующее $score_{(h,r)}(x)$.

Наконец, рассмотрим потерю, основанную на параметризации многопеременной логистической функции, предложенную в [Mozannar et al., 2020] (`SP`). В этом случае h и r моделируются с помощью одной нейронной сети с $K + 1$ выходом (где K – число классов). Выходы с 1 по K соответствуют вероятностям классов, а $(K + 1)$ -й выход соответствует перенаправлению к эксперту-человеку. Пусть $\hat{p}_i(x)$ – значение i -го выхода. В этой модели $h(x)$ определяется как $\arg \max_{k \in 1:K} \hat{p}_k(x)$, а $r(x)$ равно 1, если $\max_{k \in 1:K} \hat{p}_k(x) \geq \hat{p}_{K+1}(x)$, и 0 в противном случае. Функцию оценки $score_{(h,r)}(x)$ предлагается определить как $\max_{k \in 1:K} \hat{p}_k(x) - \hat{p}_{K+1}(x)$. Интуитивно, это отражает разницу между уверенностью классификатора и оценкой надежности эксперта-человека для рассматриваемого примера.

Экспериментальное исследование

Экспериментальное исследование проведено с использованием двух наборов данных – синтетического и набора данных с реальными метками, полученного с применением краудсорсинга (CIFAR-10H).

Синтетический набор разработан для случая бинарной классификации. Подмножество из этого набора данных представлено на рис. 3. График слева иллюстрирует два класса, а график справа использует цветовое кодирование для выделения образцов, в которых эксперт выдал правильный (зелёный) и неправильный (красный) результат. Общая идея здесь в том, что существует область (низкие значения x_2), где классы относительно хорошо разделены, и область (высокие значения x_2), где их разделение затруднено. При этом компетентность эксперта максимальна при высоких значениях x_2 и минимальна при низких. В результате можно ожидать, что при низких значениях x_2 классификация будет выполняться моделью, а при высоких — делегирована эксперту. Набор данных генерируется с помощью нормального распределения и параметр σ этого распределения управляет перекрытием между классами и ограничивает точность модели классификации (даже при низких значениях x_2). В ходе экспериментального исследования использовался синтетический набор размером 5000 образцов.

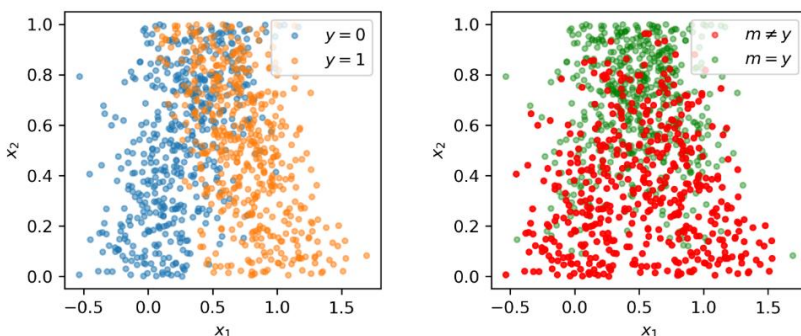


Рис. 3. Подмножество синтетического набора данных

В качестве реального набора данных используется CIFAR-10H [Peterson et al., 2019], подмножество распространенного в компьютерном зрении набора данных CIFAR-10 [Krizhevsky, 2009], содержащее 10000 изображений, для которых метки были получены с помощью краудсорсинга. Для каждого изображения в этом наборе данных имеется несколько меток, полученных от разных людей; для формирования метки m_i одна из них была выбрана случайным образом.

Разбиение на обучающее, калибровочное и тестовое множества производилось в соотношении 80/10/10. Таким образом, размеры множеств для синтетического набора данных были 4000/500/500, а для CIFAR-10H — 8000/1000/1000.

Для синтетического набора данных использовалась архитектура многослойного перцептрона (MLP) с двумя скрытыми слоями (по 40 нейронов в каждом) и функцией активации ReLU. Количество таких моделей, конфигурация выходного слоя и функция потерь варьировались в зависимости от исследуемых подходов обучения делегированию:

- для `Threshold` использовалась одна модель MLP с двумя выходными нейронами (соответствующими классам), обученная с использованием кросс-энтропии только на метках истинных значений;
- для `NatLoss` использовались две модели MLP (одна для h и одна для r), каждая с одним выходным нейроном, обученные одновременно с использованием \mathcal{L}_{nat} ;
- для `SP` использовалась одна модель MLP с тремя выходными нейронами (два из которых соответствуют классам, а один – откладыванию решения эксперту), обученная с использованием функции потерь с softmax-параметризацией из [Mozannar et al., 2020].

Во всех случаях использовался оптимизатор Adam, и обучение проводилось в пакетном режиме до сходимости (изменение значения функции потерь на обучающем множестве менее 10^{-5}).

Для набора данных CIFAR-10H была обучена модель ResNet-18 на наборе данных CIFAR-10 (исключая изображения, также входящие в CIFAR-10H), достигнув точности классификации около 86%. Затем изображения CIFAR-10H были преобразованы в их сжатые представления шириной 512, используя выход слоя, предшествующего классификационному блоку. Все модели для обучения «человек-ИИ» (классификация и стратегии делегирования) представляют собой MLP с двумя скрытыми слоями по 80 и 40 нейронов соответственно и функцией активации ReLU. Конфигурация выходного слоя и функции потерь были такими же, как описано выше (но с 10 классами).

Главный вопрос, требующий ответа в ходе эксперимента, заключается в том, позволяет ли предложенный эвристический метод накладывать ограничение на значение покрытия во время вывода. Поведение конкретной модели относительно этого требования может быть визуализировано с помощью графиков «требуемое покрытие – тестовое покрытие» (или СС-диаграммы). По оси X откладывается требуемое покрытие (задаваемое во время обучения модели), по оси Y – тестовое покрытие, оцениваемое с использованием тестового набора. Следует отметить, что в соответствии с формальным определением задачи, тестовое покрытие должно быть больше или равно необходимому покрытию, следовательно, график должен располагаться выше диагональной линии.

Рис. 4 демонстрирует примеры СС-диаграмм для синтетического набора данных (слева) и для набора данных CIFAR-10H (справа). Видно, что тестовое покрытие фактически следует за необходимым покрытием в определенном диапазоне необходимого покрытия, однако при низком необходимом покрытии тестовое покрытие оказывается значительно больше необ-

ходимого. Это происходит потому, что предлагаемый алгоритм определяет значение параметра, максимизирующее точность и удовлетворяющее ограничению на покрытие. Однако при определённом покрытии достигается максимальная точность, поэтому для всех значений необходимого покрытия, меньших этого, возвращается одно и то же тестовое покрытие (соответствующее общей наилучшей точности системы «человек-ИИ»).

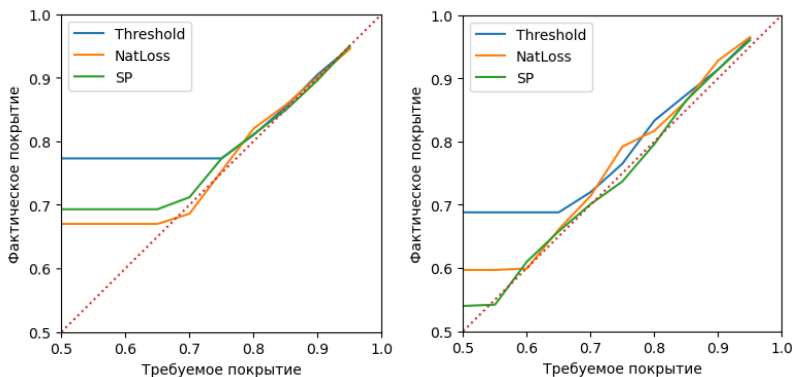


Рис. 4. СС-диаграммы для синтетического набора (слева) и CIFAR-10H (справа)

СС-диаграммы позволяют получить представление о поведении алгоритма при использовании определённой обученной модели и заданных меток, предоставленных человеком. Однако такие диаграммы обладают достаточно низкой обобщающей способностью. В связи с этим, нами также были построены диаграммы, демонстрирующие распределение максимального нарушения требования покрытия для модели. Рис. 5 иллюстрирует пример такой диаграммы, построенной для 100 сгенерированных синтетических наборов данных (и соответствующих моделей). Можно видеть, что ограничение, накладываемое на покрытие может быть нарушено, однако в подавляющем большинстве случаев это нарушение составляет менее 1%, а во всех случаях – менее 5%.



Рис. 5. Гистограмма величины нарушения ограничения на покрытие

Заключение

В статье описан эвристический метод, позволяющий накладывать ограничения на покрытие в существующих моделях обучения с делегированием. Проведена оценка предложенного метода в сочетании с тремя популярными методами обучения с делегированием: основанным на уверенности модели, на естественной функции потерь и на суррогатной функции потерь.

Вычислительные эксперименты на двух наборах данных – синтетическом и наборе CIFAR-10H, собранном методом краудсорсинга, – показали, что заданное ограничение на покрытие в основном соблюдается. Хотя возможны отдельные случаи нарушения ограничений, они встречаются нечасто и не являются значительными. Тем не менее, целесообразно провести теоретический анализ метода с целью установления теоретических границ на величину и частоту возможного нарушения задаваемого ограничения по покрытию.

Список литературы

- [Пономарев и др., 2025] Пономарев А.В., Агафонов А.А. Аналитический обзор методов распределения задач при совместной работе человека и модели ИИ // Информатика и автоматизация. – 2025. – № 1(24). – С. 229-274.
- [Alves et al., 2024] Alves J.V. et al. Cost-Sensitive Learning to Defer to Multiple Experts with Workload Constraints. Transactions on Machine Learning Research. – 2024.
- [Cordelia et al., 1995] Cordelia L.P. et al. A Method for Improving Classification Reliability of Multilayer Perceptrons // IEEE Transactions on Neural Networks. – 1995. – No. 5(6). – P. 1140-1147. – doi: 10.1109/72.410358.
- [De et al., 2020] De A., Koley P., Ganguly N., Gomez-Rodriguez M. Regression under Human Assistance // In: Proc. of the AAAI Conference on Artificial Intelligence. – 2020. – 34(03). – P. 2611-2620. – doi: 10.1609/aaai.v34i03.5645.
- [De et al., 2021] De A., Okati N., Zarezade A., Gomez-Rodriguez M. Classification Under Human Assistance // In: Proc. of the AAAI Conference on Artificial Intelligence. – 2021. – 35. – P. 5905-5913. – doi: 10.1609/aaai.v35i7.16738.
- [Gal et al., 2016] Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning // In: 33rd International Conference on Machine Learning (ICML 2016). – 2016. – P. 1050-1059.
- [Krizhevsky, 2009] Krizhevsky A. Learning Multiple Layers of Features from Tiny Images. Science Department, University of Toronto, Tech. Report. – 2009.
- [Lakshminarayanan et al., 2017] Lakshminarayanan B., Pritzel A., Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles // In: Advances in Neural Information Processing Systems. – 2017. – P. 6405-6416.
- [Leitão et al., 2022] Leitão D., Saleiro P., Figueiredo M., Bizarro P. Human-AI Collaboration in Decision-Making: Beyond Learning to Defer. In: ICML 2022, Workshop on Human-Machine Collaboration and Teaming, 2022. ArXiv: arXiv:2206.13202.

- [Madras et al., 2018]** Madras D., Pitassi T., Zemel R. Predict Responsibly: Improving Fairness and Accuracy by Learning to Defer // In: 32nd Conference on Neural Information Processing Systems. – 2018. – P. 6150-6160.
- [Mozannar et al., 2023]** Mozannar H. et al. Who Should Predict? Exact Algorithms For Learning to Defer to Humans // In: Proceedings of Machine Learning Research. – 2023. – 206. – P. 10520-10545.
- [Mozannar et al., 2020]** Mozannar H., Sontag D. Consistent estimators for learning to defer to an expert // In: Proc. of the 37th International Conference on Machine Learning. – 2020. – P. 7076-7087.
- [Peterson et al., 2019]** Peterson J. et al. Human uncertainty makes classification more robust // In: Proc. of the IEEE International Conference on Computer Visio. – 2019. – P. 9616-9625.
- [Verma et al., 2022]** Verma R., Nalisnick E. Calibrated Learning to Defer with One-vs-All Classifiers // In: Proc. of the 39th International Conference on Machine Learning. – 2022. – PMLR 162:22184-22202.
- [Wilder et al., 2020]** Wilder B., Horvitz E., Kamar E. Learning to Complement Humans // In: Proc. of the 29th International Joint Conference on Artificial Intelligence (IJCAI-20). – 2020. – P. 1526-1533.

УДК 004.89

doi: 10.15622/rcai.2025.078

АВТОМАТИЧЕСКОЕ ПОПОЛНЕНИЕ ТАКСОНОМИЙ НА РУССКОМ ЯЗЫКЕ С ПОМОЩЬЮ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ¹

Ф.А. Садковский (*sadkovsky@iling-ran.ru*)^{A,B}

Н.В. Лукашевич (*louk_nat@mail.ru*)^A

И.Ю. Гришин (*grishin@sev.msu.ru*)^A

^A Московский государственный университет им. М.В. Ломоносова,
Москва

^B Институт языкознания РАН, Москва

В настоящей работе впервые исследуется потенциал больших языковых моделей для задачи автоматического пополнения таксономий на материале русского языка. Авторы адаптировали передовую методику TaxoLLaMA, которая ранее показала высокую эффективность для английского языка, используя для этого данные русскоязычного тезауруса RuWordNet. В рамках исследования было проведено сравнение производительности мультязычных моделей и моделей, специально адаптированных для русского языка. Эксперименты подтвердили успешную применимость метода к русскоязычным данным и выявили значительное преимущество русскоязычных моделей. После дообучения модель YandexGPT-5-8B-Lite превзошла лучший предыдущий результат, достигнув значения MAP 55.42.

Ключевые слова: таксономия, пополнение таксономий, большие языковые модели, RuWordNet, TaxoLLaMA, RuAdapt.

Введение

Большие языковые модели (LLM) позволили достичь значительного прогресса в решении многих задач обработки естественного языка (NLP) благодаря их способностям к генерации контекстно релевантных терминов и выявлению семантических связей.

¹ Исследование выполнено при финансовой поддержке Междисциплинарной научно-педагогической школы Московского университета (грант № 23-Щ05-11) и государственного задания (регистрационный № 124020100068-4).

LLM показали свою эффективность и в задаче пополнения таксономий, которая ранее решалась с помощью извлечения текстовых шаблонов [Hearst, 1992], [Sabirova et al., 2014], применения векторных моделей или комбинации нескольких методов [Nikishina et al. 2020a,b, 2022]. В недавней работе [Moskvoretiskii et al., 2024] авторам удалось достигнуть передовых результатов в задачах извлечения гиперонимов и построения таксономий, а также сопоставимых с передовыми результатами в задаче пополнения таксономий благодаря предложенному методу TaxoLLaMA – обучению модели LLaMA-2-7b на материале пар гипоним–гипероним, извлеченных из тезауруса WordNet [Miller, 1990].

Тем не менее, тестирование подхода происходило почти исключительно на англоязычных данных: другие языки (испанский, итальянский) были представлены лишь в задаче извлечения гиперонимов, но и в этом случае с помощью дообучения удалось превзойти результаты предыдущих подходов.

В то же время, порождающие большие языковые модели и, в частности, подход TaxoLLaMA, ранее не применялись к данным на русском языке. Существующие подходы на основе векторных моделей [Nikishina et al. 2020a,b, 2022] не позволяют обеспечивать качество, достаточное для автоматизации задачи. Это представляется упущением, поскольку имеется ряд современных LLM, которые были специально разработаны для использования на русском языке – и среди них, модели, созданные по методике языковой адаптации RuAdapt [Tikhomirov et al., 2023, 2024], эффективность которых в других задачах NLP на русском языке выше, чем у мультязычных аналогов.

В настоящей работе исследуется возможность применения LLM к данным таксономии RuWordNet [Loukachevitch et al., 2016].

- В работе была доказана переносимость метода TaxoLLaMA на русскоязычные данные.
- Сравниваются мультязычные модели и модели, предназначенные для работы с русским языком; вторые показали большую эффективность.
- Дообучение модели YandexGPT-5-Lite-8B-pretrain [Яндекс, 2025] позволило получить передовое качество решения задачи.

1. Обзор литературы

Задача пополнения таксономии изначально формулировалась как задача извлечения гиперонима для нового понятия. Одним из первых подходов к извлечению гиперонимов для слов русского языка был предложен в работе [Sabirova et al., 2014], в которой текстовые шаблоны из [Hearst, 1992] были переведены на русский язык, дополнены синонимичными выражениями и протестированы на русской версии веб-онтологии DBPedia.

Более поздние работы, посвященные разработке и тестированию моделей на русском материале, связаны с соревнованием RUSSE'2020 [Nikishina et al., 2020a], где был впервые предложен датасет диахронического типа на основе RuWordNet [Loukachevitch et al., 2016]. Лучшие результаты были достигнуты с помощью подхода на основе большого набора признаков из внешних источников (перевод, поиск в интернете, Вики-словарь).

В последующих работах исследуется применение разнообразных векторных представлений-источников для предсказания ближайшего гиперонима в RuWordNet [Nikishina et al., 2022]. Самые высокие показатели достигаются использованием мета-эмбедингов, обученных на основе дистрибутивного (контекстного) и графового представления слов.

При этом, кроме модели-кодировщика RuBert [Kuratov et al., 2019], применение больших языковых моделей не было отмечено применительно к русским данным. Это видится серьезным упущением, поскольку в настоящий момент существует как большое количество многоязычных языковых моделей, так и моделей, специально обученных для понимания инструкций на русском языке. Среди них можно отметить модели YandexGPT компании «Яндекс», а также модели, созданные в рамках исследовательского направления по адаптации больших языковых моделей на русский язык RuAdapt [Tikhomirov et al., 2023, 2024]. Последний подход предполагает замену токенизатора на униграммный для обеспечения лучшего соответствия между токенами и русскими морфемами [Tikhomirov et al., 2023], замену входных представлений и дообучение на коллекции очищенных русскоязычных корпусов DaruLM и датасете инструкций на русском Darumeru [Tikhomirov et al., 2024].

2. Данные

Источником данных в настоящей работе служит диахронический датасет на основе RuWordNet, впервые предложенный в [Nikishina et al., 2020a] и позже расширенный в [Nikishina et al., 2020b].

В подобных датасетах в качестве обучающей подвыборки используется более ранняя версия ресурса (RuWordNet 1.0), а в качестве тестовой – слова, добавленные в более позднюю («RuWordNet 2.0–1.0»).

В [Nikishina et al., 2020a] было предложено разделение новых слов RuWordNet на две подвыборки – «public» (валидационная) и «private» (тестовая). Тестовая выборка использовалась для окончательной оценки качества предложенных решений и включает 1 525 синсетов (множеств синонимов); валидационная выборка содержит 763 синсета. В качестве правильных ответов к целевым словам были извлечены не только непосредственные гиперонимы, но и гиперонимы второго порядка.

В RuWordNet узлы графа представляют собой синсеты. Названия синсетов могут быть одним словом, словосочетанием или включать перечисление слов/словосочетаний через запятую, уточняющие слова в скобках. Другая особенность обусловлена тем, что RuWordNet наследует названия синсетов от тезауруса Рутез [Loukachevitch et al., 2002], семантическая сеть которого не подразделялась на подграфы разных частей речи. Таким образом, в зависимости от семантики слова, названия синсетов для существительных могут быть не только существительными, но и глаголами («буллинг»>«оскорбить») и прилагательными («консонанс»>«благозвучный»).

3. Метод

Оригинальный подход TaxoLLaMA, предложенный в [Moskvoretskii et al., 2024], включает следующие компоненты:

- 1) подготовка выборки на основе пар гипоним–гипероним из WordNet; тестовые понятия, входящие в датасеты для тестирования, удаляются;
- 2) дообучение модели LLaMA-2-7b [Touvron et al., 2023] на предсказание списка гиперонимов через запятую;
- 3) оценка качества на датасетах для тестирования.

В настоящей работе предлагается расширение этого метода на другие данные и модели:

1. В качестве обучающего множества берутся пары гипоним–гипероним из RuWordNet 1.0.
2. Различные 7-8 млрд-ные LLM (перечислены в след. разделе) обучаются предсказывать слова через разделитель «; », поскольку этот знак не встречается в названиях синсетов.
3. Оценка качества на диахроническом датасете «RuWordNet 2.0–1.0» (выборка «pouns-private»).

Перед оценением необходимо сопоставить предсказанной последовательности слов названия синсетов. Для этого в работе предложена следующая процедура:

1. В начало списка переносятся последовательности слов, соответствующие существующим в таксономии названиям синсетов.
2. Синсет переводится в векторный вид с помощью L2-нормализации (по [Bollegalo, Bao, 2017]) векторных представлений всех неслужебных слов, входящих в название синсета.
3. Max-Pooling: $k = 15$ ближайших соседей из множества векторных представлений синсетов были найдены для каждого из нерелевантных кандидатов, а затем среди объединенного множества размера $k \times m$ ($m \leq 15$ – число предсказанных кандидатов) были извлечены все названия синсетов, встречающиеся более 1 раза, и упорядочены по частоте и по позиции первого появления в общем списке $\langle x_{1,1}, \dots, x_{1,k}, \dots, x_{m,1}, \dots, x_{m,k} \rangle$.

Данный метод пост-обработки использует преимущества организации RuWordNet и позволяет достигнуть более высокого итогового качества, чем использование предсказаний в чистом виде. В качестве векторной модели использовалась модель FastText² [Bojanowski et al., 2017].

4. Эксперименты

4.1. Метрики

В задаче пополнения таксономии используются две основные метрики: Mean Average Precision, MAP (4.1) и Mean Reciprocal Rank, MRR (4.2):

$$MAP@K = \frac{1}{N} \sum_{n=1}^N AP@K_n, \quad (4.1)$$

$$AP@K = \frac{1}{R} \sum_{k=1}^K (Precision@k \cdot I[y_k = 1]),$$

$$Precision@k = \frac{p}{k},$$

$$MRR@K = \frac{1}{N} \sum_n \frac{1}{rank_i}. \quad (4.2)$$

Обозначения: $K = 15$ – максимальное число предсказаний; N – размер выборки; R – количество правильных ответов (гиперонимов) для целевого слова; k – индекс элемента в списке предсказаний; I – индикаторная функция; y_k – класс (верный ответ/неверный ответ) на позиции k , p – количество релевантных элементов среди k первых предсказанных; $rank_i$ – позиция первого релевантного элемента.

Используемая в ряде работ (в частности, [Moskvoretskii et al., 2024]) модификация Scaled MRR не используется, поскольку она предназначена для выборок, включающих только непосредственные гиперонимы.

4.2. Базовые методы

В качестве базовых методов используются базовый метод на основе векторной модели FastText (*fasttext*) и лучший метод на основе мета-эмбедингов (AAEME (*words* + *TADW*), усредненная сумма графовых + контекстных представлений) из [Nikishina et al., 2022].

² Предобученные векторы были взяты с официального сайта <https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.ru.300.bin.gz>.

4.3. Модели

Для экспериментов были отобраны большие языковые модели типа декодировщик с числом параметров, приблизительно соответствующим числу параметров исходной модели LLaMA-2 – 7–8 млрд. Выбирались модели в открытом доступе, не проходившие инструктивное дообучение:

- LLaMA-2-7B-hf³;
- LLaMA-3.1-8B⁴;
- Mistral-7B-v0.1 [Jiang et al., 2023];
- Mistral-7B-v0.3⁵;
- Qwen2.5-7B [Yang et al., 2024];
- Qwen3-8B-Base⁶;
- Gemma-7b [Team Google, 2024].

В список также была включена оригинальная модель TaxoLLaMA: результаты ее дообучения позволяют судить о том, насколько модель способна переносить знания о таксономических отношениях с одного языка на другой.

Другая группа моделей включала базовые модели, специально предназначенные для работы с русским языком и примерно с тем же количеством параметров:

- YandexGPT-5-Lite-8B-pretrain [Яндекс, 2025];
- Модели из серии RuAdapt [Tikhomirov et al., 2023, 2024]:
 - RuAdaptQwen2.5-7B-Lite-Beta;
 - RuAdaptLLaMA-2-7b;
 - RuAdaptMistral-v0.1;
 - RuAdaptLLaMA-3.1-8B.

4.3. Результаты

Результаты моделей⁷ приводятся в табл. 1⁸. Среди мультиязычных моделей, наилучшие результаты продемонстрировали модели семейства Qwen, в особенности Qwen3-8B-Base. Тем не менее, по метрике MAP даже лучшие модели не смогли превзойти базовые методы.

Модели для русского языка в среднем показали результаты выше: 45.02 против 38.12 по MRR и 32.8 против 23.21 по MAP. Все модели, прошедшие адаптацию на русский язык, показали более высокое качество, чем их оригинальные неадаптированные версии, см. табл. 2.

³ <https://huggingface.co/meta-llama/Llama-2-7b-hf>.

⁴ <https://huggingface.co/meta-llama/Llama-3.1-8B>.

⁵ <https://huggingface.co/mistralai/Mistral-7B-v0.3>.

⁶ <https://huggingface.co/Qwen/Qwen3-8B-Base>; <https://qwenlm.github.io/blog/qwen3/>.

⁷ Веса дообученных моделей доступны на сайте: <https://huggingface.co/TaxoLLMs>.

⁸ Здесь и далее значения метрик умножены на 100.

Модель на основе Qwen2.5-7B оказалась лидером как по абсолютному результату, так и по приросту метрик, который дает методика адаптации. Следующей моделью по значению метрик оказалась модель на основе LLaMA-3.1-8B, а следующей моделью по приросту качества – модель на основе Mistral-7B-v0.1.

Среди всех моделей самого высокого результата по метрикам MRR и MAP достигла модель YandexGPT-5-Lite-8B-pretrain. Значение метрики MAP 55.42 превосходит значение 47.4 лучшего подхода из [Nikishina et al., 2022b]. Вторую позицию занимает модель RuAdapt-Qwen2.5-7B с результатом 41.69. Этот результат не смог превзойти лучший предыдущий результат, однако оказался выше, чем у базового решения (41.4).

Самого низкого качества достигла дообученная на русских данных модель TaxoLLaMA. Значения метрик оказались даже ниже, чем у исходной для нее LLaMA-2-7b-hf, что говорит о неспособности модели эффективно переносить таксономические знания из WordNet на данные RuWordNet.

Таблица 1

Сравнение качества моделей после дообучения на «RuWordNet 1.0 – 2.0». Жирный шрифт – лучший результат, подчеркивание – второй после лучшего

модели	MRR	MAP
LLaMA-2-7b-hf	25.01	17.35
LLaMA-3.1-8B	35.11	25.35
Mistral-7B-v0.1	32.94	22.32
Mistral-7B-v0.3	32.73	21.62
Qwen2.5-7B	38.2	29.43
Qwen3-8B-Base	40.06	30.06
Gemma-7b	30.79	23.44
TaxoLLaMA	24.10	16.09
YandexGPT-5-Lite-8B-pretrain	61.55	55.42
RuadaptQwen2.5-7B-Lite-Beta	<u>49.47</u>	41.69
RuAdaptLLaMA-2-7b	30.82	25.11
RuAdaptLLaMA-3.1-8B	42.32	34.79
RuAdaptMistral-v0.1	40.95	33.57
<i>fasttext</i>	–	41.40
<i>AAEME triplet loss (words</i>	–	<u>47.40</u>

Таблица 2

Сравнение результатов моделей с адаптацией на русский язык по методике RuAdapt [Tikhomirov et al., 2023, 2024] и без

модель	оригинал		RuAdapt	
	MRR	MAP	MRR	MAP
LLaMA-2-7b-hf	25.01	17.35	30.82 _{+5.81}	25.11 _{+7.76}
LLaMA-3.1-8B	35.11	25.35	<u>42.32</u> _{+7.21}	<u>34.79</u> _{+9.44}
Mistral-7B-v0.1	32.94	22.32	40.95 _{+8.01}	33.57 _{+11.25}
Qwen2.5-7B	38.2	29.43	49.47 _{+11.27}	41.69 _{+12.26}

5. Анализ ошибок

В тестировании на русском языке можно выделить 7 основных типов ошибок. Эти типы перечислены ниже и проиллюстрированы примерами.

1. Неверно определена предметная область. Слово «авуары» (первое значение по [БТС, 2003] – «средства, за счёт которых производятся платежи и погашаются обязательства») относится к сфере финансов, однако модель Qwen3-8B-Base предсказывает гиперонимы из области этнологии: «этническая общность; этнические народы; африканцы (население); население государства; нация» при правильных «денежная единица; свободно конвертируемая валюта; банковский вклад».

2. Другие семантические отношения. Примером может служить ответ модели RuAdaptQwen2.5-7B для целевого слова «пивная»: модель через корень связала это слово с алкогольным напитком, но предсказала кандидаты, подходящие для других однокоренных слов – «пивоварни» и «пива»: «предприятие пищевой промышленности; алкогольный напиток; спиртосодержащая продукция».

3. Слишком абстрактные кандидаты. К таким примерами относится, в частности, предсказание YandexGPT-5-Lite для слова «перевязь» – «изделие легкой промышленности; изделие; полоса (кусок); кусок (отдельная часть); повязка, повязка на теле;...» при верном ответе «медицинская повязка; медицинская продукция».

4. Слишком конкретные кандидаты. Такие случаи связаны с неполнотой ресурса, а не с тем, как модель выявляет семантические связи. Примером может служить предсказание Qwen3-8B-Base для слова «бойскаут» – «член организации; участник; человек по роли; учащийся; подросток;...» при правильном ответе «мальчик; мужчина, человек мужского пола; скаут; несовершеннолетние дети». Определение этого слова по словарю [Крысин, 2018] – «мальчик или подросток – член скаутской (см. скаут) организации», из чего можно сделать вывод о релевантности кандидата на первой позиции.

5. Ошибки, связанные с многозначностью. Предсказание гиперонимов для слов, обозначающих несколько понятий одновременно, вне контекста использования также составляет проблему. В частности, в тестовую выборку входит слово «лигатура», имеющее следующие гиперонимы: (1) «хирургическое оборудование», «нить (предмет)», (2) «металл», «добавление (то, что добавлено)» и (3) «графический знак», «сочетание, комбинация». Моделям YandexGPT-5-8B-Lite и Qwen3-8B-Base удалось предсказать только последнее значение: модель серии YandexGPT смогла предсказать непосредственный гипероним «графический знак», а также «знак, обозначение», а модели серии Qwen3 — гипероним только второго порядка («изображение (результат)»).

6. Интерференция другого языка в русский текст. До обработки вывода для извлечения синсетов, среди кандидатов могут попадаться как частичные, так и полные замены слов: высший «законодательный organ» (LLaMA-3.1-8B); «органическое glass», «geometric solid», «combinatorial analysis», «травянистое植物» (Qwen2.5-7B).

7. Неверно предсказана часть речи в названии синсета. Например, модель RuAdapt-Qwen2.5-7B для слова «осиплость» предсказала гиперонимы «измениться, изменение; изменить, сделать иным;...», вероятно, связав их с событийной семантикой корня «осип-» — ‘стать хриплым’. При этом в RuWordNet для данного слова требовалось предсказать признак: «хриплый; глухой, глухо звучащий».

6. Обсуждение результатов

Предложенные методы использования мультязычных LLM с числом параметров 7–8 млрд оказались недостаточно эффективными в применении к таксономии на русском языке. В то же время, использование специализированных русскоязычных моделей продемонстрировало свой потенциал. В частности, максимальное качество, значительно превосходящее предыдущие векторные подходы, удалось получить благодаря дообучению модели YandexGPT-5-Lite-8B-pretrain.

Достоверно неизвестно, насколько этот результат является «чистым» в смысле проникновения данных RuWordNet в обучающую выборку этой модели. Однако можно привести ряд доводов против такого предположения.

- Разница в качестве между предсказанием до сопоставления с синсетами с помощью модели FastText и после (около 3 по MAP) не отличается от разницы, наблюдаемой у моделей RuAdapt, про которые достоверно известно, что RuWordNet не был включен в обучающие данные. Если бы модель специально дообучалась на RuWordNet, можно было бы ожидать меньшего выигрыша от использования нерелевантных (по форме) предсказаний.

- В официальной статье [Яндекс, 2025] нет сведений о том, что RuWordNet как-либо использовался для обучения или тестирования.
- В той же статье показано, что на основных датасетах для сравнения качества моделей YandexGPT-5-Lite достигает паритета с аналогами или превосходит их.

Таким образом, естественно ожидать, что дообучение данной модели в целом эффективно для широкого спектра задач обработки русского языка.

Лучший результат среди протестированных моделей серии RuAdapt достигнут моделью на основе Qwen2.5-7B, которая превзошла базовый метод *fasttext* по метрике MAP. Из неадаптированных моделей самого высокого качества достигла модель Qwen3-8B-Base, для которой адаптированная на русский язык версия на момент написания работы еще не была выпущена. Из адаптированных моделей ей удалось обойти только RuAdapt-LLaMA-2-7b, однако исходя из наблюдаемой тенденции можно ожидать, что качество адаптированной модели может быть сопоставимо с результатами модели от компании Yandex.

Что касается ошибок моделей, то более слабым моделям свойственно делать серьезные семантические ошибки (1–2 типы), а более сильным моделям — менее серьезные (3–4 типы), которые не являются ошибками в идеологическом смысле. Это указывает на то, что применение LLM к данной задаче весьма перспективно, хотя нуждается в дальнейшем совершенствовании методологии.

Можно предложить следующие поправки к методологии, которые будут реализованы в дальнейших исследованиях:

- 1) использовать в инструкции контекст с новым словом; многозначным словам должны соответствовать разные контексты;
- 2) использовать модель не как генератор, а как переранжировщик (англ. *Reranker*): согласно ряду исследований, эта роль — более сильная стороны генеративных больших языковых моделей, чем извлечение закономерностей из примеров (*In-context learning*) [Ma et al., 2023];
- 3) подходы 1 и 2 можно совмещать, представляя задачу пополнения таксономии как задачу генерации, дополненной поиском (англ. *Retrieval Augmented Generation*, RAG): LLM получает на вход источники информации (список потенциальных гиперонимов, контексты из корпуса, определения из внешних ресурсов), принимает решения о релевантности и выдает итоговый упорядоченный список.

Заключение

В статье была впервые рассмотрена возможность применения порождающих больших языковых моделей к решению задачи пополнения таксономии на русском языке. На диахроническом датасете RuWordNet были обучены и протестированы как применявшиеся в предыдущих частях мо-

дели, так и специально адаптированные под русский язык версии некоторых из них, а также модель YandexGPT-5-8B-Lite, также предназначенная для использования преимущественно на русском. В работе использована методология дообучения, разработанная на основе передового подхода TaxoLLaMA с учетом специфики данных.

В результате двум дообученным моделям – YandexGPT-5-8B-Lite и RuAdaptQwen2.5-7B-Lite – удалось превзойти показатели предыдущего базового метода, а модели YandexGPT-5-8B-Lite, достигшей результата 55.42 по метрике MAP, также превзойти и результат наилучшего предыдущего метода с результатом 47.4. Также на материале настоящей работы удалось дополнительно продемонстрировать преимущество методики RuAdapt для русского языка, поскольку все модели, созданные по этой методике, после дообучения на датасете RuWordNet продемонстрировали лучшее качество, чем их оригинальные версии, после аналогичного дообучения.

Список литературы

- [БТС, 2003] Большой толковый словарь / гл. ред. С. А. Кузнецов. – СПб.: Норинт, 2004. – 1534 с.
- [Крысин, 2018] Крысин Л.П. Современный словарь иностранных слов. – М.: АСТ, 2018. – 416 с.
- [Яндекс, 2025] Встречаем YandexGPT 5 – в Алисе, облаке и опенсорсе / Блог компании Яндекс [Электронный ресурс] // Хабр. – 2025. – URL: <https://habr.com/ru/companies/yandex/articles/885218s> (дата обращения: 15.06.2025).
- [Bollegala et al., 2017] Bollegala D., Bao C. Learning word meta-embeddings by autoencoding // In: Proceedings of the 27th international conference on computational linguistics. – 2018. – P. 1650-1661.
- [Bojanowski et al., 2017] Bojanowski, P., Grave, E., Joulin, A., Mikolov, T. Enriching word vectors with subword information. Transactions of the association for computational linguistics. – 2017. – Vol. 5. – P. 135-146. – doi: 10.1162/tacl_a_00051.
- [Hearst, 1992] Hearst M.A. Automatic acquisition of hyponyms from large text corpora // In: Proc. COLING 1992 volume 2: The 14th international conference on computational linguistics, Nantes, France, August 1992. – P. 539-545.
- [Jiang et al., 2023] Jiang A. Q. et al. Mistral 7b. Computer research repository. – 2023. – doi: 10.48550/arXiv.2310.06825.
- [Kuratov et al., 2019] Kuratov Y., Arkhipov M. Adaptation of deep bidirectional multilingual transformers for Russian language. Computer research repository. – doi: 10.48550/arXiv.1905.0721
- [Loukachevitch et al., 2016] Loukachevitch N.V., Gerasimova A.A., Dobrov B.V., Lashevich G., & Ivanov V.V. Creating Russian wordnet by conversion // In: Computational Linguistics and Intellectual Technologies, Moscow, Russia, June, 2016. – P. 405-415.
- [Ma et al., 2023] Ma Y., Cao Y., Hong Y., & Sun A. Large language model is not a good few-shot information extractor, but a good reranker for hard samples!. Arxiv. – 2023. – doi: 10.48550/arXiv.2303.08559

- [**Miller et al., 1990**] Miller G.A., Beckwith R., Fellbaum C., Gross D., Miller K.J. Introduction to WordNet: An on-line lexical database // International journal of lexicography. – 1990. – Vol. 3(4). – P. 235-244. – doi: 10.1093/ijl/3.4.235.
- [**Moskvoretskii et al., 2024**] Moskvoretskii V., Neminova E., Lobanova A., Panchenko A., Nikishina I. Large Language Models for Creation, Enrichment and Evaluation of Taxonomic Graphs // Semantic Web Journal (forthcoming). – 2024.
- [**Nikisina et al., 2020a**] Nikishina I., Logacheva V., Panchenko A., Loukachevitch N. RUSSE'2020: Findings of the First Taxonomy Enrichment Task for the Russian language // In: International Conference on Computational linguistics and intellectual technologies Dialog-2020, Moscow, Russia, June, 2020. – P. 579-595.
- [**Nikisina et al., 2020b**] Nikishina I., Panchenko A., Logacheva V., & Loukachevitch N. Studying taxonomy enrichment on diachronic wordnet versions // In: Proc. of the 28th International Conference on Computational Linguistics, Barcelona, Spain, December, 2020. – P. 3095-3106.
- [**Nikisina et al., 2022**] Nikishina I., Tikhomirov M., Logacheva V., Nazarov Y., Panchenko A., Loukachevitch N. Taxonomy enrichment with text and graph vector representations // Semantic Web. – 2022. – 13(3). – P. 441-475. – doi: 10.3233/SW-212955.
- [**Sabirova et al., 2024**] Sabirova K., Lukanin A. Automatic Extraction of Hypernyms and Hyponyms from Russian Texts // In: AIST (supplement). Yekaterinburg, Russia, 2014. – P. 35-40.
- [**Team Google, 2024**] Team Google. Gemma: Open models based on gemini research and technology // Arxiv. – 2024. – doi: 10.48550/arXiv.2403.08295.
- [**Tikhomirov et al., 2023**] Tikhomirov M., Chernyshev D. Impact of tokenization on LLaMa Russian adaptation // In: Ivannikov Ispras Open Conference (ISPRAS), Moscow, Russia, December, 2023. – P. 163-168.
- [**Tikhomirov et al., 2024**] Tikhomirov M.M., Chernyshev D.I. Improving Large Language Model Russian adaptation with preliminary vocabulary optimization // Lobachevskii Journal of Mathematics. – 2024. – Vol. 45(7). – P. 3211-3219. – doi: 10.1134/S1995080224604120.
- [**Touvron et al., 2023**] Touvron H. et al. Llama: Open and efficient foundation language models // Arxiv. – 2023. – doi: 10.48550/arXiv.2302.13971.
- [**Yang et al., 2023**] Yang A. et al. Qwen2. 5 technical report // Arxiv. – 2024. – doi: 10.48550/arXiv.2412.15115.

УДК 004.852

doi: 10.15622/rcai.2025.079

РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ ЛЮДЕЙ ПО ДАННЫМ С ИНФРАКРАСНОЙ И ЦВЕТНОЙ КАМЕРЫ С ИСПОЛЬЗОВАНИЕМ СЕГМЕНТНОЙ СПАЙКОВОЙ НЕЙРОННОЙ СЕТИ¹

И.С. Фомин (*i.fomin@rtc.ru*)

ГНЦ РФ Центральный научно-исследовательский
и опытно-конструкторский институт робототехники и технической
кибернетики, Санкт-Петербург

В работе рассматривается вопрос обработки изображений, получаемых при различном режиме работы камеры видеонаблюдения. В качестве примера рассматривается распознавание людей на подмножестве общедоступного набора данных. Показан алгоритм обучения сиамской нейронной сети для преобразования изображений ночной и дневной съемки в векторное пространство представлений. Приведен способ преобразования векторного представления в спайковое представление. Описан алгоритм классификации представлений инфракрасных и RGB изображений по отдельности, а также их преобразования в общий набор и совместной обработки. Проведенные экспериментальные исследования показывают высокое качество предложенных подходов. Качество автономной обработки составило 0,891 для ИК изображений, 0,965 для RGB изображений, совместной обработки – 0,957.

Ключевые слова: реидентификация людей, сверточные сети, сиамские сети, спайковые сети, нейронные сети, инфракрасные изображения, цветные изображения.

Введение

Обнаружение и распознавание объектов в системах видеонаблюдения давно известная и хорошо решаемая задача технического зрения, многие готовые решения используются в различных коммерческих системах

¹ Результаты получены в рамках выполнения государственного задания Минобрнауки России «Исследование методов анализа слабоструктурированных данных, обработки знаний и создания когнитивных агентов на базе комбинированных глубоких нейронных сетей» (FNRG-2025-0008 1024050200009-5-1.2.1;2.2.2).

обеспечения безопасности. Обнаружение людей (наличия людей) на изображении в том числе включено и в некоторые готовые решения, где это используется как один из параметров, регулирующих начало и остановку записи. Однако даже так остается ряд вопросов, который освещен недостаточно и нередко возникает в научных и практических работах. К таким задачам стоит отнести вопросы реидентификации людей при сопровождении между несколькими камерами.

Ранее был предложен универсальный подход к реидентификации людей на основе комбинации сиамской сверточной и сегментной спайковой нейронных сетей. Сиамская сеть после обучения специальным образом отвечает за преобразование четырехмерного тензора данных, полученного с цветной камеры или инфракрасной камеры в пространство представлений. Спайковая сеть для классификации, собранная из нейронов сегментной спайковой модели [Бахшиев и др., 2021], отвечает за определение того, какому из предъявленных при обучении спайковых представлений, полученных из тензоров данных людей соответствует представление, полученное из изучаемого тензора. Ранее уже было показано [Корсаков и др., 2022], что сегментная спайковая сеть хорошо справляется с задачами классификации. Также в ходе исследования [Фомин, 2025] комбинация из сиамской сверточной и сегментной спайковой сетей показала уверенные результаты в задаче классификации и реидентификации изображений людей по данным с цветной камеры.

В данной работе предлагается обобщить ранее предложенный алгоритм реидентификации с использованием сиамской сверточной и сегментной спайковой сетей на обработку изображений с камеры видеонаблюдения, работающей в ночном (инфракрасном) режиме, что позволит применять алгоритм также и в темное время суток. Также предложен подход на основе комбинации сиамской сети для цветных векторов, сиамской сети для инфракрасных векторов и объединяющей их сегментной спайковой сети с многомерным входом для одновременной реидентификации по двум разным источникам данных.

1. Обзор существующих подходов

1.1. Обзор областей применения сиамских сетей

Сравнение и сопоставление векторных представлений изображений сегодня регулярно применяются в задаче многообъектного обнаружения и сопровождения [Luo et al, 2021], [Bertinetto et al, 2016], можно обратить внимание на, то, что для решения этой задачи есть общий подход. Каждое изображение последовательности обрабатывается нейронной сетью для получения сжатого вектора представлений. Вектора представлений используются для сопоставления исходных матриц между собой на основе Евклидова расстояния, косинусного расстояния, или иных метрик дистан-

ции в n -мерном пространстве вещественных чисел R^n . Различия между методами состоят в построении способа формирования векторов представлений.

Современные подходы к сопоставлению в процессе сопровождения в основном основаны на сиамских сетях. Их основная идея [Kim et al, 2017] – обучить нейронную сеть, использующую информацию с двух или более матриц (изображений) чтобы различить их наилучшим образом. Обычно часть сети для извлечения признаков для всех изображений общая, далее идут различные способы сравнения, сопоставления, разделения. В работе [Wang et al, 2015] предлагается сиамская сеть, обученная на контрастной функции ошибки. Она принимает на вход 2 изображения, их IoU, и выдает оценку схожести. В [Zhang et al, 2016] сеть принимает на вход не полные изображения, а их фрагменты и выдает оценку схожести уже между ними. В [Son et al, 2017] рассматривается триплетная функция ошибки (triplet loss). В статье [Zhu et al, 2018] авторы делают следующий шаг, и, добавляя еще один пример, формируют четверную функцию ошибки, где на входе 3 примера одного объекта за последние 3 кадра и 1 пример другого объекта. В режиме работы используется обученная сеть-дескриптор для построения представлений и рассчитывается сопоставление уже между ними. В [Hermans et al, 2017] выполняется сравнение на 128-мерном векторе представлений, извлекаемом тройной сверточной сетью, описанной в [Zhou et al, 2019]. В [Chen et al, 2018] приводится сверточная сеть для предсказания нового положения, которая способна извлекать представления из фрагментов изображения для уточнения сопоставления при получении нового кадра. В [Dai et al, 2016] показан алгоритм сопоставления с использованием GoogLeNet, обученный с использованием triplet loss. Сеть R-FCN [Bakhshiev et al, 2022] используется для генерации предположений о положении объекта на основе траектории и признаков с предыдущих кадров, потом предположения объединяются с реальными обнаружениями и используется обученная GoogLeNet для уточнения.

Кроме простых способов определения близости векторов и отнесения объекта к одной из групп, существуют и более интересные, биоинспирированные подходы. В работе [Корсаков и др., 2022] рассмотрена авторская био-подобная сегментная спайковая модель нейрона, которая хорошо показала себя в задаче классификации (см. [Бахшиев и др., 2021]). Модель способна успешно сопоставлять даже сложные, линейно не разделимые данные.

Исходя из рассмотренных работ можно сделать однозначный вывод о том, что использование сиамских сверточных сетей для извлечения векторов признаков желаемого размера является общепринятым подходом, который актуален для применения в данной работе. С точки зрения алгоритма обучения сиамской сверточной сети выделяются два основных подхода, тройная функция ошибки (triplet loss), ориентированная на раз-

делении точек данных при обучении на основе информации о положительных и отрицательных парах объектов и контрастивное обучение (contrastive loss), который обучается на парах изображений (положительных и отрицательных) независимо.

1.2. Обзор сегментной спайковой модели нейрона

В рамках выполнения настоящей работы использовалась сегментная спайковая модель нейрона (CSNM) [Бахшиев и др., 2021]. Настоящая статья не ставит своей задачей подробное математическое описание использовавшейся модели нейрона. Такое описание можно найти в вышеуказанной работе.

Предполагается, что на вход поступают спайки, которые преобразуются в синапсах в выходную величину, отображающую влияние синаптического тока на сегмент мембраны нейрона (рис. 1).

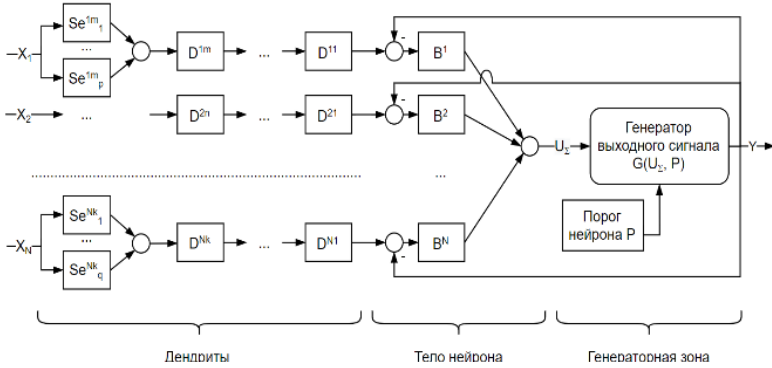


Рис. 1. Структура сегментной спайковой модели нейрона. Se_l^{ij} – входной синапс l , связанный с сегментом дендрита D^{ij} ; D^{ij} – j -й сегмент i -го дендрита; B^k – k -й сегмент тела нейрона

Представленная структурная организация модели нейрона позволяет проводить расчет систем дифференциальных уравнений для каждого нейрона независимо, поскольку нейроны обмениваются исключительно информацией о моменте времени возникновения спайка, а входной вектор системы уравнений модели нейрона формируется внутри модели как функция от моментов времени прихода спайков.

Полагаем, что нейроны обмениваются информацией через события (возникновение спайков), которые можно представить по формуле:

$$x_i = \begin{cases} 1, & t \in [t_i; t_i + \Delta t] \\ 0, & t \notin [t_i; t_i + \Delta t], \end{cases} \quad (1.1)$$

где x_i – выход нейрона в i -й интервал времени, а t – время.

Размер нейрона (число сегментов сомы), множество длин дендритов и число возбуждающих и тормозных синапсов для каждого сегмента мембраны определяют структуру нейрона, которая формирует требуемую реакцию на входные воздействия.

Данный нейрон характеризуется особенной методикой структурного обучения. Согласно ей, для заданного количества сегментов сомы, сначала выполняется процедура синхронизации, когда последовательно наращиваются длины дендритов, чтобы обеспечить одновременность прихода сигнала на сому, затем процедура нормализации, когда наращивается число синапсов, чтобы компенсировать падение уровня сигнала при прохождении по элементам дендрита, комбинация этих операций обеспечивает настройку на паттерн импульсов.

2. Предлагаемые решения

2.1. Наборы данных для экспериментальных исследований

Некоторые решения, принятые в рамках исследования предлагаемого алгоритма, учитывают отдельные характеристики набора данных, поэтому требуется начать с его общего описания.

В качестве источника данных для проведения экспериментов был выбран набор данных SYSU-MM01 [Wu et al, 2017]. Набор содержит изображения нескольких сотен разных людей (персон), полученные с 4 различных цветных и 2 различных инфракрасных камер. Для последующей обработки изображения людей были извлечены из исходных изображений и преобразованы в набор обучающих тензоров фиксированного размера, с дополнительным осреднением и случайными изменениями извлекаемой области, цветности, ориентации и т.д.. Все вектора были поделены по персонам, данные с различных камер были объединены. Оставлены только персоны, присутствующие и на цветных, и на ИК изображениях. В итоге, получено два набора серий векторов, сгруппированных по персонам, более двух сотен персон для обоих вариантов датасета.

Так как для каждой персоны представлено большое количество обучающих тензоров (особенно, извлеченных из данных, полученных с цветной камеры), для основного набора данных (табл. 1, «..., осн.») было выбрано по 50 первых персон для наборов, полученных из цветных и ИК изображений. Также были подготовлены полные наборы тензоров, извлеченных из обеих типов изображений (табл. 1, «..., полн.»), для исследования того, насколько предложенная концепция обучения и тестирования может быть успешно адаптирована кратно большему количеству классов. Параметры наборов представлены в табл. 1.

Таблица 1

Набор	Тип	Классов	Прим. мин.	Прим. макс.	Прим. среднее	Прим. медиана
ИК, осн.	Обуч.	50	32	36	32,28	32
ИК, осн.	Тест	50	8	10	8,3	8
ИК, полн.	Обуч.	233	32	36	32	32
ИК, полн.	Тест	233	8	10	8	8
Цв., осн.	Обуч.	50	56	102	75,84	76
Цв., осн.	Тест	50	14	26	19,46	19
Цв., полн.	Обуч.	233	48	115	74,4	73
Цв., полн.	Тест	233	13	29	19,07	19

2.2. Организация сиамской сети и подход к ее обучению

Так как набор данных достаточно большой, то было принято решение выстроить сиамскую сверточную сеть на основе полноценной сверточной сети, аналогично тому, как было сделано ранее для другого набора [Фомин, 2025]. В качестве основы была взята сеть ResNet-18 [He et al, 2016], предобученная на наборе данных ImageNet [Deng et al, 2009]. Конечный слой, предназначенный для обучения на данных для классификации, был отброшен и заменен на слой линейной рекомбинации с двумя выходами. Также в качестве сети для обучения рассматривалась и собственная сеть, представленная в работе [Фомин, 2025], но ряд экспериментов с разными параметрами обучения показал низкую степень сходимости, для существенно большего по размеру набора данных числа обучаемых параметров оказалось недостаточно.

В качестве функции ошибки исходно рассматривалось два варианта, triplet loss с оригинальным способом формирования обучающих отрицательных примеров [Фомин, 2025], и классический contrastive loss, представленный в том числе и в стандартной библиотеке функций ошибки pytorch, но специально подготовленный для адаптации к имевшейся программной части. Эксперименты с triplet loss продемонстрировали низкую сходимость, при том что на одну эпоху обучения (обработку полного набора данных) требовалось очень много времени в силу необходимости статистической обработки результатов для формирования отрицательных обучающих примеров на следующей эпохе. Замена triplet loss на contrastive loss позволила на треть сократить память, занимаемую одним батчем обучения, и существенно уменьшить время обсчета одной эпохи. Увеличение числа эпох позволило добиться сходимости к качественному распределению кластеров классов в пространстве представлений.

Таким образом, все обучение выполнялось на сиамской сети на базе ResNet-18, с функцией ошибки contrastive loss, до 1000 эпох обучения в обоих случаях, с фиксированными значениями скорости обучения оптимизатора и параметрами функции ошибки.

2.3. Подготовка данных и обучение сегментной спайковой сети

Для распознавания был использован подход к классификации (реидентификации), подробно описанный в работе [Фомин, 2025]. Все особенности классификации (распознавания) последовательностей импульсов подробно изложены в основополагающих работах по этой теме [Бахшиев и др., 2021], [Корсаков и др., 2022]. Кодирование результатов работы из пространства представлений сиамской сверточной нейронной сети в спайковое представление для использования в сегментной спайковой нейронной сети выполнено прямым пропорциональным преобразованием, аналогично тому, как описано в [Фомин, 2025]. Иными словами, каждая точка данных кодируется 3 импульсами задержки по 3 каналам, 1 импульс соответствует 1-й координате вектора представлений, 2 импульс 2-й координате, 3-й используется для синхронизации. Задержка импульса прямо пропорциональна соответствующей координате в векторе представлений. Кодирование выполнено единообразно для представлений тензоров, извлеченных из цветных и инфракрасных изображений.

Исходная архитектура сегментной сети – 3 (5) сегмента тела нейрона, к каждому подключен дендрит, на вход которого поступает импульс от генератора. По завершении обучения на каждом из каналов в рамках процедуры синхронизации увеличивается число сегментов дендрита обратно пропорционально значению задержки сигнала, а также число синапсов на входном дендрите в рамках процедуры нормализации. Подробнее процедура обучения описана в [Бахшиев и др., 2021].

Для экспериментов по совместной реидентификации по данным, извлеченным из цветных и инфракрасных изображений, подготовленные импульсные представления были объединены следующим образом. Для каждой пары импульсов из пространства представлений для цветных данных, подбиралась случайная пара импульсов из пространства представлений для инфракрасных данных, которых было меньше, потому пары иногда повторялись. Нейрон при обучении имел вместо двух входов данных 4 входа, по 2 для каждого из типов, плюс 1 синхронизирующий вход. В качестве обучающих данных классов брались те же центроиды, что и при автономном обучении.

3. Результаты экспериментального исследования

Экспериментальные исследования в данной работе преследовали три основные цели. Во-первых, требовалось удостовериться, что подход, подразумевающий применение сиамской сверточной нейронной сети для перевода тензоров данных, извлеченных из изображений в малоразмерное пространство признаков, может быть обобщен до существенно больших наборов данных по сравнению с тем, что применялся при первичном обучении. Во-вторых, необходимо было экспериментально подтвердить, что

данный подход может быть без существенных изменений обобщен на обработку тензоров, извлеченных из инфракрасных изображений, следовательно, метод реидентификации может быть обобщен на работу в темное время суток и в сумерках. В-третьих, необходимо было проверить гипотезу о том, что универсальный компонент классификации на основе сегментной спайковой нейронной сети принципиально может быть обобщен на совместную идентификацию результатов преобразования цветных и ИК данных, на примере объединения не связанных между собой напрямую данных. В-четвертых, был поставлен отдельный эксперимент по обучению сиамской сверточной нейронной сети на полном наборе данных цветных изображений для исследования принципиальной возможности обобщения предложенного подхода на большое количество классов и большой объем обучающих данных.

В силу вычислительной сложности модели даже с учетом переноса вычислений на GPU и использования параллельных вычислений там, где это допустимо обучение и тестирование модели занимает продолжительное время. Для сокращенного набора данных обучение и тестирование занимает порядка 8-10 часов на Nvidia 1080Ti, для полного набора – 22 тыс. примеров – 211593,59 с на том же вычислителе. Существенное влияние оказывает количество нейронов (233 нейрона на полном наборе). Обучение сверточной сети на малом наборе потребовало нескольких дней, на полном наборе – около полутора недель вычислительного времени.

Оценка качества реидентификации во всех трех случаях выполнялась одинаковым образом. Велся подсчет правильно классифицированных точек пространства представления (TP), ошибок 2 рода, когда срабатывал неправильный нейрон (FP) и ошибок 1 рода, когда не срабатывал правильный нейрон (FN). На основе этих данных по формулам (3.1) были рассчитаны показатели точности (Precision, Prec), полноты (Recall, Rec), F_1 -меры.

$$Prec = \frac{TP}{TP+FP}, Rec = \frac{TP}{TP+FN}, F_1 = \frac{2TP}{2TP+FP+FN}. \quad (3.1)$$

Приводятся показатели и для тестового набора данных, и для обучающего. Для нейронной сети в качестве обучающих примеров выступают центры кластеров классов в обучающем наборе, а все точки данных как из обучающего набора, так и из тестового выступают в качестве тестовых примеров, поэтому такая оценка корректна.

Результаты тестирования представлены в табл. 2.

Таблица 2

Набор	Обучающий набор данных			Тестовый набор данных		
	Prec	Rec	F_1	Prec	Rec	F_1
ИК	0,951	0,996	0,973	0,807	0,993	0,891
Цв., 500	0,966	0,995	0,981	0,911	0,995	0,952
Цв., 1000	0,996	1,000	0,998	0,932	1,000	0,965
ИК+Цв., 500	0,996	0,992	0,994	0,982	0,932	0,957
ИК+Цв., 1000	0,998	0,996	0,997	0,986	0,920	0,952
Цв., полный	0,556	0,999	0,714	0,518	0,999	0,682

Полученные результаты полностью подтверждают высокую результативность предложенного подхода во всех рассмотренных экспериментальных сценариях. Снижение показателей на полном наборе связано в первую очередь с тем, что несмотря на заметную сходимость использованная сверточная нейронная сеть с линейным классификатором не показала достаточной способности к обобщению. Также контрастная функция ошибки как способ обучения недостаточно хороша для малых наборов (что было показано в предыдущих работах). С увеличением количества точек данных выше некоторого предела кластеры перестают быть линейно разделимыми, поэтому часть точек оказывается ближе к другим классам, что увеличивает ошибку. Доработка функции ошибки и выбор сети для больших наборов данных есть предмет дальнейшего исследования.

Заключение

В работе исследовано обобщение ранее предложенного подхода к реидентификации людей на основе комбинации сиамской сверточной и сегментной спайковой нейронных сетей на в несколько раз больший набор данных. Предложено обобщение на обработку данных, полученных в ночном (инфракрасном) режиме камеры видеонаблюдения. Автономное качество (F_1 -мера здесь и далее) для цветных данных составило 0,932 и 0,965 для двух разных эпох, для инфракрасных 0,891. Качество на объединенном датасете составило 0,957 и 0,952 для двух пар результатов. Наконец, экспериментальное исследование на полном наборе (в десятки раз больше рассмотренного в прошлых работах) показало качество 0,714 для обучающих и 0,682 для тестовых данных, извлеченных из цветных изображений, что подтверждает работоспособность предложенного подхода в этих условиях.

Так как исследование на данных, извлеченных из изображений демонстрирует устойчивую успешную классификацию как для наборов данных малого размера, так и для наборов среднего размера, в качестве направления дальнейших исследований предполагается обобщить принцип на тензоры данных, извлекаемые из сенсоров иной физической природы (облака точек и линии с LiDAR, тепловизионные изображения, карты глубины, многоспектральные данные и др.).

Список литературы

- [Бахшиев и др., 2021] Бахшиев А.В., Демчева А.А. Сегментная спайковая модель нейрона // Нелинейная динамика в когнитивных исследованиях (Нижний Новгород, 2021): Труды VII Всероссийской конференции. – 2021. – С. 26-29.
- [Корсаков и др., 2022] Корсаков А.М., Астапова Л.А., Бахшиев А.В. Применение сегментной спайковой модели нейрона со структурной адаптацией для решения задач классификации // Информатика и автоматизация. – 2022. – Т. 21, № 3. – С. 493-520.
- [Фомин, 2025] Фомин И.С. Сравнение сжатых векторных представлений матриц с помощью сегментной спайковой модели нейрона (CSNM) // Робототехника и техническая кибернетика. – 2025. – Т. 13, № 1. – С. 33-40.
- [Bakhshiev et al, 2022] Bakhshiev A., Demcheva A., Stankevich L. CSNM: The Compartmental Spiking Neuron Model for Developing Neuromorphic Information Processing Systems // In: Advances in Neural Computation, Machine Learning, and Cognitive Research V: Studies in Computational Intelligence, eds. B. Kryzhanovsky et al. – Cham: Springer International Publishing, 2022. – P. 327-333. – doi: 10.1007/978-3-030-91581-0_43.
- [Bertinetto et al, 2016] Bertinetto L., Valmadre J., Henriques J.F., Vedaldi A., Torr P.H.S. Fully-Convolutional Siamese Networks for Object Tracking. In: Computer Vision – ECCV 2016 Workshops : Lecture Notes in Computer Science / eds. G. Hua, H. Jégou. – Cham: Springer International Publishing, 2016. – P. 850-865. – doi: 10.1007/978-3-319-48881-3_56.
- [Chen et al, 2018] Chen L., Ai H., Zhuang Z., Shang C. Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification // In: 2018 IEEE International Conference on Multimedia and Expo (ICME). – 2018. – P. 1-6. – doi: 10.1109/ICME.2018.8486597.
- [Dai et al, 2016] Dai J., Li Y., He K., Sun J. R-FCN: Object Detection via Region-based Fully Convolutional Networks // In: Advances in Neural Information Processing Systems, Curran Associates, Inc. – 2016. – P. 29. – doi:10.48550/arXiv.1605.06409.
- [Deng et al, 2009] Deng J., Dong W., Socher R., Li L.J., Li K., Fei-Fei L. Imagenet: A large-scale hierarchical image database // In: 2009 IEEE conference on computer vision and pattern recognition, 2009 Jun 20. – P. 248-255. – doi: 10.1109/CVPR.2009.5206848.
- [He et al, 2016] He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // In: Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – P. 770-778. – doi: 10.1109/CVPR.2016.90.
- [Hermans et al, 2017] Hermans A., Beyer L., Leibe B. In Defense of the Triplet Loss for Person Re-Identification [Электронный ресурс] // arXiv:1703.07737 [cs] // arXiv. – 2017. – URL: <https://arxiv.org/abs/1703.07737> (дата обращения: 24.05.2025). doi:10.48550/arXiv.1703.07737.
- [Kim et al, 2017] Kim M., Alletto S., Rigazio L.. Similarity Mapping with Enhanced Siamese Network for Multi-Object Tracking [Электронный ресурс] // arXiv:1609.09156 [cs] // arXiv. – 2017. – URL: <https://arxiv.org/abs/1609.09156> (дата обращения: 24.05.2025). – doi: 10.48550/arXiv.1609.09156.
- [Luo et al, 2021] Luo W., Xing J., Milan A., Zhang X., Liu W., Kim T.-K. Multiple Object Tracking: A Literature Review //Artificial Intelligence. – 2021. – Vol. 293. – P. 103448. – doi: 10.1016/j.artint.2020.

- [**Son et al, 2017**] Son J., Baek M., Cho M., Han B. Multi-Object Tracking With Quadruplet Convolutional Neural Networks // In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2017. – P. 5620-5629. – doi: 10.1109/CVPR.2017.403.
- [**Wang et al, 2015**] Wang B., Wang L., Shuai B., Zuo Z., Liu T., Luk Chan K., Wang G. Joint Learning of Convolutional Neural Networks and Temporally Constrained Metrics for Tracklet Association // In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. – 2016. – P. 1-8. – doi: 10.1109/CVPRW.2016.55.
- [**Wu et al, 2017**] Wu A., Zheng W.S., Yu H.X., Gong S., Lai J. RGB-infrared cross-modality person re-identification // In: Proceedings of the IEEE international conference on computer vision. – 2017. – P. 5380-5389. – doi: 10.1007/s11263-019-01290-1.
- [**Zhang et al, 2016**] Zhang S., Gong Y., Huang J.-B., Lim J., Wang J., Ahuja N., Yang M.-H. Tracking Persons-of-Interest via Adaptive Discriminative Features // In: Computer Vision – ECCV 2016: Lecture Notes in Computer Science, eds. B. Leibe et al. – Cham: Springer International Publishing, 2016. – P. 415-433. – doi: 10.1007/978-3-319-46454-1_26.
- [**Zhou et al, 2019**] Zhou H., Ouyang W., Cheng J., Wang X., Li H. Deep Continuous Conditional Random Fields With Asymmetric Inter-Object Constraints for Online Multi-Object Tracking // IEEE Transactions on Circuits and Systems for Video Technology. – 2019. – Vol. 29(4). – P. 1011-1022. doi:10.1109/TCSVT.2018.2825679.
- [**Zhu et al, 2018**] Zhu J., Yang H., Liu N., Kim M., Zhang W., Yang M.-H. Online Multi-Object Tracking with Dual Matching Attention Networks // In: Proceedings of the European Conference on Computer Vision (ECCV). – 2018. – P. 366-382. – doi: 10.1007/978-3-030-01228-1_23.

УДК 007.5:519.816:681.3.016

doi: 10.15622/rcai.2025.080

ИСПОЛЬЗОВАНИЕ МЕТОДОВ АНСАМБЛИРОВАНИЯ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ КЛАССИФИКАЦИИ ОБЪЕКТОВ, СОДЕРЖАЩИХ ШУМ¹

М.В. Фомина (*m_fomina2000@mail.ru*)

Г.В. Швец (*ShvetsGV@mpei.ru*)

Национальный исследовательский университет «МЭИ», Москва

Рассматривается проблема интеллектуального анализа зашумленных данных, полученных из реальных источников. Предлагается метод организации системы классификации объектов, подверженных влиянию шума различного типа. В основу положено сочетание ансамблевых методов с нейросетевым подходом. Предложена архитектура нейросети, решающей поставленную задачу. Приводятся результаты машинного эксперимента, даётся оценка влиянию шумов различного типа на результаты классификации.

Ключевые слова: классификация данных, модели шума, нейронная сеть, ансамблевый метод, предобработка данных.

Введение

Системы искусственного интеллекта, способные оказать помощь людям в их профессиональной сфере при принятии ответственных решений, требуют обработки и анализа больших потоков данных. В настоящее время вместе с увеличением объема информации растет также и проблема её качества. Реальные данные, получаемые из различных областей, неизбежно содержат шум: это искажения данных, или пропуски в данных [Goodfellow et al., 2016], [Гафаров и др., 2018].

Присутствие различного рода шума в данных оказывает существенное влияние на результаты их анализа, особенно в задачах классификации данных, где требуется высокая точность отнесения рассматриваемых объектов к различным категориям (классам). Эта проблема становится осо-

¹ Работа выполнена при финансовой поддержке РФФ (проект № 24-11-00285) <https://rscf.ru/project/24-11-00285>.

бенно важной в таких областях, где ошибка классификации может иметь крайне серьезные последствия, например, при разработке систем поддержки принятия решений [Еремеев и др., 2021].

В последние годы нейронные сети хорошо зарекомендовали себя как мощный инструмент решения разнообразных задач машинного обучения, включая классификацию данных. Благодаря способности изучать сложные нелинейные зависимости и высокой обобщающей способности, нейронные сети хорошо использовать для задач анализа зашумленных данных [Москвитин и др., 2023]. Современные архитектуры нейронных сетей способны демонстрировать устойчивость к определенным видам шума, выделяя значимые паттерны и игнорируя случайные искажения.

В работе рассматривается проблема улучшения классификации зашумленных данных при высоких уровнях шума за счёт использования комплексного подхода – интеграции ансамблевых методов и использования нейросети определённого типа. Приведены результаты, позволяющие оценить влияние шума различных типов на успешность решения задачи классификации. Рассмотрена организация программного комплекса, реализующего предложенные методы и алгоритмы.

1. Проблема классификации объектов в интеллектуальных системах

Задача классификации данных различного типа, является одной из фундаментальных задач машинного обучения, целью которой является отнесение объектов к одному из определенных классов на основе анализа признаков, отражающих свойства объектов [Вагин и др., 2008].

Для успешной классификации объектов требуется прежде всего решить задачу построения набора классификационных правил – классификатора, который, по сути, является алгоритмом, дающим ответ на вопрос, принадлежит ли очередной пример классу, или не принадлежит. Построение классификатора, или классификационной модели, основано на анализе конкретных примеров с целью поиска свойств, наиболее характерных для них. Примеры, которые используются для анализа, представляют собой объекты, которые могут быть описаны набором свойств (атрибутов); такое описание называется признаковым описанием.

Известны различные подходы и модели для построения классификаторов; это продукционные правила, деревья решений, методы, основанные на теории приближённых множеств, нейронные сети. Важно, что ни один из методов не является универсальным и обычно показывает наилучшие результаты на наборах данных определённого типа [Еремеев и др., 2023].

Построение хорошей классификационной модели выполняется на примерах, которые образуют обучающую выборку. В перечисленных выше методах объекты обучающей выборки снабжаются меткой – именем класса.

Если обучающая выборка размечена, речь идёт об обучении «с учителем», классификационные модели, работающие с неразмеченными данными, решают задачу обучения «без учителя».

Следует отметить, что успешность, или полезность построенного классификатора определяется его способностью правильно распознавать новые, не вошедшие в обучающую выборку примеры. Также полезным свойством классификационной модели будет устойчивость к шуму в данных. Для оценки этого свойства используется оценка *Точность классификации* или *Accuracy*, основной показатель оценки производительности модели, который измеряет долю верных ответов среди всех сделанных предсказаний [Goodfellow et al., 2016].

Для оценки эффективности классификационной модели используются следующие наборы данных:

- Обучающая выборка (Training set) – выборка, используемая для обучения модели.
- Тестовая выборка (Test set) – выборка, используемая для окончательной оценки производительности модели.
- Валидационная выборка (Validation set) – выборка, используемая для настройки гиперпараметров и выбора лучшей модели.

Шум оказывает комплексное негативное воздействие на все аспекты процесса машинного обучения, начиная от качества данных, заканчивая практическим применением моделей. Разработка методов, устойчивых к различным типам шума является важной задачей для создания надежных систем классификации.

2. Модели шума и их характеристики

Как уже говорилось ранее, шум – это нежелательные искажения данных, которые влекут за собой снижение успешности работы алгоритмов классификации [Gonsales et al., 2018]. Шум в данных представляет собой непредсказуемые вариации, поскольку искажению подвергаются истинные значения. В контексте задач классификации можно выделить несколько основных типов шума: гауссовский, импульсный, равномерный, пропущенные значения, мультипликативный, шум типа «соль и перец». Шум каждого типа по-своему влияет на результаты классификации и требует специализированных подходов для эффективной обработки данных [Fomina et al., 2013].

В рамках исследования был разработан программный комплекс, одной из функций которого является искусственное внесение в данные шума различного типа. Была реализована поддержка следующих ключевых типов шума.

Гауссовский шум (Gaussian noise). Является наиболее распространенным типом шума в реальных данных и характеризуется нормальным распределением значений с нулевым математическим ожиданием.

Импульсный шум (Impulse noise). Характеризуется редкими, но значительными по амплитуде выбросами, которые изменяют истинные значения. Импульсный шум особенно опасен тем, что может спровоцировать радикальное искажение ключевых характеристик отдельных объектов, потенциально переводя их в область другого класса.

Равномерный шум (Uniform noise). Проявляется в случайных добавках к данным, которые имеют равномерное распределение в заданном интервале [Gonsales et al., 2018].

Пропущенные значения (Missing values). Данный тип представляет собой особый шум, при котором некоторые значения признаков по какой-либо причине отсутствуют. При этом может быть известна p – вероятность того, что некоторое значение отсутствует.

Пропущенные значения представляют особую сложность для алгоритмов машинного обучения, не способных работать с неполными данными [10]. Это требует дополнительной обработки с использованием специальных методов восстановления отсутствующего значения, что вносит дополнительную неопределенность в процесс анализа.

Шум типа «Соль и перец» (Salt and pepper noise). Этот тип шума представляет собой случайную замену значений на экстремальные минимумы (*перец*) или максимумы (*соль*) x_{\min} , x_{\max} . Данный шум является частным случаем импульсного шума, при этом величины x_{\min} , x_{\max} – это минимальное и максимальное значения признака. Также известны вероятности проявления «перца» и «соли» соответственно [Gonsales et al., 2018]. Такой шум создает резкие аномалии в данных, что может привести к значительным искажениям признакового пространства. В задачах классификации подобный шум может создавать ложные кластеры или искажать истинные границы между классами.

3. Расширение возможностей нейронных сетей с помощью ансамблевых методов

Нейронные сети представляют собой мощный инструмент для создания моделей, способных эффективно аппроксимировать сложные нелинейные зависимости между входными признаками и целевыми переменными [Гафаров и др., 2018].

За счет применения специальных методов и техник, которые направлены на улучшение обобщающей способности и снижение чувствительности к искажениям в данных можно значительно улучшить эффективность работы нейросетей. Следующие ключевые подходы позволяют сде-

лать нейронные сети более устойчивыми к шуму: регуляризация, нормализация, использование глубоких нейронных сетей с резидуальными соединениями [He et al., 2016].

Регуляризация представляет собой набор методов, которые направлены на предотвращение переобучения модели путем наложения дополнительных ограничений на гиперпараметры или структуру нейронной сети [Goodfellow et al., 2016]. При наличии зашумленных данных регуляризация приобретает особую важность, поскольку предотвращает запоминание зашумленных данных нейросетью.

Методы нормализации активаций в нейросетях стабилизируют процесс их обучения, ускоряют сходимость и повышают устойчивость к различным типам шума.

В качестве базовой модели нейронной сети в работе предлагается использовать сеть с резидуальными соединениями, где передаточная функция $F(x)$ представляет последовательность двух взвешенных слоёв. При наличии зашумленных данных использование такой структуры нейросети с резидуальными соединениями даёт ряд преимуществ. Перечислим их.

1. Улучшенное распространение градиентов. При использовании резидуальных соединений локальный градиент функции потерь относительно входа слоя гарантированно не затухает даже при прохождении через множество слоев, что является особенно важным моментов при наличии шума в данных, где стабильность таких градиентов имеет критическое значение.

2. Сохранение информации. Прямое соединение позволяет беспрепятственно проходить входной информации через сеть, что позволяет сохранить исходные признаки даже в условиях шума.

3. Адаптивная глубина. Сеть имеет возможность выборочно отключать определенные резидуальные блоки (обнуляя их веса), эффективно уменьшая свою глубину в некоторых случаях.

Ансамблевые методы в машинном обучении основаны на принципе объединения множества моделей с целью повышения точности классификации, эффективности и устойчивости [Zhou, 2012]. Данный подход основан на главном, фундаментальном принципе: разнообразие независимых экспертов (моделей) улучшает качество коллективного решения. Ансамблевое обучение является особенно актуальным в контексте рассматриваемой задачи – обучение и классификация в условиях зашумленных данных, поскольку данное разнообразие позволяет нам компенсировать недостатки отдельных моделей, что способствует повышению устойчивости к различным видам шума.

Результат классификации объекта, полученный как объединение предсказаний нескольких классификаторов, формируется как комбинация результатов базовых классификаторов, входящих в ансамбль.

Ансамблевые методы, которые основываются на комбинировании предсказаний нескольких базовых моделей, позволяют существенно улучшить устойчивость и точность классификации зашумленных данных. В особенности это касается адаптивных ансамблевых моделей, учитывающие уверенность и компетентность отдельных моделей в различных областях пространства признаков и динамически адаптирующихся к различного рода шуму.

На основе проведенных исследований был спроектирован программный комплекс, способный эффективно решать задачи классификации зашумленных данных. Ключевой особенностью разработанного программного комплекса является реализация адаптивного ансамбля моделей, который динамически комбинирует результаты классификации нейронной сети и классических алгоритмов машинного обучения в случае, если главный классификатор – нейронная сеть – делает ошибки.

Проектирование адаптивного ансамбля было связано с принятием следующих решений. Ансамбль включает в себя основную нейронную сеть, а также набор вспомогательных моделей, что обеспечивает разнообразие подходов к классификации. В состав ансамбля был включен ряд моделей классификации. Среди них наибольшее влияние на результаты классификации оказали следующие модели:

1. Основная нейронная сеть с резидуальными соединениями.
2. Случайный лес (Random forest).
3. Градиентный бустинг (Gradient boosting).
4. Метод опорных векторов (SVM).
5. К-ближайших соседей (k-NN).

Особенностью ансамбля является наличие механизма адаптивного взвешивания, что позволяет динамически определять вклад каждой модели в итоговое решение на основе её уверенности и исторической точности для конкретного типа данных и шума.

В качестве базовой архитектуры выбрана глубокая нейронная сеть с несколькими скрытыми слоями и резидуальными соединениями. Её особенности:

1. Входной слой с возможностью добавления контролируемого шума.
2. Резидуальные блоки, улучшающие распространение градиентов.
3. Возможность выбора различных функций активации.
4. Многоуровневая регуляризация для предотвращения переобучения.

4. Эксперименты и оценка результатов

Методика проведения эксперимента с использованием подготовленных программных средств заключается в последовательной реализации таких этапов, как:

- этап подготовки данных;

- этап обучения моделей и оценки их качества;
- этап классификации тестовых примеров при различных уровнях шума в данных. Рассмотрим их подробнее.

Для всестороннего исследования эффективности разработанного программного комплекса были выбраны разносторонние наборы данных с разными характеристиками: размерностью, сбалансированностью классов, сложностью разделяющих поверхностей, областями применения [UCI, 1998]. Каждый набор данных разделялся на обучающую выборку и тестовую выборку.

Каждый тестовый набор данных подвергался воздействию шума различных типов с изменяющимся уровнем интенсивности от 0% до 30% и шагом 5%.

Для каждого набора данных проводилось обучение базовой нейронной сети, и остальных моделей ансамбля.

Для всех сочетаний модели шума и интенсивности шума проводилась классификация тестовых примеров с использованием базовой нейронной сети и остальных моделей ансамбля. Каждый эксперимент повторялся многократно, затем результаты усреднялись.

Для оценки результатов классификации использовалась мера Ассигаса, которая представляет долю правильно классифицированных примеров по сравнению с общим числом примеров тестовой выборки. Также исследовалось влияние на точность классификации тестовых примеров методов предварительной обработки данных, а также использование ансамблевых методов. Результаты оформлялись в виде графиков и таблиц.

Ниже представлены полученные результаты для следующих случаев.

1 Исследование влияния гауссовского шума проводилось на наборе данных Breast Cancer. Этот набор данных был выбран в качестве репрезентативного примера ввиду его сбалансированности и медицинской значимости задачи, где каждый процент точности имеет критическое значение для корректности диагностики.

В табл. 1 представлена зависимость точности классификации тестовых примеров от уровня шума при использовании гауссовской модели шума. Каждое значение в таблице представляет точность классификации тестовых примеров, показанную одним из классификаторов при наличии в данных указанного уровня шума (в процентах). Для каждого уровня шума наилучшие результаты выделены жирным шрифтом.

Как видно из таблицы, разработанная адаптивная ансамблевая модель демонстрирует существенно более высокую устойчивость к гауссовскому шуму по сравнению с отдельно взятыми алгоритмами. Особенно показательна стабильность ансамбля с главенствующей в нем хорошо разработанной нейронной сетью при высоких уровнях шума, где разрыв в точности становится наиболее заметным.

Таблица 1

Зависимость точности классификации тестовых примеров (%) от уровня шума для модели шума «Гауссовский» при работе с набором данных Breast Cancer

Уровень шума %	Ансамблевая модель %	Нейронная сеть %	Random Forest %	Gradient Boosting %	SVM %	K-NN %
0	97	92	96	91	97	94
5	97	91	94	90	97	94
10	96	90	94	90	95	94
15	96	91	94	90	95	95
20	96	91	94	91	96	95
25	95	90	93	91	95	93
30	95	91	92	90	93	93

2. Следующая таблица представляет результаты, полученные для набора данных Wine при наличии импульсного шума. Импульсный шум, как и предполагалось, оказался одним из наиболее разрушительных типов шума и снижает точность классификации тестовых примеров для большинства моделей. Результаты работы с набором данных Wine представлены в табл. 2.

Таблица 2

Зависимость точности классификации тестовых примеров (%) от уровня шума для модели «импульсный шум» при работе с набором данных Wine

Уровень шума %	Ансамблевая модель, %	Нейронная сеть, %	Random Forest, %	Gradient Boosting, %	SVM, %	K-NN, %
0	100	95	95	97	100	95
5	96	81	93	95	80	82
10	93	67	92	90	65	70
15	93	63	89	90	56	67
20	88	55	85	85	53	52
25	89	48	85	86	47	48
30	85	43	82	80	42	43

Разработанный адаптивный ансамбль демонстрирует значительное преимущество над отдельными моделями, особенно при высоких уровнях шума. Так при уровне шума равном 30% точность классификации тестовых примеров у нейронной сети, работающей без поддержки алгоритмов, входящих в ансамбль, падает на более чем 50%. Это значительно превышает аналогичные результаты при гауссовском шуме. Также при высоких уровнях шума адаптивный ансамбль превосходит в среднем на 3% такие алгоритмы, как Random Forest, и Gradient Boosting.

Методы классификации K-NN, SVM и одиночная нейронная сеть показали наименьшую устойчивость к импульсному шуму, что объясняется их чувствительностью к экстремальным значениям.

3. Шум, который проявляется в виде пропущенных значений, представляет собой особую проблему, поскольку многие алгоритмы машинного обучения не способны работать с неполными данными. Эксперименты проводились на наборе данных Heart Disease с последовательным увеличением вероятности появления в данных пропущенных значений от 0 до 30%. Результаты эксперимента на обучающих выборках представлены в табл. 3.

Таблица 3

Зависимость точности классификации тестовых примеров от уровня шума для модели шума «пропущенные значения» при работе с набором данных Heart Disease

Уровень шума %	Ансамблевая модель %	Нейронная сеть %	Random Forest %	Gradient Boosting %	SVM %	K-NN %
0	86	86	80	78	84	80
5	86	86	80	81	85	82
10	85	84	79	79	84	80
15	83	83	77	79	82	81
20	81	79	76	75	81	77
25	82	80	75	74	80	77
30	79	78	74	74	79	77

5. Предобработка данных

Важной частью разработанного программного комплекса является *модуль предобработки данных*. Для рассмотренных моделей шума – гауссовский шум, импульсный шум, шум «отсутствие значений» были предложены и программно смоделированы следующие методы обработки.

Известно, что гауссовский шум характеризуется нормальным распределением, что в свою очередь позволяет применять статистически оптимальные методы фильтрации, к ним относится адаптивный фильтр Винера. Данный метод предобработки позволяет снизить уровень шума без потери важной информации, что обеспечивает качество данных для их последующего анализа и классификации. [Gonsales et al., 2018].

Импульсный шум требует специальных методов, способных обнаруживать и корректировать локальные выбросы без размывания значимой информации; к таким методам относится адаптивный медианный фильтр. Метод автоматически подстраивается под локальные характеристики, обеспечивая оптимальный баланс между подавлением шума и сохранением полезной информации. [Hendrycks et al., 2019].

Обработка пропущенных значений и алгоритмы восстановления этих значений рассмотрены в [Fomina et al. 2013]. Прежде всего это метод К ближайших соседей: пропущенные значения заполняются на основе анализа значений у К ближайших, наиболее похожих объектов. Сходство объектов определяется с помощью евклидова расстояния в пространстве признаков, не содержащих шума.

Помимо методов и алгоритмов обработки зашумленных данных предложен механизм автоматического определения типа шума на основе статистического анализа.

Заключение

Проведенные экспериментальные исследования, выполненные с помощью разработанного программного комплекса для классификации зашумленных данных на основе нейронных сетей, показали эффективность предложенного подхода. Предложенный адаптивный ансамбль с главенствующей нейронной сетью продемонстрировал высокую точность классификации для всех исследованных типов шума. Особенно это касается ситуаций, когда уровень шума достигает высоких значений. Экспериментально получены оценки влияния шумов разного типа на успешность классификации.

Разработанный и реализованный механизм ансамблирования успешно расширяет возможности нейронной сети, сохраняя её в качестве основного классификатора и динамически подключая вспомогательные алгоритмы только в случаях, когда нейросеть демонстрирует недостаточную уверенность в своих предсказаниях. Такой подход позволил повысить устойчивость нейросетевого решения к различным типам шума без необходимости кардинального изменения архитектуры самой нейронной сети. Предложенная архитектура нейронной сети с резидуальными соединениями в сочетании с методами предобработки данных и адаптивным ансамблированием, способна обеспечить эффективное решение даже в наиболее сложных случаях.

Реализованные методы предобработки данных, подверженных влиянию шума, способны обеспечить в большинстве случаев значительное повышение точности классификации, прежде всего за счёт сглаживания импульсных выбросов и применения методов восстановления неизвестных значений в данных.

Список литературы

- [Goodfellow et al., 2016] Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge: MIT Press, 2016.
- [Гафаров и др., 2018] Гафаров Ф.М., Галимьянов А.Ф. Искусственные нейронные сети и приложения: учебное пособие. – Казань: Изд-во Казан. Ун-та, 2018.

- [Еремеев и др., 2021] Еремеев А.П., Варшавский П.Р., Поляков С.А. Программная реализация модуля анализа данных на основе прецедентов для распределенных интеллектуальных систем // Программные продукты и системы. – 2021. – № 34(3).
- [Москвитин и др., 2023] Москвитин В.М., Семёнова Н.И. Влияние шума на рекуррентные нейронные сети с нелинейными нейронами // Известия вузов. Прикладная нелинейная динамика. – 2023. – Т. 31, № 4.
- [Вагин и др., 2008] Вагин В.Н., Головина Е.Ю., Загорянская А.А., Фомина М.В. Достоверный и правдоподобный вывод в интеллектуальных системах / под ред. В.Н. Вагина, Д.А. Поспелова. – 2-е изд. – 2008.
- [Еремеев и др., 2023] Еремеев А.П., Сергеев М.Д., Петров В.С. Интеграция методов обучения с подкреплением и нечеткой логики для интеллектуальных систем реального времени // Программные продукты и системы. – 2023. – № 4.
- [Gonsales et al., 2018] Gonzalez R.C., Woods R.E. Digital Image Processing. – 4th ed. – London: Pearson, 2018.
- [Fomina et al. 2013] Marina Fomina, Alexander Eremeev, Vadim Vagin. Noise models in Inductive Concept Formation // Proceedings of ICEIS 2013, 15th International Conference on Enterprise Information Systems, 2013, Angers – France. – 2013. – Vol. 1. – P. 413-419.
- [He et al., 2016] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). – 2016.
- [Zhou, 2012] Zhou Zhi-Hua. Ensemble Methods: Foundations and Algorithms. – Boca Raton: Chapman & Hall/CRC, 2012.
- [UCI, 1998] UCI Repository of Machine Learning Datasets. Available: <http://archive.ics.uci.edu/ml/>.
- [Hendrycks et al., 2019] Hendrycks D., Dietterich T. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations // Int. Conf. On Learning Representations (ICLR). – 2019.

УДК 004.89

doi: 10.15622/rcai.2025.081

**ОПТИМИЗАЦИЯ ГЕНЕРАТИВНЫХ СИСТЕМ
С ПОМОЩЬЮ UNIKERNEL¹**

Э.К. Берхеев (*auzza@bk.ru*)

А.А. Романов (*romanov73@gmail.com*)

Ульяновский государственный технический университет,
Ульяновск

В работе описывается подход, обеспечивающий ускорение работы нейросетевых моделей за счет использования unikernel-среды. Проведен анализ существующих архитектур глубокого обучения, таких как CNN, ResNet и DenseNet, с точки зрения их применимости к легковесной виртуализации. Предложена методика оптимального развёртывания моделей в unikernel-среде, включая выбор подходящей архитектуры и реализацию окружения с минимальными издержками. Проведены экспериментальные исследования, демонстрирующие преимущества предложенного подхода, включая улучшение

¹ Исследование выполнено в рамках государственного задания № 075-03-2023-143 по проекту «Исследование интеллектуального предиктивного мультимодального анализа больших данных, конструирование признаков гетерогенных динамических данных для машинного обучения».

времени отклика и снижение потребления ресурсов. Сформулированы выводы и рекомендации для проведения дальнейших исследований в области генеративного искусственного интеллекта и серверлесс-архитектур. Новизна подхода заключается в методике интеграции современных методов генеративного искусственного интеллекта с технологией unikernel.

Ключевые слова: unikernel, генерация изображений, сервисы, нейронные сети.

Введение

Машинное обучение и глубокие нейронные сети представляют собой ключевые инструменты для анализа данных и решения различных задач, таких как классификация изображений, обработка естественного языка и предсказательная аналитика. Однако реализация этих методов требует значительных вычислительных ресурсов, особенно в тех случаях, когда необходимо развертывание больших моделей для обработки данных в реальном времени. Данная проблема особенно характерна при использовании традиционных методов развертывания моделей, таких как контейнеры или виртуальные машины.

Современные задачи компьютерного зрения требуют эффективного развертывания моделей глубокого обучения, таких как сверточные нейронные сети (CNN), ResNet и DenseNet, для решения задач классификации изображений в условиях ограниченных вычислительных ресурсов. Несмотря на высокую точность и распространённость этих моделей, традиционные методы их развертывания, основанные на полноценных операционных системах и контейнерах, обладают значительным временем инициализации и избыточным потреблением ресурсов [Anari, 2025].

Технология unikernel представляет собой перспективный подход к оптимизации процессов развертывания, позволяющий создать минималистичное и изолированное окружение, включающее только необходимые компоненты для выполнения модели. Это снижает накладные расходы и ускоряет инстанцирование сервисов на базе нейросетей, что особенно важно при работе в реальном времени или в распределённых системах с ограниченной мощностью [Chen et. al., 2021], [Imes, 2018].

В данной работе рассматриваются модели глубокого обучения CNN, ResNet и DenseNet как объект исследования, а также методы их оптимального развертывания с использованием unikernel. Основное внимание уделяется анализу производительности, времени запуска и потребления ресурсов при различных сценариях использования. Предложенный подход направлен на создание методологии, обеспечивающей быстрое и эффективное развертывание нейросетевых моделей с сохранением их функ-

циональных характеристик. Практическая ценность исследования заключается в возможности использования результатов в задачах реального времени, облачных вычислениях и встраиваемых системах.

1. Методы оптимизации вычислительных ресурсов

Наиболее значимые результаты представлены в зарубежных работах. В статье [Ahuja, 2021] подробно описаны преимущества использования unikernel для развёртывания моделей машинного обучения. Отмечается, что unikernel уменьшает объём памяти, ускоряет запуск до нескольких секунд, повышает уровень безопасности и упрощает развёртывание через инструмент OPS, который позволяет преобразовать существующие ML-модели в unikernel-образы. Применение таких решений особенно актуально в распределённых и ограниченных по ресурсам средах, таких как пограничные вычисления и IoT. Дополнительный акцент сделан на способности unikernel функционировать в изолированной среде, что критично для задач с высокими требованиями к надёжности.

Исследования показывают, что unikernel, такие как MirageOS и IncludeOS, обладают высокой производительностью и малыми накладными расходами, что делает их привлекательными для задач машинного обучения. В частности, использование MirageOS демонстрирует быстрое время загрузки и низкую нагрузку на память по сравнению с традиционными виртуальными машинами [Imada, 2018], [Kurek, 2024].

Однако существует ряд ограничений. Например, исследование [Mitra, 2019] показывает, что интеграция популярных библиотек машинного обучения, таких как TensorFlow и PyTorch, с unikernel сталкивается с трудностями. В частности, возникают проблемы совместимости библиотек, требующих динамической компоновки и использования внешних зависимостей, что усложняет разработку и эксплуатацию таких решений.

Кроме того, исследование применения unikernel в окружениях с поддержкой безсерверных архитектур [Tsiourvas, 2021] указывает на их высокую эффективность для задач с короткими жизненными циклами процессов. В таких случаях unikernel обеспечивают минимальную задержку при вызове функций, что делает их подходящими для использования в рамках пограничных вычислений и FaaS (Function-as-a-Service) решений.

Исследование разработки Unikraft демонстрирует, что unikernel могут значительно улучшить производительность приложений за счет минимизации компонентов операционной системы и создания легковесных виртуальных машин. Это позволяет сократить время загрузки до миллисекунд и уменьшить объём занимаемой памяти, что делает unikernel перспективной технологией для высокопроизводительных задач машинного обучения [Kuenzer, 2021].

Контейнеры против unikernel: Сравнение контейнеров с unikernel в рамках пограничных вычислений показало, что unikernel обеспечивают более низкое время инстанцирования и меньшую нагрузку на ресурсы по сравнению с контейнерами и виртуальными машинами. Это делает их подходящими для использования в распределенных вычислительных системах, где критичны быстрый отклик и минимизация потребления ресурсов [Kong, 2022].

В статье [Ahuja, 2021] рассматривается сравнительная эффективность unikernel по отношению к контейнерам и виртуальным машинам. Результаты показывают значительное сокращение времени инстанцирования, снижения потребления памяти и процессорных ресурсов, а также улучшение изоляции и безопасности за счёт минималистской архитектуры. Это делает unikernel подходящим для систем реального времени и критически важных приложений, таких как автономный транспорт и промышленная автоматизация. Также подчёркивается совместимость unikernel с популярными ML-библиотеками, хотя и отмечаются трудности, связанные с ограниченной поддержкой динамических компонентов и необходимостью адаптации инструментов сборки. Одним из наиболее продвинутых инструментов для генерации unikernel-систем является фреймворк Unikraft, предлагающий оптимизированный путь к созданию высокопроизводительных unikernel-окружений [Kuenzer et. al., 2021].

Применение AutoML-алгоритмов в контексте автоматической настройки моделей может значительно снизить входной порог и упростить развертывание [Бабкина и др., 2024]. Особое внимание при построении системы было уделено интеллектуальной оптимизации гиперпараметров, что находит обоснование в отечественных работах по AutoML [Донской, 2020]. Отмечается высокий потенциал технологии для задач с ограниченным временем отклика и ресурсами, однако также указывается на трудности интеграции нейросетевых моделей из-за несовместимости с текущими системами виртуализации и отсутствия гибких средств компиляции и связывания библиотек. В частности, рассматриваются направления, близкие по задачам: машинное обучение в робототехнике, где акцент делается на снижение затрат и ускорение процессов; AutoML и интеллектуальная оптимизация, предполагающие автоматизацию конфигурации и развертывания моделей; проектирование нейроморфных систем и работа с несбалансированными данными, где целью является эффективное использование ресурсов. В исследованиях по переходу от контейнеров к unikernel в условиях edge-вычислений подтверждены преимущества такого подхода для промышленного применения [Chen et al., 2021].

Современные исследования указывают на высокую перспективность использования unikernel в серверлесс-инфраструктурах, особенно на границе сети [Moebius et al., 2024]. Однако для широкого применения необ-

ходима доработка экосистемы, создание инструментов автоматизации и поддержка интеграции с существующими ML-средствами. Однако наряду с преимуществами, unikernel имеет и ограничения, связанные с поддержкой динамических библиотек и сложностью отладки, как подчёркивается в [Leon, 2020].

В качестве метода решения задачи оптимизации развертывания нейросетей через unikernel можно предложить разработку инструментов для автоматизации сборки и управления зависимостями библиотек машинного обучения. Эти улучшения позволят значительно сократить время инстанцирования и упростить процесс интеграции моделей машинного обучения в unikernel, делая их более доступными для широкого применения. Этот подход обеспечит гибкость при развертывании систем глубокого обучения в облачных и распределенных средах, снижая нагрузку на ресурсы и повышая производительность систем.

2. Предлагаемый подход оптимизации

Предлагаемая методика оптимизации выполняется по двум направлениям:

Методики настройки и оптимизации:

1. Ресурсно-ориентированная оптимизация. Этот подход направлен на минимизацию потребления вычислительных ресурсов при сохранении приемлемого уровня производительности модели. Ключевые методы включают квантование весов, структурную и неструктурную обрезку параметров (pruning), выбор эффективной архитектуры и регулирование параметров выполнения (например, batch size).
2. Адаптация модели под задачу. Здесь основное внимание уделяется подбору и корректировке гиперпараметров, таких как скорость обучения, коэффициенты регуляризации, размер обучающих пакетов и т.п., с целью достижения баланса между точностью, скоростью выполнения и устойчивостью модели к переобучению. Подобные настройки позволяют адаптировать модель к специфике используемых данных и вычислительной среды.

Пошаговое описание конвертации библиотек:

Шаг 1: Подготовка библиотек к конвертации. Первый этап заключается в оптимизации и адаптации исходных библиотек машинного обучения под требования специализированной среды исполнения.

Для PyTorch производится удаление лишних зависимостей, таких как torchvision, ONNX, CUDA и динамические загрузчики, чтобы минимизировать размер и сложность итогового пакета. Важной частью становится использование сериализованных форматов моделей – TorchScript – которые интегрируются через компактную бинарную структуру, встроенную непосредственно в исполняемый образ.

Для TensorFlow базой инференса служит легковесная версия – TensorFlow Lite. Все операции, связанные с файловым вводом/выводом, заменяются на обработку данных исключительно в оперативной памяти, что исключает обращение к файловой системе во время выполнения. Поддерживаются только вычисления с типом данных float32 без использования GPU, что упрощает и ускоряет процесс выполнения.

Для scikit-learn применяется компиляция кода с использованием Cython, а также статическая линковка ключевых библиотек numpy и scipy. Из пакета удаляются все интерфейсы, связанные с обучением, оставляя только функционал инференса. Такой подход значительно уменьшает конечный размер и улучшает стабильность работы.

Шаг 2: Внедрение MirageOS. Для последующей интеграции и исполнения подготовленных моделей используется операционная система MirageOS, которая ориентирована на работу в режиме unikernel. MirageOS устанавливается на машину с Linux-базированной системой, поскольку конвертация и развертывание на Windows не поддерживаются из-за особенностей архитектуры unikernel и требуемых низкоуровневых модификаций.

Шаг 3: Использование специализированных скриптов для трансфера и обновления образов. Для автоматизации процесса обновления и управления моделями в системе разработан набор скриптов. На стороне сервера запускается скрипт, ответственный за создание и отправку новых образов с конвертированными моделями.

В результате оптимизации и конвертации достигается значительное сокращение размера библиотек и моделей, что положительно сказывается на скорости развертывания и экономии ресурсов.

Уменьшения объема в результате оптимизации:

- PyTorch: исходный размер около 3.4 ГБ сокращается до 1.4 ГБ – почти в 2.5 раза.
- TensorFlow: 0.9 ГБ уменьшается до 0.3 ГБ – в 3 раза.
- scikit-learn: 0.7 ГБ снижается до 0.22 ГБ – более чем в 3 раза.

Использование подхода рассмотрим на примере задачи классификации изображений. Рассмотрим множество архивов A , где каждый архив a_i содержит изображения дефектных и недефектных деталей:

$$A = \{a_1, a_2, \dots, a_n\},$$

где a_i – i -й архив, представляющий собой множество изображений: $a_i = \{I_1, I_2, \dots, I_k\}$, где I_j – j -й файл изображения в архиве. Изображения делятся на два класса:

- Дефектные детали C_1 .
- Недефектные детали C_2 .

Каждое изображение имеет следующие атрибуты:

- Размер изображения S в кб.
- Разрешение изображения R : количество пикселей.
- Тип файла T .

$$I_j = \{S, R, T\}.$$

Для обработки изображений используется несколько моделей машинного обучения, представленных в модуле ModelFactory. Модуль ModelFactory создает различные модели для классификации изображений, имея в себе функцию build для каждой модели:

$$M = \{CNN, ResNet, DenseNet\}.$$

В качестве выбора среды исполнения используется $Z = \{VM, Docker, Unikernel\}$.

Каждая модель M_k имеет следующие параметры:

- A_k – точность: точность классификации.
- T_k – время обработки всех изображений в коллекции.
- P_k – объем памяти, необходимый для работы модели.

Процесс обучения и оптимизации модели включает работу следующих модулей (рис. 1):

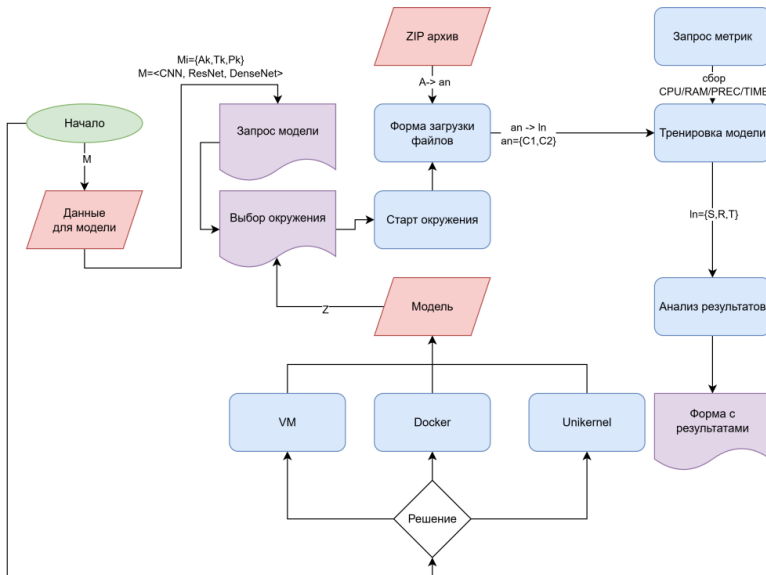


Рис. 1. Схема алгоритма

1. Модуль Trainer реализует функцию обучения и оптимизации с учетом гиперпараметров с помощью Keras Tuner. Алгоритм оптимизации включает в себя настройку гиперпараметров для каждой модели; обучение модели: с использованием тренировочного набора данных.
2. Модуль ModelFactory отвечает за создание моделей. Для каждой модели реализована функция сборки:

$$ModelFactory.build(M_i) \rightarrow M_z,$$

где M_z – модель, инициализированная для классификации изображений.

3. Модуль Trainer обучает модель и оптимизирует её параметры с помощью Keras Tuner:

$$T(M_z, D_{train}) \rightarrow O_z,$$

где O_z – оптимизированная модель. Тогда Целевая функция:

$$F(M_i) = \frac{1}{N_i} \sum_{n=1}^N \frac{1}{Z} \sum_{j=1}^Z L(O_z),$$

где L – функция потерь, Z – число моделей, N – число вариантов реализации, i – число сред исполнения. При этом $F(M_i)$ рассчитывается для каждой Z .

Оценка производительности для модели M_k : среднее время обработки: T_k ; точность A_k ; память P_k .

Оптимизация рассматривается с точки зрения использования памяти и процессора при развертывании, увеличение скорости обучения.

3. Результаты экспериментов

Целью экспериментов являются: демонстрация сокращения времени инстанцирования модели при использовании unikernel; сравнение потребления ресурсов (памяти и процессора) unikernel и традиционных методов развертывания (контейнеров, виртуальных машин); проверка корректности работы развернутой модели машинного обучения.

Процедура эксперимента включает в себя следующие шаги:

1. Подготовка моделей:
 - a. Простая CNN реализована и обучена локально.
 - b. ResNet50 и DenseNet121 загружаются предобученными из библиотеки Keras.
2. Развертывание моделей:
 - a. На unikernel: каждую модель упаковывают в unikernel с минимальными ресурсами (512 МБ памяти, 1 ядро).
 - b. В Docker: модели запускаются в контейнере с установленным TensorFlow.

- с. На VM: модели запускаются на полноценной операционной системе.
3. Обработка данных:
 - а. Каждой модели подают 100 изображений из тестового набора.
 - б. Замеряется среднее время обработки одного изображения.
4. Сбор данных: Для каждой среды фиксируются: время инстанцирования, время обработки, потребление ресурсов, точность классификации.

Входные данные:

- Датасет: Subset из ImageNet (100 изображений для классификации). При построении датасета особое внимание уделялось проблеме несбалансированных классов, что типично для производственных задач [Сааков, 2023].
- Модели:
 - CNN: Простая сверточная сеть с 2 сверточными слоями.
 - ResNet50: Предобученная модель на ImageNet.
 - DenseNet121: Предобученная модель на ImageNet.

Среды выполнения:

1. Unikernel (U): Использован NanoVMs OPS.
2. Docker (D): Использован стандартный контейнер Python с TensorFlow.
3. Виртуальная машина (VM): Среда RHEL8 с 4 ГБ памяти.

Метрики:

1. Время инстанцирования модели: время запуска до готовности принимать запросы.
2. Время обработки одного изображения.
3. Потребление ресурсов (память, процессор).
4. Точность классификации (Тор-1 Accuracy).

В табл. 1 приводятся значения метрик, полученных в ходе эксперимента по классификации изображений.

Таблица 1

Результаты эксперимента

	CNN			ResNet			Dense Net		
	U	D	VM	U	D	VM	U	D	VM
Время инстанцирования (с)	0.15	5.0	12.0	2.0	6.0	14.0	1.0	7.0	16.0
Время обработки (м)	26	47	42	32	52	50	22	39	37
Потребление памяти (ГБ)	7	12	12	7	12	12	6	12	12
Загрузка процессора (%)	100	100	95	100	100	95	100	100	95

Анализ результатов:

1. Время инстанцирования: Unikernel показывает наименьшее время запуска для всех моделей (0.15–0.25 с), тогда как Docker и VM значительно медленнее из-за сложности окружений.
2. Время обработки: Unikernel быстрее обрабатывает изображения (на 30–50% быстрее VM). Это обусловлено минимизацией системных накладных расходов.
3. Потребление ресурсов: Unikernel использует значительно меньше памяти и процессора, чем Docker и VM.
4. Точность классификации: Точность одинакова для всех сред, так как модели используют одну и ту же архитектуру.

Сравнение производительности моделей в разных окружениях

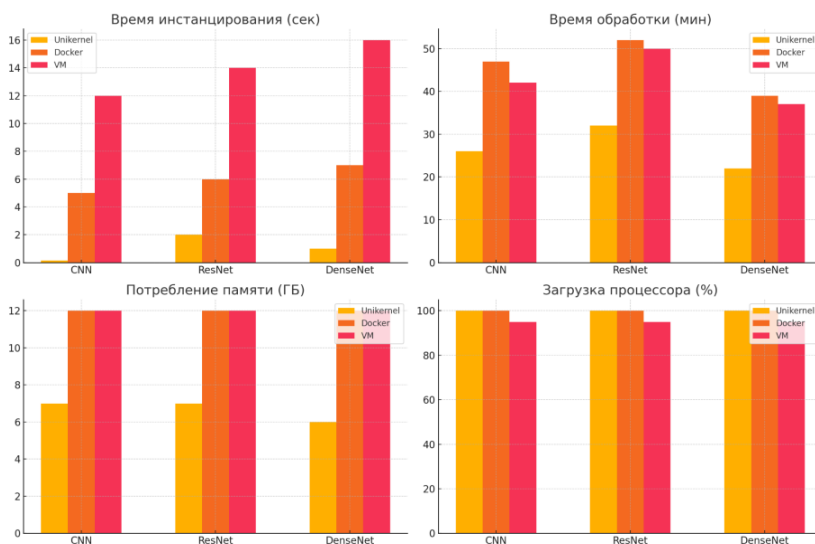


Рис. 2. Сравнение метрик использования ресурсов

1. Unikernel демонстрирует значительно более быстрое время инстанцирования по сравнению с традиционными подходами, что делает его подходящим для систем с требованиями реального времени (рис. 2).
2. Потребление памяти и загрузка процессора в среде unikernel ниже, что подтверждает его эффективность для оптимизации ресурсов.
3. Точность классификации идентична для всех сред, что подтверждает корректность работы unikernel.

Unikernel подходит для задач с высокими требованиями к быстродействию и низкими требованиями к памяти. Однако сложность настройки unikernel может быть препятствием для применения в широкомасштабных проектах без специализированных инструментов. Экспериментальные результаты данного исследования сопоставимы с выводами работы IEEE, в которой подтверждается превосходство unikernel-сред по времени инстанцирования и потреблению ресурсов [Ahuja, 2021].

Заключение

Проведённые тесты подтвердили корректность и эффективность разработанного подхода. Время запуска моделей составило от 0.6 до 1.2 секунд в зависимости от архитектурной сложности, потребление памяти снизилось на 40–60% по сравнению с контейнерными аналогами, а производительность в задачах классификации изображений осталась на высоком уровне.

К недостаткам описанной методики относятся: сложность интеграции с популярными библиотеками; ограниченная поддержка аппаратного ускорения; трудности отладки и мониторинга из-за минималистичной архитектуры; для развёртывания моделей в unikernel требуются специализированные навыки, что ограничивает круг пользователей. Также в данной работе тестировались только классические архитектуры (CNN, ResNet, DenseNet), но не рассматривались современные генеративные модели (например, GAN или Transformer), которые могут иметь уникальные требования к ресурсам.

Перспективами дальнейшего совершенствования методики являются доработка инструментов автоматизации, расширение поддержки библиотек и аппаратного ускорения, а также упрощение процессов отладки и интеграции для повышения степени универсальности.

Список литературы

- [Бабкина и др., 2024] Бабкина Е.А., Гаев Л.В. Автоматическое машинное обучение (automl): алгоритмы и инструменты для снижения порога входа // Международный журнал гуманитарных и естественных наук. – 2024. – № 6-1(93). – С. 175-178.
- [Донской, 2020] Донской В.И. Интеллектуальная оптимизация на основе машинного обучения: современное состояние и перспективы (обзор) // Таврический вестник информатики и математики. – 2020. – № 1(46). – С. 32-63.
- [Сааков, 2023] Сааков Д.В. Улучшение производительности алгоритмов машинного обучения с несбалансированными данными // Экономика строительства. – 2023. – № 4. – С. 73-77.
- [Ahuja, 2021] Ahuja A., Jain V. Challenges and Opportunities for Unikernels in Machine Learning Inference // 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). – IEEE, 2021. – С. 1-5.

- [Anari, 2025] Anari S. et al. Explainable attention based breast tumor segmentation using a combination of UNet, ResNet, DenseNet, and EfficientNet models // Scientific Reports. – 2025. – Vol. 15, No. 1. – P. 1027.
- [Chen et al., 2021] Chen S., Zhou M. Evolving container to unikernel for edge computing and applications in process industry // Processes. – 2021. – Vol. 9, No. 2. – P. 351.
- [Imada, 2018] Imada T. Mirageos unikernel with network acceleration for iot cloud environments // Proceedings of the 2018 2nd International Conference on Cloud and Big Data Computing. – 2018. – P. 1-5.
- [Imes, 2018] Imes C., Hofmeyr S., Hoffmann H. Energy-efficient application resource scheduling using machine learning classifiers // Proceedings of the 47th International Conference on Parallel Processing. – 2018. – P. 1-11.
- [Kong, 2022] Kong L. et al. Edge-computing-driven internet of things: A survey // ACM Computing Surveys. – 2022. – Vol. 55, No. 8. – P. 1-41.
- [Kurek, 2024] Kurek T., Niemiec M., Lason A. Performance evaluation of a firewall service based on virtualized IncludeOS unikernels // Scientific Reports. – 2024. – Vol. 14, No. 1. – P. 557.
- [Kuenzer, 2021] Kuenzer S. et al. Unikraft: fast, specialized unikernels the easy way // Proceedings of the Sixteenth European Conference on Computer Systems. – 2021. – P. 376-394.
- [Leon, 2020] Leon M. The dark side of unikernels for machine learning // arXiv preprint arXiv:2004.13081. – 2020.
- [Mitra, 2019] Mitra S. et al. Deepplace: Learning to place applications in multi-tenant clusters // Proceedings of the 10th ACM SIGOPS Asia-Pacific Workshop on Systems. – 2019. – P. 61-68.
- [Moebius et al., 2024] Moebius F., Pfandzelter T., Bermbach D. Are Unikernels Ready for Serverless on the Edge? // 2024 IEEE International Conference on Cloud Engineering (IC2E). – IEEE, 2024. – P. 133-143.
- [Tsiourvas, 2021] Tsiourvas A. et al. A mechanism design and learning approach for revenue maximization on cloud dynamic spot markets // 2021 IEEE 14th International Conference on Cloud Computing (CLOUD). – IEEE, 2021. – P. 427-432.

УДК 004.94

doi: 10.15622/rcai.2025.082

РАЗРАБОТКА ВИРТУАЛЬНОГО АНГАРА ДЛЯ ПРОЕКТИРОВАНИЯ И ИЗУЧЕНИЯ АВИАЦИОННЫХ КОНСТРУКЦИЙ

Н.М. Боргест (*borgest@yandex.ru*)

С.А. Власов (*vlasov.ssau@mail.ru*)

Самарский национальный исследовательский университет
имени академика С.П. Королёва, Самара

Статья посвящена разработке виртуального ангара – интерактивной среды для проектирования и изучения авиационных конструкций с использованием технологий виртуальной (VR) и дополненной реальности (AR), штатных систем автоматизированного проектирования (САПР). Приводится анализ технических, эргономических и образовательных аспектов, а также перспективы применения AR/VR в инженерной практике. Описаны этапы разработки виртуального ангара, включая создание 3D-моделей, алгоритмов конвертации объектами и интеграцию с интеллектуальным помощником проектанта. Проведена апробация решений в образовательном процессе, подтвердившая их эффективность. Предложены рекомендации по масштабированию и внедрению разработанных технологий в авиастроение и образование.

Ключевые слова: виртуальная реальность, дополненная реальность, авиационные конструкции, САПР, робот-проектант, виртуальный ангар.

Введение

В настоящее время активно развиваются технологические инновации в проектировании и образовании с использованием технологии виртуальной (VR), дополненной (AR) и смешанной реальностей (XR). На рис. 1 показан прогноз размера мирового рынка по AR/VR на 2025-2030 годы, составленный на основе оценочных данных, приведенных в работах [Dimiter et al., 2019],

[Radianti et al., 2023], [Pellas et al., 2021], [Cheng & Tsai, 2023], а также многочисленных обзорах^{1,2,3,4,5}.

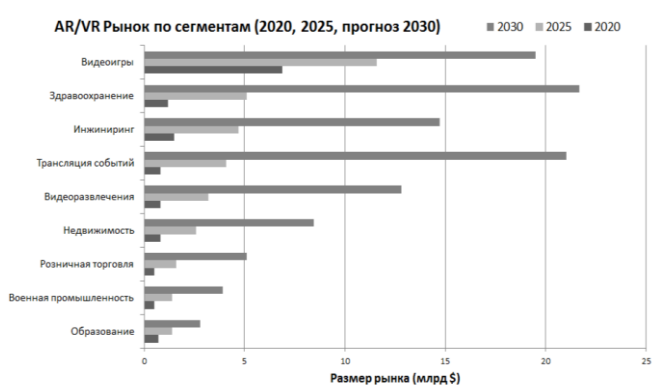


Рис. 1. Прогноз размера мирового рынка AR/VR

Актуальность исследований обусловлена широким распространением мобильных систем в виде смартфонов, планшетов, шлемов AR/VR и потребностью в создании программного обеспечения для синтеза и работы с виртуальными объектами по технологии XR с целью внедрения новых методик и технологий интерактивного обучения [Курасов et al., 2020], а также новых способов моделирования сложных объектов и процессов.

Работа является развитием исследований, проводимых на кафедре конструкции и проектирования летательных аппаратов Самарского университета, в области проектирования самолетов и подготовки специалистов для авиационной отрасли. В частности, основой для данной работы послужил интеллектуальный помощник проектанта самолета, в котором были заложены принципы описания предметной области в форме тезауруса, баз данных и знаний с использованием систем автоматизированного проектирования (САПР) [Боргест и др., 2015].

¹ Объём рынка виртуальной реальности вырастет до 51 млрд долларов к 2030 году. 9.11.2021. <https://skillbox.com/media/gamedev/obyem-rynka-virtualnoy-realnosti-vyrastet-do-51-mlrd-dollarov-k-2030-godu/>.

² Прогноз по рынку AR/VR 2025-2030. Май 2025. <https://tenchat.ru/media/3345922-prognoz-po-rynku-arvr-20252030>.

³ Как путешествие рынка виртуальной реальности к оценке в 57 миллиардов долларов к 2030 году изменит обучение. 25.10.2024 г. <https://www.binance.com/ru/square/post/15351397751137>.

⁴ Финансовые перспективы рынка виртуальной реальности. <https://vr-app.ru/blog/finansovye-perspektivy-rynka-virtualnoi-realnosti/>.

⁵ Инвестиции в VR-рынок: тренды, прогнозы, перспективные ниши. 27 июня 2025. <https://dzen.ru/a/aF5RDbjRpFDxvtAe>.

1. Интеллектуальный помощник проектанта

Разработанный интеллектуальный помощник – робот-проектант самолета – интегрирован в создаваемый виртуальный ангар. Робот поддерживает два режима: «инженер» (проектирование) и «студент» (обучение). Основные функции робота связаны с предварительным проектированием самолета на основе технического задания и включают процедуры выбора аэродинамической схемы, расчёта параметров, компоновки и центровки самолёта, разработку 3D модели самолета.

Интерфейс Робота-проектанта состоит из нескольких рабочих экранов (рис. 2): таймлайн проекта (1), интерактивный экран изменения проектных параметров (2), матрицы проекта, как набор данных, описывающий исследуемый проект (3) и окна предварительного просмотра 3D модели самолета (4) [Боргест et al., 2015].

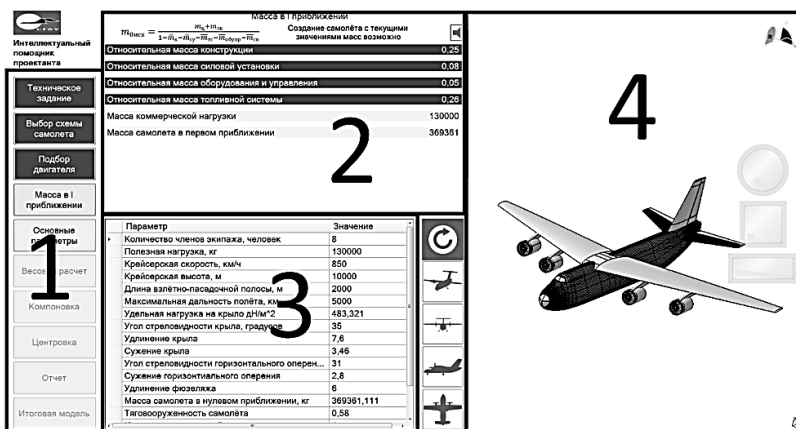


Рис. 2. Интерфейс робота-проектанта

Разработанная параметризованная 3D модель самолета позволяет масштабировать проект и автоматически перестраивать его под выбранные (заданные) или оптимизируемые (в случае применения процедур оптимизации) проектные параметры самолета. Полученная 3D модель проектируемого самолёта в дальнейшем используется в качестве основы для последующего конечно-элементного анализа конструкции самолета, проведения аэродинамических исследований в инженерных системах, а также для проведения физических экспериментов на модели, полученной на 3D принтере.

2. Виртуальный ангар авиационных конструкций

При разработке виртуального ангара, как информационной системы (ИС), учитываются контекст её использования и состав акторов – группы лиц или систем, взаимодействующих с ИС. На рис. 3 показан фрагмент UML диаграммы ИС.

В данной модели пользователь, пройдя предварительный тест на знание конструкции, погружается в искусственно созданную среду и знакомится с инструкцией, изучая способы перемещения и взаимодействия. Затем он открывает дверь ангара и перемещается в ангар. В виртуальном мире нет конкретного сценария и чёткой последовательности действий, пользователь может перемещаться куда угодно и начинать с любой точки ангара. При этом существует виртуальный гид, помогающий пользователю ознакомиться с тем или иным экспонатом. На каждом стенде имеется возможность взаимодействия с экспонатом: взять экспонат, разобрать или собрать его или выполнить измерение объекта. Пользователь может построить свой самолёт в виртуальном пространстве с использованием робота-проектанта.

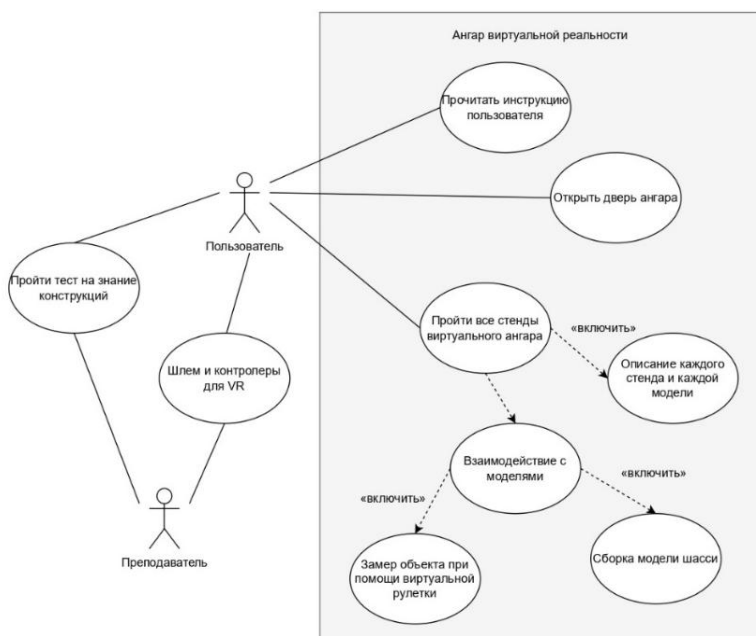


Рис. 3. Фрагмент диаграммы вариантов использования ИС

На рис. 4 в виде диаграммы IDEF0 изображены основные компоненты процесса создания ангара виртуальной реальности.



Рис. 4. Диаграмма IDEF0 с позиции разработчика

Для создания виртуального ангара авиационных конструкций использована кроссплатформенная среда разработки Unity 3D, а также Unity Asset Store – платформа от Unity с различными материалами необходимыми для разработки, где доступны инструменты, наборы моделей, шаблоны проектов, анимации, специализированные системы и др. Вид ангара в среде Unity показан на рис. 5.



Рис. 5. Вид ангара в среде Unity

На рис. 6 изображен стенд с интерактивной стойкой шасси самолета Ту-4, информацию о которой можно изучить как визуально, так и прослушать с помощью аудиогuida. Рядом со стойкой расположены её детали

и «призрак» её сборки, который может помочь пользователю при сборке. При взятии деталей над ними возникает надпись с названием, которая позволяет определить, куда и в какой последовательности устанавливать деталь в сборку.

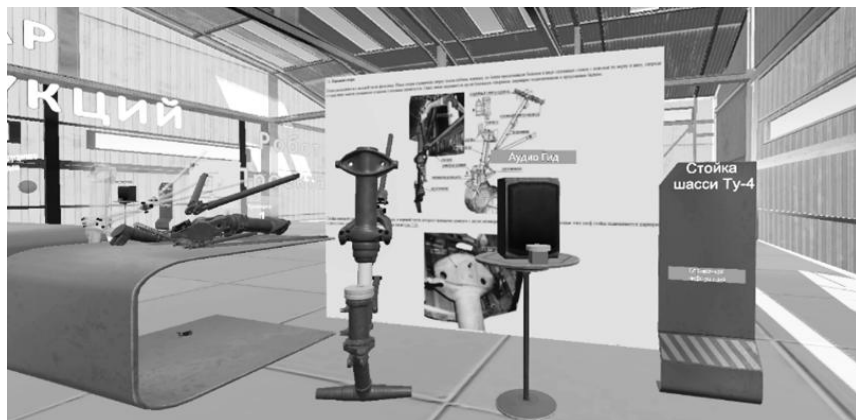


Рис. 6. Стенд "Стойка шасси Ту-4"

Пользователь может перемещаться по всей территории ангара за счёт использования специального компонента Steam VR⁶, TeleportArea⁷, но у каждого стенда в ангаре есть ещё TeleportPoin. Это один из заранее готовых объектов, поставляемых вместе с SteamVR, в нем уже есть графические объекты, и компоненты, которые позволяют переместиться в конкретную точку в пространстве. Объект TeleportArea схож с ним по своим функциям, но даёт возможность перемещаться в любую точку заранее заданной прямоугольной области, границы которой становятся видны при инициализации телепортации [Vlasov et al., 2020].

Для закрепления наименования основных деталей стойки шасси был разработан специальный скрипт, позволяющий производить сборку передней опоры стойки шасси с использованием контроллеров и шлема виртуальной реальности.

⁶ SteamVR – это универсальный инструмент для использования VR с любым оборудованием.

⁷ TeleportArea – это место телепортации, которое переносит пользователя в указанное им место на поверхности.

3. Конвертирование модели для автоматизированного проектирования в виртуальной реальности

Для интеграции 3D-моделей, созданных в САПР, в VR требуется преобразование данных из формата WRL (VRML, Virtual Reality Modeling Language) в OBJ – полигональный формат, используемый в Unity [Zhang, 2023]. Формат WRL, применяемый для представления сложных 3D-сцен, включает геометрию, текстуры, материалы, освещение и иерархию объектов. Однако для VR-приложения требуются упрощенные модели с низкой полигональностью и оптимизированными текстурами для обеспечения высокой производительности (60–90 кадров/с). Формат OBJ, поддерживающий только статическую геометрию и базовые материалы, подходит для этой задачи благодаря своей простоте и совместимости. Поэтому в данной работе для виртуального ангара разработан алгоритм конвертации WRL в OBJ в контексте автоматизированного проектирования [Patel et al., 2022].

Конвертация из WRL в OBJ направлена на преобразование параметризованных или иерархических данных WRL в упрощенный полигональный формат, сохраняя ключевые элементы модели – геометрию и текстуры. Основные задачи:

- минимизация потерь данных при преобразовании;
- оптимизация модели для VR (число полигонов < 100 000);
- обеспечение совместимости с VR-движками и САПР.

WRL-файл представляет собой текстовую или бинарную структуру [Müller, 2024], состоящую из узлов (nodes), таких как Shape, IndexedFaceSet (геометрия), Appearance (материалы и текстуры), Transform (трансформации). В отличие от WRL, OBJ хранит только вершины (v), нормали (vn), текстурные координаты (vt) и грани (f), а материалы описываются в сопутствующем MTL-файле (см., листинг 1). Конвертация включает следующие этапы.

- Парсинг WRL-файла: Извлечение геометрии (IndexedFaceSet), текстурных координат (TextureCoordinate) и материалов (Material, ImageTexture).
- Преобразование геометрии: Конвертация вершин и граней в формат OBJ, удаление ненужных узлов (анимации, освещение).
- Обработка текстур: Привязка текстурных координат и экспорт материалов в MTL.
- Оптимизация: Упрощение полигональной сетки для снижения нагрузки на VR-систему.
- Экспорт: Формирование OBJ и MTL файлов.

Конвертация позволила интегрировать модели из САПР в виртуальный ангар, обеспечив кадровую частоту (производительность) 90 FPS на конфигурации Intel Core i7, NVIDIA RTX 3060. Геометрия и текстуры сохраняются с минимальными потерями.

```

Shape{
  geometry IndexedFaceSet {
    coord Coordinate {
      point[0 0 0, 1 0 0, 1 1 0, ...]
    }
    coordIndex[0 1 2 - 1, ...]
    texCoord TextureCoordinate {
      point[0 0, 1 0, 1 1, ...]
    }
  }
  appearance Appearance {
    material Material { diffuseColor 0.8 0.8 0.8 }
    texture ImageTexture { url "texture.png" }
  }
}

```

Фрагмент управления роботом показан на рис. 7.

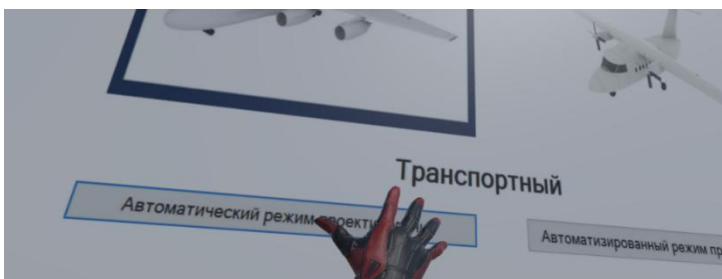


Рис. 7. Выбор режима проектирования проектантом в VR

На рисунке видно, как проектант выбирает опцию «Автоматический режим проектирования» и переходит к процессу проектирования военнотранспортного самолёта. На рис. 7 показана полигональная 3D модель самолёта, которая появляется в виртуальном пространстве за счёт кода конвертирования параметризированной модели. Все эти процессы происходят в фоновом режиме и не мешают нахождению проектанта в виртуальном ангаре.

Заключение

В работе рассмотрены основные направления применения технологии виртуальной реальности в САПР, разработан программный комплекс для проектирования самолета с использованием технологии VR. В виртуальный ангар вошли созданные трёхмерные модели самолётов с использованием технологий реверс-инжиниринга, а также виртуальный помощник

проектанта, позволяющий строить параметризованную 3D модель по заданному техническому заданию. Разработан алгоритм, позволяющий конвертировать параметризованную модель в полигональную, что позволило интегрировать модели САПР в VR среду.

Благодарности. Авторы выражают благодарности студентам Самарского университета им. С.П. Королёва Никите Кондрашину и Сухайли Ширинбегзода за помощь в создании виртуального ангара.

Список литературы

- [Боргест и др., 2015] Боргест Н.М., Власов С.А., Громов А.Г. и др. Робот-проектант: на пути к реальности // *Онтология проектирования*. – 2015. – Т. 5, № 4(18). – С. 429-449.
- [Курасов и др., 2020] Курасов Д.А., Подвальная Е.Ф. Использование инновационных образовательных технологий виртуальной реальности (VR) в обучении и промышленности // *Инновационные наукоемкие технологии*. – 2020. – № 4. – С. 190-192.
- [Dimitier et al., 2019] Dimitier V., Plamena Z. Analysis of v-Commerce as the New Online Sales Channel // *International Journal of e-Education, e-Business, e-Management and e-Learning (IJEEEE)*. – 2019. – Vol. 96 No. 2. – P. 131-137. – doi: 10.17706/ijeeee.2019.9.2.131-137. 1.
- [Radianti et al., 2023] Radianti J., Majchrzak T. A., Fromm J., Wohlgenannt I. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda // *Computers & Education*. – 2023. – Vol. 187. – P. 104528. – doi: 10.1016/j.compedu.2022.104528.
- [Pellas et al., 2021] Pellas N., Mystakidis S., Kazanidis I. Immersive Virtual Reality in K-12 and Higher Education: A systematic review of the last decade scientific literature // *Virtual Reality*. – 2021. – Vol. 25, No. 3. – P. 835-861. – doi: 10.1007/s10055-020-00489-9.
- [Cheng & Tsai, 2023] Cheng K.H., Tsai C.C. The interaction of child–parent shared reading with an augmented reality (AR) picture book and parents' conceptions of AR learning // *British Journal of Educational Technology*. – 2023. – Vol. 54, No. 1. – P. 396-414. – doi: 10.1111/bjet.13271.
- [Garcia et al., 2023] Garcia M., Ivanov D. Automated 3D Model Processing for Immersive Virtual Reality: From WRL to OBJ and Beyond // *International Conference on Computer Graphics and Interactive Techniques*. Tokyo, Japan, 15–17 Nov. 2023. Proceedings. Vol. 2. – Tokyo: ACM Press, 2023. – P. 89-98. – doi: 10.1145/3587421.3587432.
- [Patel et al., 2022] Patel R., Singh A., Kumar V. Optimizing VRML to OBJ Conversion for Real-Time Rendering in Virtual Environments // *SIGGRAPH Asia 2022 Technical Papers*. – 2022. – Vol. 41, No. 6. – P. 123-134. – doi: 10.1145/3550469.3550501.
- [Vlasov et al., 2020] Vlasov S., Borgest N. Application of VR and AR Technologies in Educational Process // *Advances in Intelligent Systems Research*. – 2020. – Vol. 174. – P. 78-85.
- [Zhang et al., 2023] Zhang L., Kim H. Advances in 3D File Format Conversion for Virtual Reality Applications // *IEEE Transactions on Visualization and Computer Graphics*. – 2023. – Vol. 29, No. 5. – P. 2567-2578. – doi: 10.1109/TVCG.2022.3156789.

УДК 004.89

doi: 10.15622/rcai.2025.083

ОНТОЛОГИЧЕСКИЙ ПАТТЕРН ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ РЕАБИЛИТАЦИОННОЙ КОМАНДЫ¹

В.В. Грибова (*gribova@iacp.dvo.ru*)

Е.А. Шалфеева (*shalf@dvo.ru*)

Институт автоматики и процессов управления ДВО РАН,
Владивосток

Одним из ключевых направлений оптимизации реабилитационного процесса является внедрение интеллектуальных компьютерных систем, способных обеспечивать поддержку принятия клинических решений на основе комплексного анализа медицинских данных. Использование онтологического инжиниринга и методов логического вывода на основе формализованных знаний создают основу для агрегирования экспертных и нормативных знаний и применения методов объяснимого искусственного интеллекта. Целью исследования является разработка онтологического паттерна для создания интеллектуальных компьютерных помощников в сфере реабилитации, способных к анализу клинических документов и медицинских знаний. Разработанный онтологический паттерн включает структуру электронной медицинской карты с информацией о пациенте; онтологии знаний для клинических задач; специализированные решатели; медицинский терминологический словарь реабилитологов; компоненты графического пользовательского интерфейса (GUI-компоненты); инструментарий для наполнения и развития знаний. Построение систем поддержки принятия решений с помощью паттерна состоит в наполнении баз знаний и их интеграции со специализированными решателями и GUI-компонентами. Такие системы призваны оказывать эффективную помощь в решении актуальных проблем современной реабилитологии от диагностики функциональных нарушений до планирования лечебных программ.

Ключевые слова: онтологический паттерн, цели реабилитации, МКФ-коды, план реабилитационных мероприятий.

¹ Работа выполнена при финансовой поддержке РНФ (проект № 25-11-20021).

Введение

Современная система здравоохранения сталкивается с растущей необходимостью повышения эффективности реабилитационных мероприятий, что обусловлено увеличением числа пациентов с хроническими заболеваниями, последствиями травм и возрастными нарушениями. Одним из ключевых направлений оптимизации реабилитационного процесса является внедрение интеллектуальных компьютерных систем, способных не только автоматизировать рутинные задачи, но и обеспечивать поддержку принятия клинических решений (прежде всего, диагностики, планирования и контроля реабилитационных вмешательств) на основе комплексного анализа медицинских данных. Однако применение новейших методов искусственного интеллекта, в частности нейросетевых технологий, в медицинской реабилитации остаётся ограниченным в силу ряда факторов. Во-первых, для их эффективного функционирования требуются обширные массивы обучающих данных, которые зачастую отсутствуют или недостаточно формализованы. Во-вторых, критически важным аспектом является обеспечение доверия со стороны медицинских специалистов, что предполагает необходимость объяснимости результатов, выдаваемых системами поддержки принятия решений (СППР). В этой связи перспективным направлением представляется использование онтологических паттернов [Ghosh et al., 2021], [Боровикова, 2017] и методов логического вывода на основе формализованных знаний, которые позволяют закладывать в СППР свойства прозрачности, согласованности и интеграции с другими эффективными технологиями поиска решений. Онтологический подход обеспечивает унификацию клинических данных, нормативных документов и экспертных знаний, что создаёт основу для применения методов объяснимого искусственного интеллекта (ХАИ).

Целью данного исследования является разработка онтологического паттерна для создания интеллектуальных помощников и СППР в сфере реабилитации, способных агрегировать экспертные и нормативные знания.

1. Материалы

Для систематизации существенных понятий предметной области и установления взаимосвязей между ними требуется анализ клинических рекомендаций, методов принятия решений и практик реабилитационного ведения пациентов [Николаев и др., 2024]. В рамках исследования были выбран раздел медицины «травматология» и отобраны нозологии, характеризующиеся высокой распространённостью и наличием регламентированных клинических рекомендаций (КР), включая: переломы проксимального отдела бедренной кости, переломы диафиза и дистальных отделов бедренной кости (за исключением проксимальных), перелом шейки

бедра, повреждения связочного аппарата коленного сустава, переломы и вывихи грудного, поясничного и крестцового отделов позвоночника [Дубров и др. 2021], [Клинические, 2024].

Дополнительно рассмотрены варианты оперативных вмешательств, в частности гемиартропластика тазобедренного сустава [Ямщиков и др., 2024], [Дубров и др., 2021], а также ограниченный датасет, содержащий клинические описания пациентов и индивидуальные планы реабилитации.

В ходе анализа выявлены задачи, требующие интеллектуальной поддержки:

- Формирование кодов Международной классификации функционирования (МКФ) как структурных компонентов комплексного диагноза, отражающего не только нозологию, но и уровень функциональных нарушений [Ivanova et al., 2022].
- Установление целей и задач реабилитации на основе интеграции клинических данных, стандартов лечения и индивидуальных особенностей пациента.
- Подбор оптимальных методик и инструментов реабилитации с учётом доказательной базы и ресурсных ограничений.

Интеграция поддержки разных задач основана на аккумуляции всех значимых данных и принятых предварительных и окончательных решений в едином структурированном документе – электронной медицинской карте (ЭМК).

Полученные результаты позволяют определить направления расширения разработанной ранее онтологии предметной области, интегрированной в платформу MedPortal IACPaaS [Gribova et al., 2023], за счёт включения модулей, ответственных за планирование реабилитационных мероприятий. Это предполагает также разработку специализированных онтологических решателей, способных осуществлять логический вывод на основе структурированных клинических знаний и формализованных нормативных документов. Реализация данного подхода призвана обеспечить повышение эффективности поддержки принятия решений в клинической практике реабилитологии.

2. Методы

Для создания систем поддержки принятия решений, способных к непрерывному совершенствованию знаний в процессе эксплуатации при сохранении прозрачности и верифицируемости, нами применяется онтологический подход. Его преимущества заключаются в: обеспечении структурной согласованности знаний; поддержке логического вывода; возможности развития системы экспертами предметной области.

Для генерации множества специализированных СППР, решающих актуальные задачи медицинских учреждений (в частности, для реабилитационных команд), нами разработана концепция **онтологических медицинских паттернов**. Каждый паттерн включает:

- Унифицированную структуру информации о пациенте – ЭМК или реабилитационную карту.
- Предметно-ориентированные онтологии знаний (формализующие структуру знаний о связях понятий для решаемых задач, сопровождаемую онтологическими соглашениями об их интерпретации).
- Специализированные решатели (алгоритмы логического вывода) и GUI-компонент.
- Медицинский терминологический словарь.
- Инструментарий для наполнения и развития знаний.

Структура информации о пациенте – ЭМК обеспечивает: формирование и хранение реальных клинических случаев (создание формализованных датасетов), которая может выступать в дальнейшем для методов машинного обучения, формирования прецедентной базы, создания тестовых примеров для валидации СППР.

Предметно-ориентированная онтология знаний определяет: семантические связи между понятиями, структуру хранения знаний, правила и ограничения интерпретации данных.

Специализированные решатели – программные модули, реализующие формальные механизмы обработки знаний для решения конкретных классов медицинских задач в рамках СППР. Они обеспечивают прозрачность механизмов вывода, которые критически важны для клинического доверия и валидации решений.

Медицинский терминологический словарь в структуре онтологического паттерна представляет собой формализованную систему описания терминов, обеспечивающую семантическую согласованность данных и знаний.

Инструментарий для наполнения и развития знаний представляет собой комплекс программных средств, обеспечивающих формирование, непрерывное обновление и валидацию знаний в СППР. Он включает следующие ключевые компоненты: редакторы актуализации онтологий, редакторы знаний, управляемые онтологиями, для экспертного формирования баз знаний и средства коллективной работы экспертного сообщества, программные средства автоматического извлечения знаний из текстов клинических рекомендаций с использованием больших языковых моделей (LLM); для снижения трудоемкости создания интерфейсов паттерн дополняется средствами автоматической генерации пользовательских интерфейсов СППР на основе шаблонов (GUI-компонент).

Технологическим стеком реализации является платформа IACPaas, обеспечивающая: повторное использование апробированных онтологий; интеграцию существующих решателей; применение готовых паттернов (например, для генерации МКФ-кодов реабилитационного диагноза).

3. Результаты

Структурно-содержательные характеристики элементов паттерна соответствуют задачам работы междисциплинарной команды реабилитологов, а именно: формирование комплексного МКФ-диагноза; целей и задач реабилитации; плана реабилитации как комплекса реабилитационных методик.

Структура информации о пациенте (ЭМК) основана на принципе детерминированного размещения информационных элементов в строго определенных структурных единицах документа (рис. 1). Такая организация обеспечивает однозначную локализацию данных для всех категорий пользователей, включая как людей, так и программных агентов, что гарантирует сохранение целостности информационных потоков при выполнении операций чтения и записи.

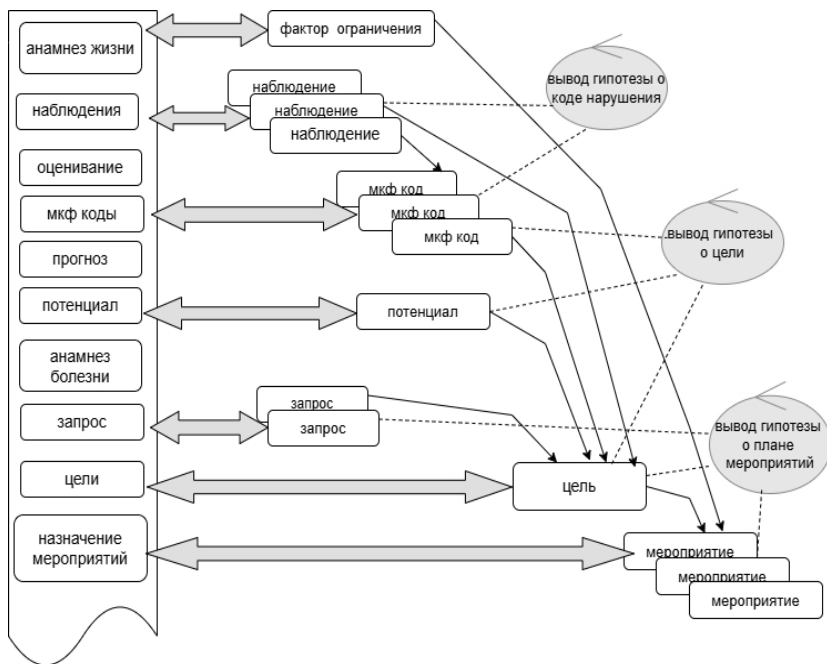


Рис. 1. Схема связи задач и документируемых сведений при реабилитации

Анализ порядка ведения пациентов, формирования документации и существующей взаимозависимости мнений и решений участников команды реабилитологов дал основание для определения *комплекта онтологических моделей*, достаточных для интеллектуальной поддержки их деятельности (вышеперечисленных трех важных этапов).

Онтология знаний о связях понятий для формирования полного МКФ-диагноза на основе тестов со шкалами и наблюдений определена ранее и прошла апробацию [Грибова и др., 2025].

Онтология знаний о связях понятий для формирования целей реабилитации после перенесенного заболевания или операции такова:

Заболевание | Проведенная операция ('set')

Цель на весь курс реабилитации ('set')

ЗАВИСИТ ОТ

АЛЬТЕРНАТИВА

- Факт ('list')

Тип факта

(МКФкод,

запрос,

реабилитационный потенциал,

Элемент анамнеза)

АЛЬТЕРНАТИВА

Элемент-Характеристика факта ('set')

Значение ('set')

(Качественное,

Числовое или Балл)

+ реабилитационный потенциал,

- Комплекс ('set') {

Факт ('list')

Тип факта}

совместимость элементов.

ДОСТИГАЕТСЯ ЧЕРЕЗ

Цель краткосрочная ('set')

Онтология знаний о связях понятий для формирования плана реабилитационных мероприятий состоит из двух частей: для конструирования индивидуального плана и для описания условий применения типовых планов.

На онтологическом портале эксплуатируется онтология для описания зависимости лечебно-восстановительного мероприятия (с подробностями: средство, режим | способ применения) от условий: особенности организма, результаты осмотров, проведенные вмешательства-операции, цели.

Онтология представления знаний о типичных планах реабилитационных мероприятий имеет следующий вид:

План мероприятий

мероприятие | методика ('set')

средство | аппарат, инструмент

режим | способ применения аппарата, инструмента

упражнение | процедура

ПРИВЯЗАН К

Цель краткосрочная ('set')

ПРОТИВОПОКАЗАНИЕ

- Факт ('list')

Тип факта

(реабилитационный потенциал,

элемент анамнеза)

Онтологические соглашения для интерпретации знаний и фактов.

Правила интерпретации структурированных знаний и данных отражают соглашения экспертов предметной области, документируются (записываются явно) и являются одной из опор программирования решателей.

В рамках исследования были зафиксированы:

- правило о выдвижении гипотезы о *Цели на весь курс реабилитации* (из базы знаний), если факты в реабилитационной карте (истории болезни) «согласуются» со знаниями, т.е. соответствуют фактам (одному или Комплексу Фактов), связанным в базе знаний с этой целью связью «ЗАВИСИТ ОТ»; Комплекс условий считается выполненным, если указаны (и совпадают) все перечисленные в нем факты;
- правило о выражении *Цели на весь курс реабилитации* через ограниченный набор *Целей краткосрочных*;
- правило о выдвижении гипотезы о необходимости назначения *Плана мероприятий*, если в базе знаний он хранится в связке «ПРИВЯЗАН К» с *Целью краткосрочной*, которая зафиксирована в реабилитационной карте.

Специализированные решатели соответствуют выделенными этапами процесса и реализованы на базе формализованных онтологических моделей, включая механизмы обработки онтологических соглашений:

- решатель выдачи гипотез-кодов МКФ (по наблюдениям и шкалам);
- решатель выдачи гипотез о целях и задачах в соответствии с правилом выдвижения гипотезы о *Цели на весь курс реабилитации*;
- решатель выдачи гипотез о плане реабилитации или его компонентах в соответствии с правилом выдвижения гипотезы о необходимости назначения *Плана мероприятий*, и правилом учета ограничений по возможности назначения пациенту активности / мероприятия.

Обработка имеет несколько этапов: поиск в описании текущей ситуации имен фактов для подтверждения условий для гипотезы, поиск в описании текущей ситуации значений фактов для подтверждения варианта

либо опровержения (попутно производится вывод в отчет-объяснение каждого подтверждаемого условия проверяемого процесса и всех фактов, соответствующих подтверждаемому условию). Решатели проходят апробацию на медицинском портале IACPaas.

Медицинский терминологический словарь. Интегрируемость с клиническими документами, другими программными средствами и источниками знаний (экспертных, нормативных, прецедентных) стимулирует к развитию возможностей словаря, входящего в онтологический базис. Разработанный ранее коллективом терминологический словарь дополнен дополнительными описаниями терминов:

- Словарь-тезаурус наблюдений с синонимией расширен ссылками на коды-идентификаторы из перечней Минздрава и международных номенклатур клинических терминов.
- Словарь-тезаурус активностей (мероприятий, воздействий) расширен иерархическим представлением мероприятий от общих методов до конкретных упражнений, используемых аппаратов, процедурных аспектов, режимов применения.

Инструментарием для наполнения и развития знаний являются инструменты платформы IACPaas, а также набор специализированных инструментов медицинского портала, включая средства генерации компонентов адаптивного пользовательского интерфейса на основе шаблонов (GUI-компонентов).

Обновленный паттерн включает:

- структуру ЭМК с информацией о пациенте;
- онтологии знаний для трех подзадач (структуры для хранения знаний и соглашения экспертов о правилах их интерпретации для вывода гипотезы о решении);
- три специализированных решателя (алгоритмы логического вывода);
- медицинский терминологический словарь реабилитологов;
- повторно используемый (адаптируемый к терминологии и профилю знаний) GUI-компонент для координации и распределения совместной работы (по разным направлениям) на «вкладках» единого «диалогового пространства»;
- инструментарий для наполнения и развития знаний, инструментарий для оценивания качества баз знаний (по контрольным прецедентам).

Построение версии СППВР на основе паттерна состоит в наполнении баз знаний и ее интеграции со специализированными решателями и GUI-компонентами. ЭМК с информацией о пациенте, базы знаний и информация на «вкладках» GUI выражаются в терминах единого медицинского словаря реабилитологов.

Заключение

Обсуждение представленного исследования позволяет выделить ключевые достижения и перспективные направления развития предложенного подхода к созданию интеллектуальных систем поддержки принятия решений в области медицинской реабилитации. Разработанный авторами онтологический паттерн демонстрирует эффективное решение актуальных проблем современной реабилитологии, связанных с необходимостью персонализации лечебных программ, обеспечением прозрачности и обоснованности выдаваемых рекомендаций, а также интеграцией разнородных источников медицинских знаний. Важным преимуществом исследования является его комплексный характер, охватывающий весь цикл реабилитационного процесса – от первичной диагностики функциональных нарушений до планирования индивидуальных вмешательств.

Практическая значимость работы заключается в апробировании предложенного решения на конкретных нозологиях, таких как переломы проксимального отдела бедренной кости и повреждения позвоночника, а также учетом реальных клинических сценариев, включая послеоперационную реабилитацию после эндопротезирования. Техническая реализация на платформе IACPaaS обеспечивает важные преимущества с точки зрения масштабируемости решения и возможности его адаптации к различным медицинским учреждениям. Особого внимания заслуживает разработка специализированных решателей для конкретных задач реабилитолога, что существенно повышает практическую ценность системы.

Перспективы дальнейшего развития исследования видятся в нескольких направлениях. Во-первых, требуется расширение базы прецедентов для контроля качества баз знаний. Во-вторых, представляет интерес интеграция предложенного подхода с большими языковыми моделями для обработки неструктурированных клинических записей. В-третьих, необходима комплексная клиническая валидация системы, включающая контролируемые испытания ее эффективности, оценку влияния на рабочую нагрузку специалистов и анализ улучшения клинических исходов пациентов. Следует отметить и существующие практические ограничения, прежде всего, связанные с необходимостью значительных усилий по интеграции с существующими электронными системами здравоохранения. Дальнейшее развитие системы должно быть направлено на повышение степени автоматизации обработки знаний по диагностике, назначению персонализированных реабилитационных мероприятий и прогнозу реабилитационного потенциала.

Список литературы

- [Боровикова и др., 2017] Боровикова О.И., Загоруйко Г.Б., Загоруйко Ю.А., Шестаков В.К. Использование паттернов для разработки онтологии информационно-аналитического интернет-ресурса «поддержка принятия решений» // Информационные и математические технологии в науке и управлении. – 2017. – № 3(7). – С. 171-183.
- [Грибова и др., 2025] Грибова В.В. [и др.]. Комплекс онтологий как модель системы интеллектуальной поддержки в реабилитации пациентов, перенесших инсульт // Программные системы: теория и приложения. – 2025. – Т. 16, № 1(64). – С. 61-82. – doi: 10.25209/2079-3316-2025-16-1-61-82.
- [Дубров и др., 2021] Дубров В.Э. [и др.]. Переломы проксимального отдела бедренной кости. Клиника, диагностика и лечение (Клинические рекомендации, в сокращении) // Вестник травматологии и ортопедии им. Н.Н. Приорова. – 2021. – Т. 28, № 4. – С. 49-89. – doi: 10.17816/vto100763.
- [Клинические, 2024] Клинические рекомендации. переломы бедренной кости (кроме проксимального отдела бедренной кости). – 2024. – URL: https://legalacts.ru/doc/klinicheskie-rekomendatsii-perelomy-bedrennoi-kosti-krome-proksimalnogo-otdela-bedrennoi_1.
- [Николаев и др., 2024] Николаев Н.С., Преображенская Е.В., Петрова Р.В., Андреева В.Э. Полный цикл медицинской реабилитации пациентов после травматолого-ортопедических операций на примере профильного федерального центра // Национальное здравоохранение. – 2024. – Т. 4, № 4. – С. 23-36. – doi: 10.47093/2713-069X.2023.4.4.23-36.
- [Ямщиков и др., 2024] Ямщиков О.Н., Емельянов С.А., Чумаков Р.В. Гемиартропластика тазобедренного сустава биполярным эндопротезом при несостоятельном остеосинтезе шейки бедренной кост // Уральский медицинский журнал. – 2024. – Т. 23(1). – С. 112-120. – doi 10.52420/2071-5943-2024-23-1-112-120.
- [Ghosh et al., 2021] El Ghosh M. et al. Modeling logical definitions in biomedical ontologies by reusing ontology design patterns // ICBO'21: International Conference on Biomedical Ontologies. – 2021. – Vol. 3073. – P. 56-62.
- [Gribova et al., 2023] Gribova V.V., Moskalenko P.M., Timchenko V.A., Shalfeeva E.A. The IACPaaS Platform for Developing Systems Based on Ontologies: A Decade of Use // Scientific and Technical Information Processing. – 2023. – Vol. 50(5). – P. 406-413.
- [Ivanova et al., 2022] Ivanova G.E., Bodrova R.A., Builova T.V., Karimova G.M., Komarnitsky V.S. Algorithm for formulation a rehabilitation diagnosis using the international classification of functioning in a patient with a stroke: clinical case // Physical and rehabilitation medicine, medical rehabilitation. – 2022. – Vol. 4(1). – P. 37-54. – doi: 10.36425/rehab96918.

УДК 620.9:004

doi: 10.15622/rcai.2025.084

СЕРВИСЫ ДЛЯ ОБРАБОТКИ ДАННЫХ И ЗНАНИЙ В ЭКОСИСТЕМЕ ЗНАНИЙ В ЭНЕРГЕТИКЕ¹

Л.В. Массель (*massel@isem.irk.ru*)

А.Г. Массель (*amassel@isem.irk.ru*)

В.Р. Кузьмин (*kuzmin_vr@isem.irk.ru*)

Т.Г. Мамедов (*mamedowtymur@isem.irk.ru*)

Д.А. Гаськова (*gaskovada@isem.irk.ru*)

Институт систем энергетики им. Л.А. Мелентьева СО РАН,
Иркутск

В работе описывается предлагаемый подход к интеграции сервисов в рамках архитектуры экосистемы знаний в энергетике. В коллективе, который представляют авторы, разработаны программные решения для реализации семантического моделирования, программный комплекс для прогнозных исследований топливно-энергетического комплекса с точки зрения энергетической безопасности, а также информационно-вычислительная система для проведения вычислительных экспериментов и решения задач по оценке выбросов загрязняющих веществ от объектов энергетики. В работе рассматривается проектирование взаимодействия существующих решений как сервисов разрабатываемой экосистемы знаний в энергетике.

Ключевые слова: экосистема знаний, энергетика, семантическое моделирование, сервисы, обработка знаний.

Введение

В современной научной практике решение комплексных исследовательских задач требует применения разнообразных программных комплексов и решений. Спектр используемых решений охватывает специализированные проприетарные продукты с закрытым исходным кодом, про-

¹ Результаты получены в рамках выполнения проекта по госзаданию ИСЭМ СО РАН FWEU-2021-0007 № AAAA-A21-121012090007-7.

граммные комплексы с открытым исходным кодом, универсальные решения общего назначения и авторские научные разработки, созданные для решения конкретных исследовательских задач.

Особую актуальность необходимость применения разнообразных вычислительных продуктов приобретает в контексте междисциплинарных научных исследований, к которым относится и энергетический сектор. Сложность выполнения таких исследований зачастую выходит за рамки одной системы. Целью настоящей работы является разработка подхода к интеграции существующих научных разработок в области энергетики, выполненных в коллективе авторов, в рамках единой среды для обмена знаниями и результатами исследований – экосистемы знаний. Однако, перед тем как переходить к описанию предлагаемого подхода к интеграции, в работе будут рассмотрены аналогичные российские и зарубежные разработки.

В настоящее время, существует ряд решений связанных с построением баз знаний и схожих программных решений, как в России, так и зарубежом, однако, провести сравнение с ними затруднительно, так как они являются проприетарными, а также решают несколько иные задачи. В России в качестве примера можно привести «Систему управления знаниями Росатома» (ЕИП РТТН) [Управлять знаниями, 2021], предназначенную для мониторинга и контроля реализации Комплексной программы РТТН, а также создания единого хранилища и организации информационной поддержки участников реализации программы «Развитие техники, технологий и научных исследований в области использования атомной энергии в Российской Федерации на период до 2024 года». Также стоит отметить информационную базу знаний «Россети Сибирь», одной из задач которой является повысить эффективность основных бизнес-процессов за счет формирования системного подхода к управлению знаниями. Проект направлен на совершенствование существующей системы управления и развития, управления корпоративными знаниями [Россети Сибирь разработали и внедрили систему управления знаниями, 2022]. За рубежом также разработан ряд решений в этой отрасли, например, References+ – это корпоративная соцсеть для сотрудников компании Siemens, которая позволяет находить экспертов, обмениваться опытом в режиме чата и снижать время решения проблем [Muller et.al, 2012]. Остальные зарубежные решения, описанные в [Top 10: Energy Management Platforms, 2025], направлены больше на цифровизацию энергетики и оптимизацию работы с точки зрения энергоэффективности с использованием ИИ, чем на создание программных продуктов, связанных со знаниями. Для рассмотренных решений отсутствуют описания архитектуры, а также применённых методов к интеграции данных и знаний.

Далее в работе рассматривается интеграция программных комплексов по двум направлениям:

- Исследования направлений развития топливно-энергетического комплекса (ТЭК) страны проводятся в подавляющем большинстве случаев с применением программных средств, реализующих сценарный подход, позволяющий проводить исследования в условиях неопределённости, а оценка результатов зачастую осуществляется на экспертном уровне. Такие исследования проводятся, например, с помощью модели NEMS (США) [Gabriel, 2001], предназначенной для прогнозирования развития экономики США с особым акцентом на энергетический сектор, программного комплекса SCANNER (ИНЭИ РАН) для исследования энергетики страны и мира [Макаров, 2011] или программного комплекса ИНТЭК (ИСЭМ СО РАН), обеспечивающего поддержку научных расчетов и моделирования сценариев развития ТЭК с учётом энергетической безопасности.

- Комплексные исследования по оценке загрязнения окружающей среды объектами энергетики подразумевают взаимодействие специалистов из разных областей, а сами исследования зачастую связаны с нормативными методиками проведения необходимых расчётов. Эти факторы, во многих случаях ограничивают применение таких зарубежных программных комплексов как CollectER (европейское решение) [European Environment Agency, 2023] или программных реализаций методик HYSPLIT [Stein, 2015] и AERMOD [EPA, 2004].

1. Компоненты экосистемы знаний

Сам термин «экосистема знаний» появился более 20 лет назад [Shrivastava, 1998], но строгого определения не обрёл. Авторы рассматривают экосистему знаний как специализированное направление развития более проработанной технологии цифровых экосистем [Массель, 2023]. Под этим термином обычно понимают сеть взаимосвязанных цифровых технологий, платформ и услуг, взаимодействующих друг с другом для создания ценности для бизнеса и потребителей. Однако, ключевой ценностью в рассматриваемой концепции являются знания и обмен ими между пользователями.

В работе рассматривается подход, развиваемый в авторском коллективе, в котором предлагается разработка архитектуры цифровой платформы экосистемы знаний. В соответствии с определением Основных направлений реализации цифровой повестки Евразийского экономического союза до 2025 года [Решение Высшего Евразийского экономического совета, 2017] авторы предлагают, что архитектура цифровой платформы экосистемы знаний в энергетике должна включать не только компьютерные и математические модели, но и базы данных и знаний, информационную и интеллектуальную подсистемы, подсистему визуализации и вспомогательные сервисы.

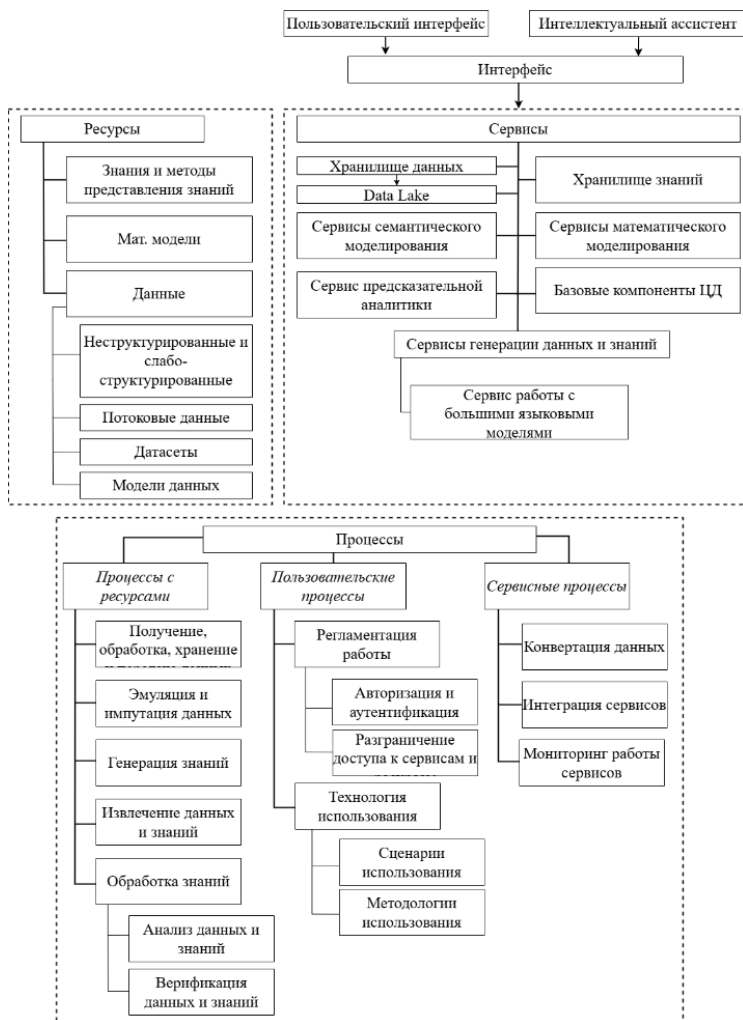


Рис. 1. Архитектура цифровой платформы экосистемы знаний в энергетике

Предлагаемая архитектура экосистемы знаний в энергетике представлена на рис. 1, подробное описание приведено в работе Л.В. Массель, А.Г. Масселя и В.Р. Кузьмина «Методы построения экосистемы знаний на примере энергетики» в этом же сборнике, здесь же отметим, что в ней выделяются три основных группы:

1. Ресурсы – включают в себя знания (а также модели их представления), математические модели и данные.

2. Процессы – включают процессы с ресурсами, пользовательские процессы, сервисные процессы.
3. Сервисы – включают хранилища данных и знаний, сервисы математического и семантического моделирования, а также сервисы предсказательной аналитики.

Ключевым аспектом реализации этого подхода является интеграция существующих в коллективе научных программных комплексов рамках этой архитектуры. Такая интеграция подразумевает определение и формализацию сценариев взаимодействия сервисов и последовательность решения научных задач, а также технические соглашения разработки сервисов. Программные комплексы, рассмотренные в рамках этой работы:

- Программный комплекс (ПК) ИНТЭК-SAW предназначен для оценки влияния внешних угроз на ТЭК страны и регионов, а также определения мероприятий по их предупреждению или ликвидации последствий их наступления в контексте принятия стратегических решений. Основные функциональные возможности компонентов ПК ИНТЭК-SAW представлены на рис. 2, более подробно в [Массель, 2021].

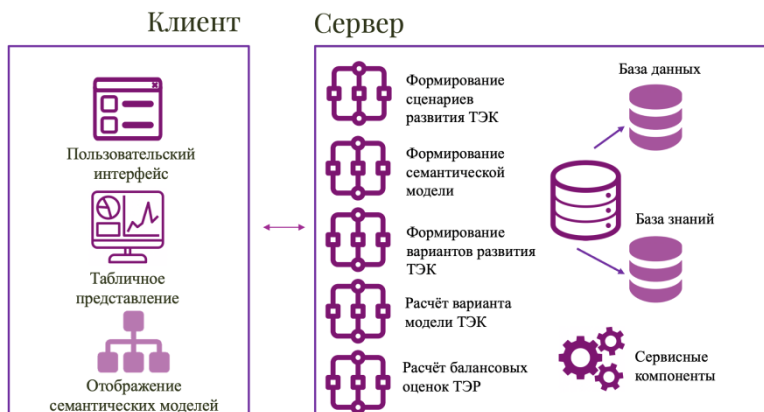


Рис. 2. Базовые компоненты и задачи ПК ИНТЭК-SAW

- Информационно-вычислительная система (ИВС) WICS разработана для проведения комплексных исследований по оценке загрязнения окружающей среды объектами энергетики и позволяет оценивать объём выбросов загрязняющих веществ (ЗВ) от объектов энергетики, рассчитывать их распространение в атмосферном воздухе с учётом метеорологических условий, а также оценивать экономический ущерб, нанесённый окружающей среде выбросами ЗВ от объектов энергетики. Основные функциональные возможности компонентов ИВС WICS представлены на рис. 3, подробнее в [Kuzmin, 2023].

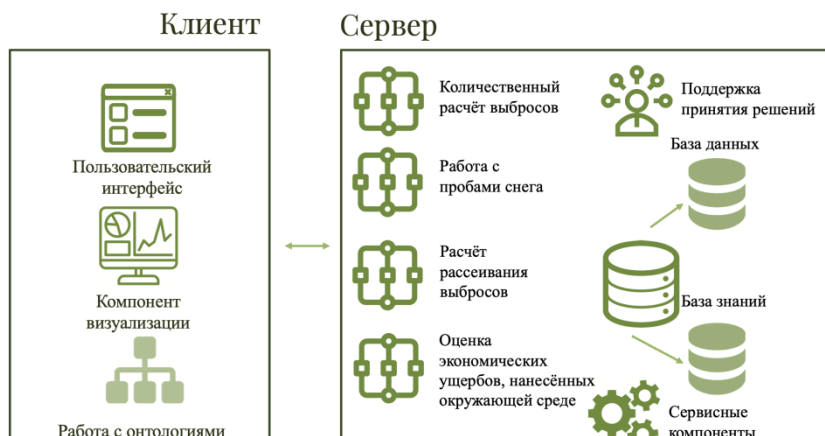


Рис. 3. Базовые компоненты и задачи ИBC WICS

Представленные решения имеют клиент-серверную архитектуру, собственные вычислительные компоненты и локальные базы данных и знаний. Для обеспечения единообразного хранения данных и знаний в формализованном виде с использованием онтологий, предоставить возможность работы с онтологиями, обеспечить пополнение и актуализацию данных и знаний из внешних источников, предоставить интерфейс с возможностью поиска знаний и данных предлагается разработка онтологического портала.

Архитектура в обобщенном виде предлагаемого онтологического портала представлена на рис. 4.



Рис. 4. Базовые компоненты и задачи онтологического портала

2. Интеграция компонентов экосистемы знаний

ПК ИНТЭК-SAW позволяет решать классическую задачу линейного программирования при помощи модели оптимизации топливно-энергетических ресурсов (ТЭР) по федеральным округам с выделением блоков добычи, транспорта и производства ТЭР по отраслям энергетики (газ, уголь, мазут, электроэнергетика, теплоэнергетика, нефть). Модель описывается системой линейных уравнений и целевой функцией минимизации суммы приведенных затрат и потерь от дефицита топливно-энергетических ресурсов среди потребителей. Исследования выполняемые при помощи ПК заключаются в формировании сценария развития ТЭК как в нормальных условиях, так и в условиях реализации угроз энергетической безопасности. Сценарий представляет упорядоченную последовательность вариантов модели ТЭК. Для каждого варианта оценивается наличие дефицитов ТЭР и стоимость реализации. Также одним из результатов расчётов ПК является объем потребленного ТЭР каждой энергетической технологией или потребление неэнергетических отраслей по федеральным округам. Превышение предельно допустимых концентраций (ПДК) ЗВ напрямую не учитываются и не рассчитываются в модели, однако, могут быть учтены при формировании сценариев.

ИВС WICS реализует интеграцию методик расчётов выбросов и распространения загрязнений и позволяет проводить расчеты выбросов и распространения загрязнений без привязки к федеральным округам. Результаты вычислительных экспериментов, выполненных с применением ИВС WICS, могут быть использованы при создании и эксплуатации объектов энергетики, и оценки влияния этих решений на финансово-экономические и инвестиционные показатели, региональную экономику и экономику природопользования.

Для оценки вариантов развития ТЭК с учётом расчётов выбросов и распространения загрязнений от объектов энергетики предлагается интегрировать ПК и ИВС. Результаты расчетов, полученные в ПК ИНТЭК-SAW, а именно объемы потребленного ТЭР энергетической технологией или неэнергетическими отраслями передаются в ИВС WICS для оценки загрязнения окружающей среды и, в зависимости от цели исследования, в случае превышения ПДК по результатам вычислительного эксперимента, возвращаются в качестве входных параметров обратно в ПК. Схема такого взаимодействия представлена на рис. 5.

Для подготовки данных в ИВС WICS используются технические характеристики объектов энергетики, характеристики сжигаемого топлива, метеоданные, сведения о рельефе, нормативы ПДК, нормативы предельно допустимых выбросов. ПК ИНТЭК-SAW рассчитывает объемы потребленного ТЭР и содержит информацию о видах топлив.

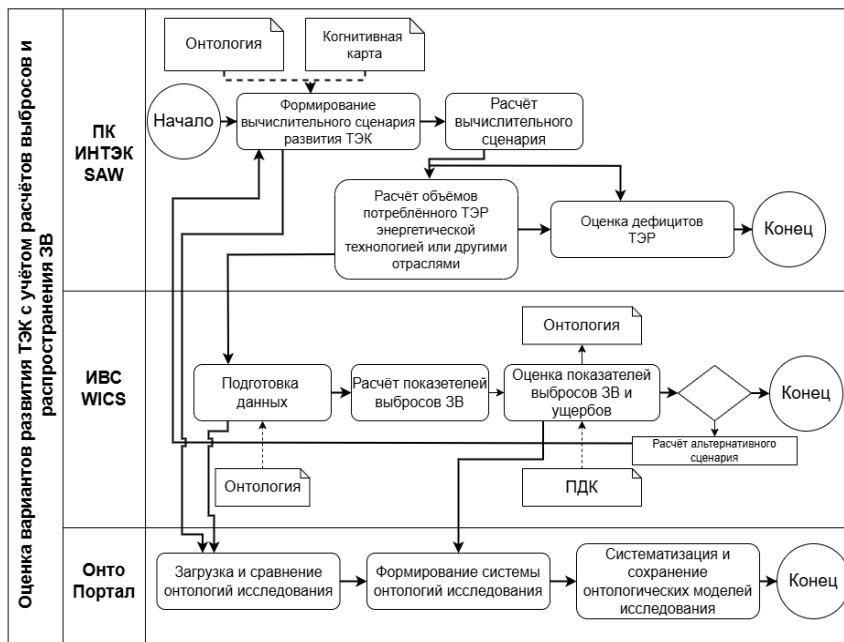


Рис. 5. Диаграмма интеграции ПК ИНТЭК-SAW, IBC WICS и онтологического портала для оценки вариантов развития ТЭК с учётом расчётов выбросов и распространения загрязнений от объектов энергетики

Для интеграции и обмена знаниями между ПК ИНТЭК-SAW и IBC WICS предлагается использовать язык управления знаниями (Knowledge Management Language – KML) [Massel et al., 2019]. Онтология KML представлена на рис. 6. KML является модификацией и развитием языка ситуационного управления CML (Contingency Management Language). KML позволяет выполнять формализацию семантических моделей для их хранения в базе знаний, преобразования между моделями (например, онтологий в когнитивные карты), а также интеграцию с другими программными комплексами. Для интеграции ПК ИНТЭК-SAW и IBC WICS предлагается использовать когнитивные карты и онтологии, описывающие объекты энергетики, для которых проводится расчёт в ИНТЭК-SAW с требуемыми параметрами, и на основе которых в IBC WICS проводятся расчёты количественных показателей выбросов и их рассеивание в атмосферном воздухе.

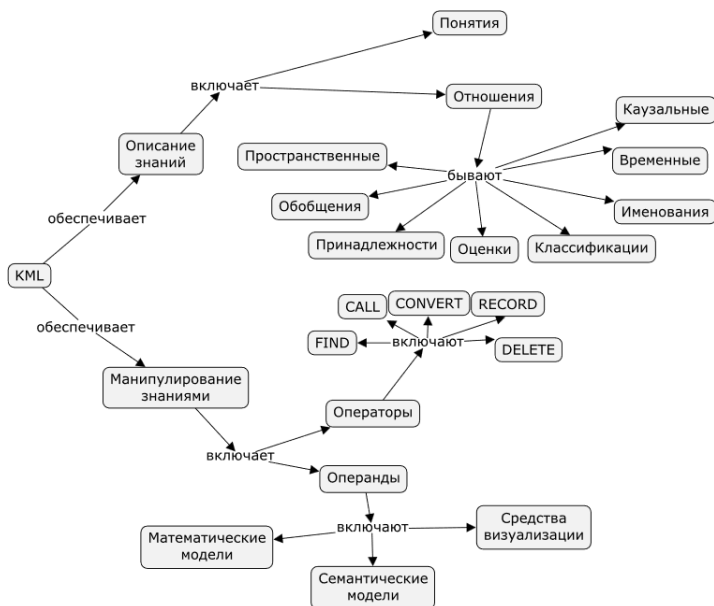


Рис. 6. Онтология языка управления знаниями KML

Для интеграции ПК ИНТЭК-SAW и ИВС WICS предлагается использовать когнитивные карты и онтологии, описывающие объекты энергетики, для которых проводится расчёт в ИНТЭК-SAW с требуемыми параметрами, и на основе которых в ИВС WICS проводятся расчёты количественных показателей выбросов и их рассеивание в атмосферном воздухе. В настоящий момент, согласован набор параметров для проведения расчётов и ведётся работа над программной реализацией. Аналогичным образом, с использованием KML, будет выполнена интеграция ИВС WICS и Онтологического портала. KML планируется интегрировать в представленные программные продукты в виде библиотеки, которая будет автоматически преобразовывать данные для взаимодействия программных продуктов между собой, а также обеспечивать доступ к знаниям, формализованным в виде онтологий и хранящихся в том числе в Онтологическом портале.

Представленные программные комплексы предоставляют возможность формализовать проводимые исследования с помощью семантических моделей: формирование сценария развития ТЭК с использованием когнитивной карты и использование системы «лёгких» онтологий для оценки загрязнения окружающей среды объектами энергетики. Проектирование онтологического портала для работы с семантическими моделями позво-

лит интегрировать сервисы на уровне работы со знаниями о предметных областях, формирования наборов необходимых данных для выполнения различных задач исследования.

Заключение

Интеграция специализированных программных комплексов в рамках архитектуры цифровой платформы экосистемы знаний направлена на решение междисциплинарных задач энергетического сектора. Подход к построению архитектуры цифровой платформы экосистемы знаний, включающей математические модели, базы данных, интеллектуальные подсистемы и сервисы визуализации, позволяет преодолеть ограничения изолированного использования специализированных решений. Построение различных сценариев совместного использования программных решений обладает потенциалом для адаптации в других междисциплинарных областях, где требуется координация разнородных вычислительных ресурсов и стандартизация данных. В настоящий момент, Цифровая платформа Экосистемы знаний в энергетике находится в стадии активной разработки, поэтому некоторые аспекты интеграции компонентов могли быть не раскрыты в рамках статьи.

Список литературы

- [Макаров, 2011] Макаров А.А., Веселов Ф.В., Елисеева О.А., Кулагин В.А., Митрова Т.А., Филиппов С.П. SCANNER – Модельно-информационный комплекс. – М.: ИНЭИ РАН, 2011.
- [Массель, 2021] Массель А.Г., Мамедов Т.Г., Пяткова Н.И. Технология вычислительного эксперимента в исследованиях работы энергетических отраслей при реализации угроз энергетической безопасности // Информационные и математические технологии в науке и управлении. – 2021. – No. 3(23). – С. 62-73.
- [Массель, 2023] Массель Л.В. Экосистема знаний как развитие и специализация цифровой экосистемы // Труды Межд. научно-техн. конгресса «Интеллектуальные системы и информационные технологии–2023». Научн. изд. в 2-х т. Т. 2. Таганрог: Издатель Ступин С.А., 2023. – С. 155-164.
- [Решение Высшего Евразийского экономического совета, 2017] Решение Высшего Евразийского экономического совета от 11 октября 2017 г. N 12 «Об основных направлениях реализации цифровой повестки Евразийского экономического союза до 2025 года» [Электронный ресурс]. – URL: <https://www.alta.ru/tamdoc/17vr0012/> (дата обращения: 20.05.2025).
- [Россети Сибирь разработали и внедрили систему управления знаниями, 2022] Россети Сибирь разработали и внедрили систему управления знаниями | Новости энергетики «ЭЛЕКТРОЭНЕРГИЯ. Передача и распределение» [Электронный ресурс]. – URL: <https://eepir.ru/new/rosseti-sibir-razrabotali-i-nbsp-vnedrili-sistemu-upravleniya-znaniyami/> (дата обращения: 12.08.2025).
- [Управлять знаниями, 2021] Управлять знаниями [Электронный ресурс]. – URL: <https://atomvestnik.ru/2021/10/26/upravljat-znaniyami/> (дата обращения: 12.08.2025).

- [EPA, 2004] AERMOD: description of model formulation. [Электронный ресурс]. – URL: <https://nepis.epa.gov/Exe/ZyPDF.cgi/P1009OXW.PDF?Dockey=P1009OXW.PDF> (дата обращения: 05.08.2025).
- [European Environment Agency, 2023] EMEP/EEA air pollutant emission inventory guidebook – European Environment Agency, 2023 [Электронный ресурс]. – URL: <https://www.eea.europa.eu/themes/air/air-pollution-sources-1/emep-eea-air-pollutant-emission-inventory-guidebook> (дата обращения: 05.08.2025).
- [Gabriel, 2001] Gabriel S.A., Kydes A.S., Whitman P. The National Energy Modeling System: A Large-Scale Energy-Economic Equilibrium Model [Электронный ресурс]. – URL: <https://doi.org/10.1287/opre.49.1.14.11195> (дата обращения: 05.08.2025).
- [Kuzmin, 2023] Kuzmin V.R., Vorozhtsova T.N., Massel L.V. Design and Development of Information and Computational System for Energy Facilities' Impact Assessment on Environment // Eng. Proc. – 2023. – 33, 21.
- [Massel et al., 2019] Massel L., Kuzmin V. Knowledge management language in the information and analytical system for impact assessment of the energy on the geoecology. Advances in Intelligent Systems Research T. 166 // 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS), 28-29 May 2019. – Ufa, Russian Federation. – P. 321-325.
- [Muller et al., 2012] Müller Johannes & Krchnavi Jaren & Stocker Alexander. Global knowledge exchange using References+ within the Siemens intranet. In: Innovation and Transformation Through Knowledge Management. Eds: Serventi, Evie. 2012.
- [Shrivastava, 1998] Shrivastava P. Knowledge Ecology: Knowledge Ecosystems for Business Education and Training [Электронный ресурс]. – URL: <https://web.archive.org/web/20170825081451/http://www.facstaff.bucknell.edu/shrivast/KnowledgeEcology.html> (дата обращения: 20.05.2025).
- [Stein, 2015] Stein A.F., Draxler R.R., Rolph G.D. [et al.] NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System // Bulletin of the American. – 2015. – Vol. 96, Is. 12. – P. 2059-2077.
- [Top 10: Energy Management Platforms, 2025] Top 10: Energy Management Platforms|Sustainability Magazine. [Электронный ресурс]. – URL: <https://sustainabilitymag.com/top10/top-10-energy-management-platforms>.

УДК 004.89

doi: 10.15622/rcai.2025.085

ТЕНЗОРНОЕ ПРЕДСТАВЛЕНИЕ НЕОДНОРОДНЫХ СЕМАНТИЧЕСКИХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧИ СЛИЯНИЯ БАЗ ЗНАНИЙ

А.И. Пальчевский (*apalchev@gmail.com*)

А.И. Молодченков (*aim@tesyan.ru*)

Федеральный исследовательский центр
«Информатика и управление» РАН, Москва

В работе рассмотрен способ преобразования баз знаний, построенный на основе неоднородных семантических сетей, в семейство тензорных структур. Введена метрика оценки близости узлов базы знаний. Преобразование базы знаний в семейство тензоров является одним из этапов разработки алгоритма автоматизации слияния баз знаний. Предложенный подход может быть применен и к онтологиям.

Ключевые слова: тензоры, неоднородные семантические сети, машинное обучение, КИИ-2025.

Введение

Задача слияния графовых структур, к которым относятся семантические и неоднородные семантические сети, является не новой, но актуальной в настоящее время. Так, например, когда несколько экспертов работают над одной базой знаний или необходимо объединить несколько онтологий в одну.

Существует множество подходов к решению данной задачи, варьирующихся от простых эвристических правил до сложных методов машинного обучения. Одним из перспективных направлений является использование графовых нейронных сетей (GNN) [Li, 2023], позволяющих обучать модели, непосредственно работающие с графовой структурой данных. Для эффективного обучения GNN необходимо разработать способы представления графов в виде векторных или тензорных эмбедингов. Среди существующих методов генерации таких эмбедингов можно выделить несколько ключевых направлений:

1. CP-разложение (CANDECOMP/PARAFAC) [Faber, 2003], [Yang, 2021], [Zhoubao, 2003], [Geña Hahn, 1997] представляет собой тензорное разложение высокого порядка, позволяющее декомпозировать граф на латентные компоненты. CP-разложение обладает относительно простой реализацией и интерпретируемостью, что делает его привлекательным для анализа графов с простой структурой. Однако, его эффективность может снижаться при работе со сложными графовыми структурами, содержащими большое количество связей и иерархических отношений.

2. Knowledge-Enriched Tensor Factorization [Padiá, 2019] представляет собой расширение методов тензорного разложения, позволяющее интегрировать внешние знания, такие как метаданные и семантическая информация, в процесс разложения. Интеграция внешних знаний позволяет улучшить качество получаемых эмбедингов и повысить точность анализа графов. Однако, эффективность этого подхода напрямую зависит от качества и доступности внешних знаний, а также от сложности их интеграции в тензорную модель.

3. Message Passing (MP) [Bordes, 2013] является основой многих современных GNN и позволяет узлам графа обмениваться информацией с соседними узлами, учитывая как локальные, так и глобальные зависимости в графе. Алгоритмы, основанные на этом подходе, часто превосходят стандартные методы матричного разложения и алгоритмы семейства TransX по качеству, особенно при наличии сложных зависимостей между узлами и небольшом объеме данных для обучения. Однако, MP требует значительных вычислительных ресурсов и чувствителен к качеству интеграции знаний, требуя преобразования текстовой или другой неструктурированной информации в матричный вид.

Несмотря на существенный прогресс в разработке методов слияния графовых структур, большинство существующих подходов ориентированы на работу с однородными графами, где все узлы и ребра имеют одинаковый тип и свойства. Однако, есть задачи, где часто встречаются неоднородные графы, содержащие узлы и ребра различных типов, характеризующиеся разными свойствами и семантикой. В данной работе представлен новый подход к преобразованию графовых структур в семейство тензоров, который подходит для работы с разнородной структурой узлов и свойств. Кроме этого, в работе представлена метрика оценки близости таких узлов.

1. Преобразование семантической сети в семейство тензоров

Неоднородная семантическая сеть – семейство графов, имеющих общее множество вершин; вершинам сопоставлены объекты моделируемой действительности, ребрам элементы некоторых бинарных отношений на множестве вершин; им же сопоставлены процедуры, предназначенные для

проверки корректности сети и порождения различного рода гипотез, повышающих эффективность процесса построения сети [Осипов, 2015]. В данной главе представлен метод преобразования неоднородных семантических сетей (НСС) в структуру, пригодную для обработки методами тензорной алгебры, обучения графовых, и не только, нейронных сетей и разработки алгоритмов слияния графовых структур. Ключевой особенностью подхода является представление НСС в виде семейства разделенных тензоров, позволяющих эффективно кодировать разнородную информацию, содержащуюся в структуре графа. В отличие от традиционных методов, использующих единый тензор для представления графа, наш подход разделяет информацию о связях, атрибутах узлов, текстовых дескрипторах и типах узлов на отдельные тензорные компоненты, что обеспечивает большую гибкость и выразительность представления.

1.1. Структура тензоров

Предложенная структура разделенных тензоров включает в себя три основных компонента:

1. Тензор связей – $A \in \mathbb{R}^{N \times N}$ – кодирует направленные связи между узлами в графе. N представляет собой общее количество узлов в сети. Элемент A_{ij} представляет собой значение, ассоциированное с направленной связью типа k от узла i к узлу j .

Например, $A_{3,5} = 0$ означает наличие связи типа "всегда наблюдается" от узла с идентификатором 3 к узлу 5. Соответственно связь $A_{5,3} = 1$ обозначает, обратную связь от узла с идентификатором 5 к узлу 3 с типом связи «Может наблюдаться».

2. Тензор узлов – $X \in \mathbb{R}^{N \times F}$. Этот тензор кодирует атрибуты, характеризующие каждый узел в графе. N представляет количество узлов, а F – количество постоянных свойств, описывающих каждый узел. Для нашего примера $F = 3$ (название, тип узла, описание). Тензор узлов может включать в себя как числовые, так и категориальные свойства. Числовые свойства нормализуются для приведения к единому масштабу.

Категориальные свойства кодируются с использованием методов one-hot или multi-hot кодирования. Например, $X_0 = [0.3, 1, 0, 1]$ может обозначать узел с численным атрибутом, имеющим значение 0.3, и дополнительными категориальными свойствами, представленными вектором $[1, 0, 1]$. Дополнительно, в нашем подходе мы используем расширенный набор признаков, включая статистические показатели, полученные из окрестности узла в графе, такие как центральность по степени, посредничеству и близости, что позволяет более полно охарактеризовать роль и положение узла в сети.

3. Тензор названий узлов – $E \in \mathbb{R}^{N \times D}$. Этот тензор кодирует семантическое значение названий узлов. N представляет количество узлов, а D – размерность векторного представления (эмбединга). Для генерации

эмбедингов используются предварительно обученные NLP-модели, такие как BERT, Word2Vec или GloVe. Использование NLP-моделей позволяет учитывать контекст и семантические связи между названиями узлов, что повышает качество представления графа. Мы также рассматриваем возможность использования трансформерных архитектур для генерации более контекстуально-зависимых эмбедингов.

1.2. Метод преобразования НСС в семейство тензоров

Рассмотрим пример структуры подсети НСС (рис. 1).

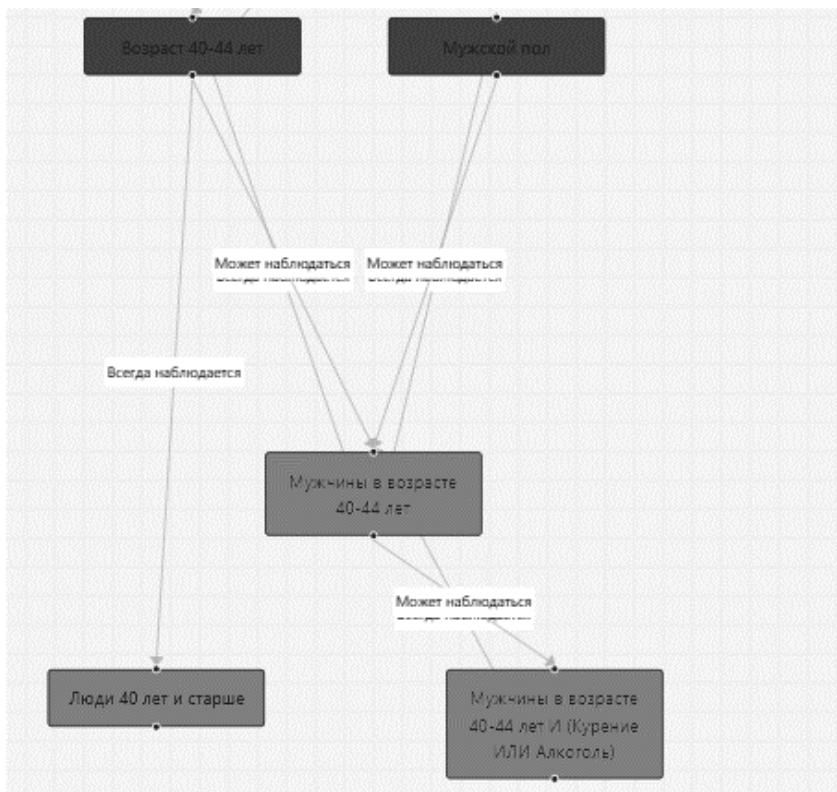


Рис. 1. Пример подсети. С тремя типами узлов

Метод состоит из следующих этапов.

На первом этапе создается матрица узлов. Название узла векторизуется, а затем связывается с соответствующим полем в матрице. Записываются значения типа узла и векторизованное описание.

Например, (табл. 1).

Таблица 1

Индекс	Название	Тип узла	Векторное представление названия	Векторное представление описания
0	Мужчины в возрасте 40-44 лет	0	[векторизованный текст]	[векторизованный текст]
1	Возраст 40-44 лет	1	[векторизованный текст]	[векторизованный текст]

На втором этапе преобразование динамических свойств узла в тензор свойств. Например,

$$\begin{pmatrix} \text{Пол} & \text{Возраст} & \text{Артериальное давление} \\ [M, F] & [18, \dots, 80] & [70, \dots, 190] \end{pmatrix}. \quad (1.1)$$

На третьем этапе строится матрица связей узлов с их динамическими свойствами. Эта матрица связывает индексы узлов и динамических свойств. В качестве значений выступают области значений этих свойств для узла. В примере выше свойства Пол, Возраст, Артериальное давление имеют индексы 0, 1, 2 соответственно. У нас имеются узлы «Мужской пол», «Возраст 40-44» которые имеют индексы в своей матрице 0, 1. Тогда для связи узла «Мужской пол» со свойством будет в матрице будет записана следующая строка

$$\begin{pmatrix} \text{Индекс узла} & \text{Индекс свойства} & \text{Область значений} \\ 0 & 0 & [M] \end{pmatrix}. \quad (1.2)$$

На четвертом этапе строится матрица связей, где указаны типы отношений между узлами. Например, создаем матрицу α размерностью $[N \times N]$, где N – это номер узла.

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 2 & 2 & 0 \end{bmatrix}, \quad (1.3)$$

где

а) Узлы:

- «Возраст 40-44 лет» – индекс 0;
- «Мужской пол» – индекс 1;
- «Мужчины в возрасте 40-44» – индекс 2;

б) Связи и их типы кодируются значениями в промежутке от 0 до 4 включительно, где 1 – Всегда наблюдается, 2 – Может наблюдаться, 3 – Может отсутствовать, 4 – всегда отсутствует. Соответственно матрица читается как: От узла 0 связь «Всегда наблюдается» к узлу 2 и от узла 3 связь «Может наблюдаться» к узлу 0.

2. Метрики близости для оценки близости узлов с гетерогенными свойствами

В задачах слияния графовых структур, особенно в контексте неоднородных семантических сетей (НСС), критически важным является точная оценка близости между узлами, характеризующимися гетерогенными свойствами. Традиционные метрики, такие как евклидово расстояние или косинусное сходство, часто оказываются неэффективными при сравнении узлов с разнородными атрибутами. В данной главе мы представляем взвешенную комбинированную метрику, специально предназначенную для оценки близости узлов в НСС, учитывающую гетерогенность свойств и структурные особенности графа.

Предлагаемый подход состоит в объединении нескольких специализированных метрик, каждая из которых предназначена для оценки сходства в определенном векторном пространстве, с последующим взвешенным агрегированием результатов. Это позволяет эффективно обрабатывать разнородные атрибуты узлов, такие как числовые значения, категориальные признаки, текстовые описания и интервальные данные, а также учитывать структурную информацию о графе.

Общая формула для оценки близости между узлами N и K выглядит следующим образом:

$$Sim_{total}(N, K) = \left(\sum_{i=1}^n \beta_i \times Sim_{K_i}(K_i, K'_i) \right) \times (Sim_{cos}(NodeName) \times X(N, K)), \quad (2.1)$$

где

1. Sim_{name} – косинусное сходство векторных представлений имен свойств.
2. Sim_{value} – метрика близости гетерогенных скалярных значений.
3. Sim_{type} – категориальная метрика близости для свойств типа интервальная, качественная и т.д.
4. $Sim_{interval}$ – комбинированная метрика (перекрытие + расстояние).
5. $\beta_i \in [0,1], \sum \beta_i = 1$ – адаптивные веса для корректировки важности признака.

Структурные свойства узлов:

1. $Sim_{cos}(NodeName)$ – Косинусное сходство векторных представлений имен узлов.
2. X – функция, которая используется как бинарный множитель, зависящий от типа узла где 1 это совпадение типа узла и 0 для разных типов узлов.

Рассмотрим более подробно каждый компонент данной формулы.

Для оценки близости названий можно использовать косинусное сходство:

$$Sim_{Name/NodeName} = \cos(\theta) = \frac{A \times B}{||A|| ||B||} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}. \quad (2.2)$$

Для интервальных значений можно применить комбинированную метрику, которая будет учитывать, как перекрытие, так и расстояние интервалов:

$$Sim_{interval}(A, B) = \alpha \times \frac{|A \cap B|}{|A \cup B|} + (1 - \alpha) \times \left(1 - \frac{|mid(A) - mid(B)|}{max_range}\right), \quad (2.3)$$

где $\alpha \in [0, 1]$ – вес перекрытия.

Для работы со скалярными значениями нужно ввести универсальную метрику, учитывающую, что скалярное значение может быть нескольких видов, например, непрерывная величина, категориальное значение или порядковая величина.

Формально метрику можно представить в виде:

$$Sim_{value}(X, Y) = \omega \times (Sim_{interval}(X, Y)) + (1 - \omega) \times \delta_{XY}. \quad (2.4)$$

Обоснование создания новой метрики для разных типов узлов заключается в том, что традиционные метрики не учитывают динамическую размерность и гетерогенность признаков. Поэтому предложено решение: каждое векторное пространство признаков K_i – обрабатывается отдельно специализированной метрикой, уменьшая размерность итогового пространства к 2 размерностям. Веса β_i отражают относительную важность одной конкретной метрики признаков в домене. Например для признаков из разных векторных пространств $\beta_{name} > \beta_{value}$.

Оптимизация β_i проводится на этапе экспериментов. Веса изначально инициализируются как $\frac{1}{n}$. При подборе значений весов можно использовать L2 регуляризацию для предотвращения перекоса в пользу одного признака. Также использование относительных весов дает устойчивость к неполноте данных. Например: если некоторый K_i отсутствует, то соответствующий ему β_i будет равен 0.

Структурные модуляторы мы можем использовать для корректировки общей относительной близости узла. $Sim_{cos}(NodeName)$ – использует контекстные эмбединги (Bert или Sentence-BERT), учитывающие близость имён. Это позволяет нам корректировать итоговое значение близости двух узлов при совпадении их типов и близости динамических свойств.

Функция $X(N, K)$ – вводит штраф за несовпадение типов. Мы берем упрощённую модель, где при разных типах мы можем сказать, что даже при совпадении свойств узлы разные.

Все Sim_{K_i} , нормированы и $Sim_{total} \in [0,1]$, где 1 – это полное совпадение узлов. Базовая настройка порогового значения $Sim_{total} > 0.8$. При достижении порогового значения мы можем сказать, что узлы N и K являются одним узлом и мы можем провести операцию слияния.

Заключение

В настоящей работе был представлен подход к решению задачи преобразования графовых структур в семейство тензоров, ориентированный на работу с неоднородной структурой графов и семантических сетей.

Преобразование НСС в семейство тензоров позволило эффективно закодировать разнородную информацию, содержащуюся в базах знаний, путем разделения ее на отдельные тензорные компоненты, соответствующие связям, атрибутам узлов, текстовым дескрипторам и типам узлов.

Взвешенная комбинированная метрика, разработанная для оценки близости узлов, учитывает гетерогенность свойств и структурные особенности графа.

В дальнейших исследованиях планируется разработка алгоритма результатов для создания итеративного алгоритма слияния, который будет автоматически определять наиболее близкие узлы на основе разработанной метрики и выполнять их слияние с учетом структуры графа, а также планируется проведение экспериментальных исследований.

Список литературы

- [Осипов, 2015] Осипов Г.С. Методы искусственного интеллекта. – 2-е изд. – М.: ФИЗМАТЛИТ, 2015. – 296 с.
- [Abadal, 2021] Abadal, Sergi, et al. Computing graph neural networks: A survey from algorithms to accelerators // ACM Computing Surveys (CSUR). – 2021. – 54.9. – P. 1-38. – doi:10.1145/3477141.
- [Bordes, 2013] Bordes A., Usunier N., Garcia-Durán A., Weston J., & Yakhnenko O. Translating Embeddings for Modeling Multi-relational Data // Advances in Neural Information Processing Systems (NeurIPS). – Dec. 2013. – Vol. 26. – P. 2787-95. – <https://proceedings.neurips.cc/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf>.
- [Li, 2023] Li C., Wu S., Chen T., Wang R. and Cao J. Knowledge Fusion Algorithm Based on Entity Relation Mining and Graph Neural Network // 5th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qiangdao, China, 2023. – P. 413-417. – doi: 10.1109/ICFTIC59930.2023.10456136.
- [Faber, 2003] Faber, Nicolaas Klaas M., Rasmus Bro, and Philip K. Hopke. Recent developments in CANDECOMP/PARAFAC algorithms: a critical review // Chemometrics and Intelligent Laboratory Systems. – 2003. – 65.1. – P. 119-137. – doi: 10.1016/S0169-7439(02)00089-8.

- [**Geña Hahn, 1997**] Geña Hahn, Gert Sabidussi. Graph symmetry: algebraic methods and applications. – Springer, 1997. – Vol. 497. – P. 116. – (NATO Advanced Science Institutes Series). – ISBN 978-0-7923-4668-5.
- [**Padia, 2019**] Padia, Ankur, et al. Knowledge graph fact prediction via knowledge-enriched tensor factorization // Journal of Web Semantics. – 2019. – 59. – P. 100497. – doi: <https://doi.org/10.1016/j.websem.2019.01.004>.
- [**Yang, 2021**] Yang Han, and Junfei Liu. Knowledge graph representation learning as groupoid: unifying TransE, RotatE, QuatE, ComplEx // Proceedings of the 30th ACM international conference on information & knowledge management. – P. 2311-2320. – doi: 10.1145/3459637.3482442.
- [**Zhoubao, 2003**] Zhoubao Sun, Xiaodong Zhang, Haoyuan Li, Yan Xiao, Haifeng Guo. Recommender Systems Based on Tensor Decomposition // Tech Science Press Computers, Materials & Continua January 2020. – 66(1). – P. 621-630. – DOI: 10.32604/cmc.2020.012593.

УДК 004.89

doi: 10.15622/rcai.2025.086

ОНТОЛОГИЧЕСКАЯ ПРОГРАММНАЯ СРЕДА ИНТЕЛЛЕКТУАЛЬНОГО ОБУЧЕНИЯ НА ОСНОВЕ ИНТЕГРИРОВАННЫХ ЭКСПЕРТНЫХ СИСТЕМ: МЕТОДЫ ПОСТРОЕНИЯ НА ПРИНЦИПАХ СИТУАЦИОННОГО УПРАВЛЕНИЯ

Г.В. Рыбина (*GVRybina@yandex.ru*)

А.А. Григорьев (*grigandal625@gmail.com*)

Национальный исследовательский ядерный университет «МИФИ»,
Москва

Описываются результаты исследований, связанных с построением адаптивной онтологической программной среды (АОЕ) интеллектуального обучения на основе использования обучающих ИЭС и веб-ИЭС, разработанных на базе задачно-ориентированной методологии и средств интеллектуальной программной среды инструментального комплекса АТ-ТЕХНОЛОГИЯ. Архитектура АОЕ разработана в виде гибридной интеллектуальной системы управления (ИСУ) процессами функционирования обучающих ИЭС и веб-ИЭС с использованием принципов ситуационного управления сложными организационно-техническими системами. Рассматриваются предложенные модели и методы реализации ситуационного подхода к решению задачи управления АОЕ с использованием обобщенной онтологии представления знаний о направлении/специальности подготовки и базовых средств вывода.

Ключевые слова: интеллектуальные обучающие системы, интеллектуальное обучение, интегрированные экспертные системы, ИЭС, обучающие ИЭС, веб-ИЭС, задачно-ориентированная методология, комплекс АТ-ТЕХНОЛОГИЯ, адаптивная онтологическая среда, обобщенная модель онтологии, ситуационное управление, интеллектуальная система управления, ИСУ.

Введение

Методы и технологии интеллектуальных обучающих систем (ИОС), ведущих свое начало от пионерских работ Дж. Карбонелла [Carbonell, 1970] и П.Л. Брусиловского [Брусиловский, 1990], сегодня активно вос-

требованы в системе вузовского образования, поскольку обеспечивают возможность автоматизированной поддержки базовых для ИОС моделей *обучаемых* и разнообразных моделей *обучения*.

Быстрое развитие веб-технологий и облачных платформ позволило значительно расширить возможности средств *интеллектуального обучения* [Рыбина, 2023b] на основе разработки и использования ИОС и веб-ИОС, обладающих разнообразной архитектурой и способных интегрировать методы и средства ИИ с методами и подходами из других областей на основе концептуальной и логической взаимосвязи процессов *индивидуализации, интеллектуализации и веб-ориентации*. Классификация и примеры реализации ИОС с различной архитектурной типологией, включая инструментальные средства (ИС) для поддержки их разработки приводятся в [Рыбина, 2014], [Рыбина, 2023b], а также в работах [Алещенко и др., 2016], [Bonner et al., 2015], [Gribova et al., 2018], [Nye, 2015], [Rahman et al., 2016], [Сычев и др., 2025] и др.

В последние годы возросла роль онтологического подхода в ИОС и веб-ИОС для построения онтологий различного назначения [Фролов, 2024], [Гаврилова, 2016], [Рыбина, 2023a], [Rybina et al., 2022], [Rybina et al., 2023], при семантической интеграции различных веб-ИОС в рамках общей проблемной области (ПрО) [Sosnovsky et al., 2008] и т.д. Однако, при этом в ИОС отсутствуют комплексные исследования и разработки, направленные на создание системного подхода к автоматизированному построению и использованию доступного онтологического пространства знаний и умений обучаемых, формируемого на протяжении всего конкретного образовательного цикла.

Для решения этой актуальной проблемы достаточно перспективной научной и технологической базой оказался комплекс моделей и методов автоматизированного построения *обучающих* интегрированных экспертных систем (ИЭС) и веб-ИЭС (далее ИЭС/веб-ИЭС), представляющих собой полнофункциональные ИОС, и разработка которых осуществляется на основе задачно-ориентированной методологии построения ИЭС [Рыбина, 2008] и поддерживающего методологию инструментария класса Workbench в виде интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ [Рыбина, 2014], [Rybina et al., 2019].

Многолетний опыт (с 2008 г.) разработки и использования в учебном процессе НИЯУ МИФИ по направлениям подготовки «Прикладная математика и информатика» и «Программная инженерия» нескольких обучающихся ИЭС/веб-ИЭС привел к необходимости эффективного использования значительных ресурсов в виде хорошо апробированных взаимосвязанных моделей, методов и программных средств для решения типовых задач интеллектуального обучения на различных этапах образовательного процесса [Рыбина, 2014], [Рыбина, 2023b].

Поэтому актуальной стала задача создания единого автоматизированного *онтологического пространства знаний и умений* обучаемых [Рыбина, 2023b], [Rybina et al., 2022], [Rybina et al., 2023] как совокупности информационно-образовательных процессов и программных ресурсов по реализации типовых задач интеллектуального обучения на основе обучающих ИЭС/веб-ИЭС, используемых в течение полного цикла непрерывного обучения под управлением обобщенной онтологии и средств интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ, с целью автоматизированного построения компетентностно-ориентированных моделей выпускников по конкретным направлениям/специальностям подготовки за конкретный период.

В работе обсуждаются результаты исследований, связанных с разработкой методов и средств построения адаптивной онтологической среды интеллектуального обучения на основе обучающих ИЭС/веб-ИЭС.

1. Краткая характеристика обучающих интегрированных экспертных систем и процессов их функционирования

Детальное описание методов и средств разработки обучающих ИЭС/веб-ИЭС содержится в базовых работах [Рыбина, 2023b], [Рыбина, 2014], [Рыбина, 2008] и целом ряде публикаций разных лет, поэтому ниже приводится краткая характеристика только тех особенностей, которые наиболее важны в данной работе.

1. Обучающие ИЭС/веб-ИЭС обладают масштабируемой архитектурой, позволяющей расширять компонентную функциональность ИЭС до уровня ИОС за счет построения комплекса базовых моделей, что включает: компетентностно-ориентированную модель обучаемых, адаптивную модель обучения, модель прикладной онтологии курса/дисциплины, модель обобщенной онтологии представления знаний о направлении/специальности подготовки, расширенную модель объяснения.

2. Проходят постоянную апробацию в учебном процессе методы и средства реализации трех типовых задач интеллектуального обучения [Рыбина, 2023b], а именно: индивидуальное планирование методики изучения учебных курсов/дисциплин; интеллектуальный анализ решений учебных задач; интеллектуальная поддержка принятия решений.

3. Динамически поддерживаются: процессы веб-тестирования обучаемых при выявлении текущего уровня *знаний* обучаемых с использованием прикладных онтологий курсов/дисциплин и генетического алгоритма; веб-взаимодействие с компонентами выявления уровня *умений* обучаемых решать учебные НФ-задачи [Рыбина, 2023b], [Рыбина, 2014] и обучающими *воздействиями* типа «Тренинг с ЭС/ИЭС» [Rybina et al., 2024]

4. Обеспечивается мониторинг функционирования обучающих веб-ИЭС, включающий автоматизацию процессов, возникающих в ходе обучения и контроля знаний/умений обучаемых (с учетом требований информационной безопасности).

2. Общая структура и конкретизация информационно-образовательных и программных ресурсов интеллектуального обучения на основе интегрированных экспертных систем

На рис. 1 представлена общая структура [Рыбина, 2023а] основных информационно-образовательных процессов и программных ресурсов нескольких обучающих ИЭС/веб-ИЭС, задействованных для автоматизированного выявления уровня знаний и умений обучаемых при реализации представленных на рис. 1 типовых задач интеллектуального обучения, с учетом использования всех ресурсов и средств на всех и/или отдельных этапах полного образовательного цикла под управлением обобщенной онтологии представления знаний о конкретном направлении/специальности подготовки.

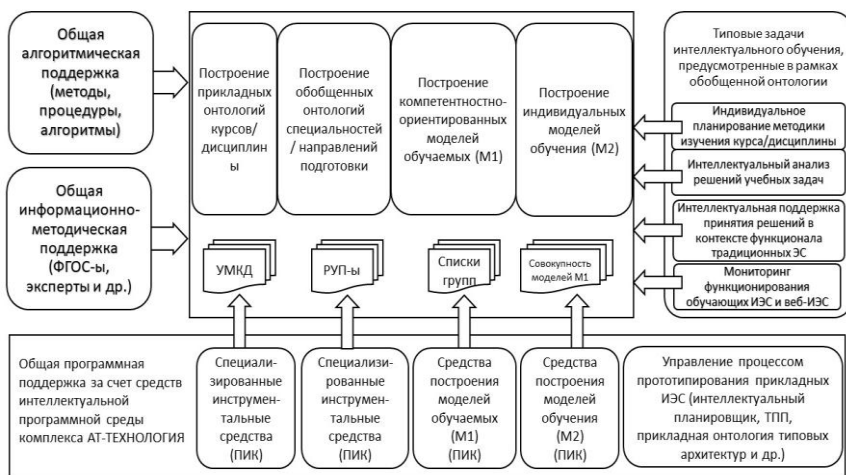


Рис. 1. Общая структура информационно-образовательных и программных ресурсов интеллектуального обучения с использованием обучающих ИЭС [Рыбина, 2023а]

В этом случае общая структура адаптивной онтологической среды интеллектуального обучения на основе обучающих ИЭС (далее АОЕ от англ. Adaptive Ontological Environment) конкретизируется в виде кортежа:

$$M_{AOE} = \langle G, T, D, M1^{kn}, M1^{sk}, M2, K, P, W, I, DB, IES, SP, DOC \rangle, \quad (1)$$

где $G = \langle O, F \rangle$ – модель обобщенной онтологии представления знаний о направлении/специальности подготовки, где $O = \bigcup_{i=1}^5 O_i$ – прикладные онтологии курса/дисциплины, $F = \{F_1, F_2, F_3, F_4\}$ – множество базовых операций (процедур) построения элементов онтологии G , реализованных в виде программных компонентов (F_1 – объединение онтологий, F_2 – выборка элементов онтологии, F_3 – отсечение элементов онтологии, F_4 – поиск элементов онтологии);

$O_i = \langle M_o, F_o \rangle$ – прикладная онтология курса/дисциплины, где M_o – базовая модель онтологии в виде семантической сети, $F_o = \{F_z, F_q, F_{am}, F_k, F_{kr}\}$ – множество операций построения элементов O_i (F_z – построение структуры курса/дисциплины, F_q – формирование вопросов к контролируемым элементам, F_{am} – реализация адаптивного метода репертуарных решеток, F_k – построение целевых компетенций, F_{kr} – построение связей между компетенциями и элементами курса.);

$T = \{T_j\}, j = 1..3$ – типовые задачи интеллектуального обучения, где T_1 – индивидуальное планирование методики изучения учебного курса/дисциплины, T_2 – интеллектуальный анализ решений учебных задач, T_3 – интеллектуальная поддержка принятия решений;

$D = \{D_n\}$ – курсы/дисциплины в соответствии с УМКД;

$M1^{kn} = \{M1_k^{kn}\}$ – модели обучаемых (знаний обучаемых) по конкретным курсам/дисциплинам, $k \in D$;

$M1^{sk} = \{M1_l^{sk}\}$ – модели обучаемых (умений обучаемых) по конкретным курсам/дисциплинам, $l \in D$;

$M2 = \{M2_m\}$ – модели обучения по конкретным курсам, $m \in D$;

$K = \{K_q\}$ – целевые профессиональные компетенции в соответствии с УМКД, $q \in D$, $C_q \in \{\text{ПК-2, ПК-5, ПК-8, ОПК-3, ОПСК-1}\}$;

$P = \{P_{M1}, P_{M2}, P_G, P_O\}$ – средства построения $M1, M2, G, O$, включенные в состав интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ;

$W = \{W_U, W_T\}$ – базовые средства вывода, включенные в состав интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ, где W_U – универсальный АТ-РЕШАТЕЛЬ, W_T – темпоральный решатель;

I – средства имитационного моделирования внешней среды, включенные в состав интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ;

$DB = \{DB_{M1}, DB_{M2}\}$ – единая база данных моделей обучаемых и моделей обучения (более 2500 моделей), содержащая информацию, накопленную в процессе функционирования обучающихся ИЭС по конкретным курсам/дисциплинам;

IES – динамическая ИЭС «Мониторинг информационной безопасности процессов функционирования обучающихся веб-ИЭС»;

$SP = \{SP_i\}, i = 1..7$ – совокупность параметров аналитико-статистической обработки результатов мониторинга функционирования обучающихся ИЭС;

DOC – совокупность отчетных документов (ведомости, отчеты и др.).

Таким образом, структура АОЕ представляет собой сложный *функционально-интегрированный* комплекс взаимодействующих логико-лингвистических, математических и информационно-программных средств автоматизации, предназначенных для мониторинга и управления процессами функционирования обучающихся ИЭС/веб-ИЭС. Это позволяет рассматривать *АОЕ* как отдельный тип архитектуры гибридной интеллектуальной системы управления (ИСУ), основываясь на базовых принципах организации современных ИСУ [Макаров и др., 2006].

3. Применение принципов ситуационного управления для построения архитектуры интеллектуальной системы управления функционированием обучающихся ИЭС

В соответствии с классической теорией автоматического управления под *управлением* всегда подразумевается организованное специальным образом взаимодействие объекта и устройства (системы) управления. Однако, нетрадиционность и слабая формализуемость *интеллектуального* объекта управления (ОУ) организационно-технического типа в виде обучающей ИЭС/веб-ИЭС, базирующейся на интеграции моделей ИИ и ИОС и технологиях, основанных на знаниях (включая НЕ-факторы знаний), делает неприемлемым использование классических методов оптимального управления сложными объектами.

Поэтому в работе используется подход, основанный на принципах ситуационного управления сложными техническими и организационными системами, предложенный Д.А. Поспеловым [Поспелов, 1986], в основе которого лежат семиотические и логико-лингвистические модели ИИ для представления знаний об ОУ и способах управления им. Идеи ситуационного управления [Клыков, 1974], [Поспелов, 1986], [Поспелов, 2020] развивались в нашей стране одновременно с развитием работ в области ИИ, взаимно обогащая как теорию, так и практические разработки интеллектуальных систем для различных ПрО и классов задач управления [Болотова, 2012], [Башлыков и др., 2025], [Цветков, 2023] и др.

Поскольку данная работа представляет собой современный виток развития ситуационного управления применительно к ИСУ, то для дальнейшего анализа метода ситуационного управления на рис. 2 приводится общая схема решения задачи правления сложными объектами в том виде, как она была представлена в [Поспелов, 1986]. Здесь введены основные определения: текущая ситуация на ОУ – Q_j , полная ситуация на ОУ – S_i ,

управляющее воздействие на ОУ (одношаговое решение) – U_k . Элементарный акт управления задается в виде $S_i; Q_j \rightarrow Q_l$, где Q_l – новая ситуация, возникшая после применения допустимого воздействия U_k .

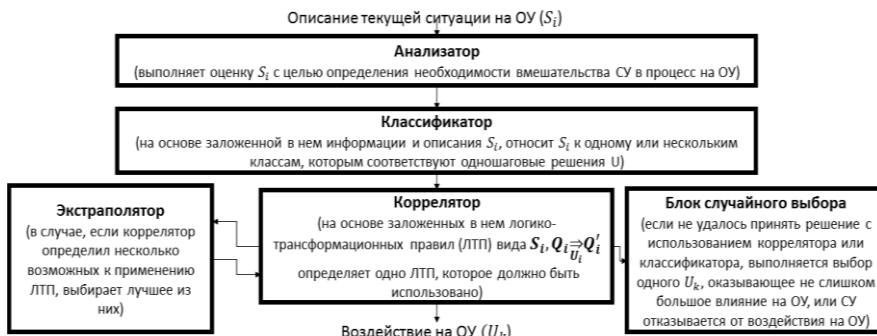


Рис. 2. Общая схема решения задачи управления [Поспелов, 1986]

Правила преобразования называются *логико-трансформационными* правилами (ЛТП) или *корреляционными*, а полный список ЛТП определяет возможности системы управления (СУ) воздействовать на процессы в ОУ. Таким образом, решение задачи управления – это последовательность выполнения основных процедур ситуационного управления, т.е. *анализа* текущей проблемной ситуации, ее *классификации*, *назначения конкретного ЛТП*, *выбора решения* и *экстраполяции* последствий принятия решений.

Используя принципы ситуационного управления, задача принятия управляющих решений на любом уровне управления системой формулируется как поиск такого разбиения множества ситуаций на классы, при котором каждому классу соответствует решение, оптимальное с точки зрения критерия функционирования. Если такое решение получено, то управление ОУ на конкретном уровне реализуется следующим образом: по ситуации $s(t)$, зафиксированной в момент времени t , определяется класс, к которому принадлежит $s(t)$, и выбирается соответствующая команда управления, на основе которой ситуация $s(t)$ преобразуется в ситуацию $s(t + 1)$ и т.д. до тех пор, пока очередная ситуация не попадет в заранее фиксированный класс или не будет исчерпано заданное число этапов.

В ранних работах по ситуационному управлению предлагались различные семиотические языки описания ситуаций и средства их анализа и интерпретации [Клыкков, 1974], [Болотова, 2012] и др. С 1980-х гг. стали создаваться интеллектуальные системы, в архитектуре которых использовались многие принципы ситуационного управления [Поспелов, 2020]: экспертная система объединила в себе функции *анализатора*, *классификатора*, *коррелятора*, *экстраполятора* и базы знаний, в которой накап-

ливается история функционирования ОУ; блок имитационного моделирования осуществлял моделирование развития ситуаций (прогноз); интеллектуальный интерфейс обеспечивал общение с пользователем.

В данной работе при выборе метода ситуационного управления для практической реализации ИСУ учитывались: используемые в обучающих ИЭС/веб-ИЭС технологии инженерии знаний и онтологического инжиниринга; концептуально-функциональная многокомпонентность и неоднородность информационно-программных ресурсов пяти обучающих ИЭС/веб-ИЭС, а также динамической ИЭС для мониторинга информационной безопасности; различные модели представления знаний, разнообразные средства вывода и моделирования внешней среды; способы использования отдельных средств поддержки построения обучающих веб-ИЭС (инструментальный комплекс АТ-ТЕХНОЛОГИЯ) и др.

На основе исследования и развития базовых положений и принципов ситуационного управления применительно к современным интеллектуальным системам типа ИСУ [Макаров и др., 2006] была разработана архитектура *гибридной* ИСУ, реализующей целый ряд интеллектуальных и вычислительных функций и способной к адаптации и изменениям внешних условий функционирования ОУ, в качестве которого в данной работе выступают обучающая ИЭС/веб-ИЭС по конкретному курсу/дисциплине и процессы ее функционирования в течение полного цикла (или отдельных этапов) интеллектуального обучения.

Поскольку речь идет о новом уровне автоматизации и дальнейшего развития обучающих ИЭС/веб-ИЭС в контексте современных ИСУ, то построение *общей модели управления* по ситуациям, складывающимся на интеллектуальном ОУ, полностью определяется тем, насколько существующие информационно-программные ресурсы, используемые в обучающих ИЭС/веб-ИЭС, можно соотнести с множеством возможных ситуаций (состояний) S процессов функционирования ОУ, возможностью выделения классов ситуаций и соответствующих управляющих решений R .

Изначально структура управления интеллектуальным ОУ в виде обучающей ИЭС/веб-ИЭС имеет *иерархический* характер, причем ситуации S на каждом уровне управления обладают различной полнотой и точностью описания реализованных задач, а число допустимых управляющих решений R меньше числа ситуаций S . В настоящее время разработано полное описание и общая структура объемной (свыше 90 вершин) ситуационной сети (StS), отражающей несколько уровней управления – *мониторинговый, обобщенный, два компонентных* (верхний и нижний), где на соответствующих уровнях находятся системы, программные средства и компоненты, обеспечивающие реализацию конкретных управляющих решений R для типовых задач интеллектуального обучения и мониторинга процессов функционирования обучающих веб-ИЭС.

На основе разработанной ситуационной сети StS и с использованием средств поддержки построения прикладных онтологий, входящих в состав инструментального комплекса АТ-ТЕХНОЛОГИЯ, построена *онтологическая модель* Z и разработаны процедуры и алгоритмы, реализующие взаимодействия отдельных элементов онтологии.

Соответственно, новая модель *обобщенной онтологии* представления знаний о направлении/специальности подготовки описывается в виде:

$$G = \langle O, Z, F, RV, Sem, Goal \rangle, \quad (2)$$

где $O = \bigcup_{i=1}^5 O_i$ – прикладные онтологии курса/дисциплины;

$Z = \langle N, S, Ax \rangle$ – онтологическая модель ситуационной сети управления StS ;

$N = \{MN, GN, HC, LC\}$ – типы уровней ситуаций управления, где MN – уровень мониторинга; GN – обобщенный уровень управления; HC – компоненты верхнего уровня; LC – компоненты нижнего уровня;

$S = \{S_i\}$ – множество ситуаций управления, где $S = \bigcup_{k \in N} S^k$, где $S^{MN} = \{S_j^{MN}\}, j = 1..5$ – множество ситуаций управления, связанных с уровнем мониторинга, $\{S_p^{GN}\}, p = 1..7$ – множество ситуаций управления, связанных с обобщенным уровнем, $\{S_l^{HC}\}, l = 1..7$ – множество ситуаций управления, связанных с компонентами верхнего уровня, $\{S_h^{LC}\}, h = 1..14$ – множество ситуаций управления, связанных с компонентами нижнего уровня.

$Ax = \{Ax_i\}, i = 1..q$ – множество аксиом и/или правил обобщения ситуаций, q – количество аксиом/правил;

$F = F_B \cup F_S$ – операции с элементами онтологии Z , где $F_B = \{F1, F2, F3, F4\}$ – базовые операции с элементами онтологии O_i , $F_S = \{F_{Si}\}$ – множество операций с элементами онтологии Z , реализующих процедуры типа ЛТП для отображения текущей ситуации на онтологию и выбора управляющего воздействия (решения) и модификации моделей;

$RV = \{RV_i\}, i = 1..t$ – множество правил взаимодействия элементов Z с элементами O , t – количество правил;

$Sem = \{Sem_i\}$ – множество этапов образовательного цикла, определенных на конкретном интервале времени (за семестр, курс и др.);

$Goal = \{Goal_j^{Sem_i}\}$ – множество целей по достижению обучаемыми целевых профессиональных компетенций на конкретном этапе.

Таким образом, разработанная обобщенная онтология представления знаний о направлении/специальности подготовки (G) совместно с базовыми средствами вывода, функционирующими в составе интеллектуальной программной среды комплекса АТ-ТЕХНОЛОГИЯ, обеспечивают реализацию классической схемы ситуационного управления (анализ, кластеризация, корреляция) в рамках архитектуры современной ИСУ. Для

описания онтологии G и ее отдельных компонентов использовались предусмотренные в рамках задачно-ориентированной методологии базовые средства представления и обработки знаний [Рыбина, 2008], [Рыбина, 2014] в виде расширенного языка представления знаний (ЯПЗ), универсального АТ-РЕШАТЕЛЯ и темпорального решателя.

Основной функцией универсального АТ-РЕШАТЕЛЯ при взаимодействии с онтологией Z является анализ и классификация текущей ситуации управления S_i (используется режим конфигурации АТ-РЕШАТЕЛЯ) и поиск (выбор) решения на основе базы ЛТП, описанной на ЯПЗ. Все ЛТП реализуются на основе процедур выбора управляющего воздействия R_i , выполнение которого приводит к переходу к новой ситуации, а R_i реализуется как запуск готовых программ. Отметим, что целью управления является достижение для каждого обучаемого целевых профессиональных компетенций (описанных в прикладных онтологиях курсов/дисциплин [Рыбина, 2023b]), в течение образовательного цикла.

4. Особенности программной реализации адаптивной онтологической среды интеллектуального обучения в виде архитектуры интеллектуальной системы управления

Программная реализация АОЕ выполнена в виде архитектуры ИСУ (рис. 3), реализованной на принципах ситуационного подхода к управлению, где в качестве ОУ выступает обучающая веб-ИЭС и ее процессы функционирования.

Для моделирования случайных воздействий внешней среды на ОУ в составе архитектуры предусмотрены средства имитационного моделирования. Система управление реализуется на основе взаимодействия средств вывода и обобщенной онтологии представления знаний о направлении/специальности подготовки (G), в состав которой включена онтологическая модель ситуационной сети управления (Z).

Поскольку, в отличие от сложных ОУ технической природы, обучающая веб-ИЭС и процессы ее функционирования рассматриваются как ОУ программно-организационной природы, то роль исполнительных механизмов заключается в обеспечении программной поддержки управляющих воздействий.

Общая сложность реализации ИСУ заключается использовании большого количества (свыше 90) разнородных программных компонентов с различными уровнями функциональности, между которыми реализуются достаточно сложные механизмы взаимодействия, определяемые логикой решения типовых задач интеллектуального обучения с использованием обучающих ИЭС/веб-ИЭС. Поэтому важной частью исследований являлся выбор адекватного подхода к программной реализации ИСУ.



Рис. 3. Адаптивная онтологическая среда интеллектуального обучения в виде архитектуры ИСУ, реализованной на принципах ситуационного подхода

В данном случае наиболее предпочтительным является *сборочное программирование* [Лаврищева, 2025а], т.к. выполняются все необходимые условия: большое число разнообразных программных продуктов, использующихся как объекты сборки; возможность паспортизации программных объектов сборки; достаточное число стандартных правил сопряжения объектов и алгоритмов их реализации и средств автоматизации процесса сборки; наличие технологий применения разработанных объектов для использования в более сложных программах.

Из современных методологий сборочного программирования на основе сравнительного анализа был выбран *компонентно-ориентированный (CBD)* подход к разработке, где базовым понятием является программный компонент – независимый, самостоятельно реализованный программный объект, доступ к которому возможен с помощью интерфейсов, определяющих его функции и порядок обращения к его операциям.

Каждый интерфейс описывает механизм удаленного вызова компонентов, располагающихся в разных узлах сети или среды (сеть строится на основе стандартной семиуровневой модели открытых систем OSI). В рамках CBD реализована стратегия объединения «Каркас», обеспечивающая интеграцию программных компонентов в ИСУ и позволяющая отделить функции частей системы от управления, т.е. определять процедуры управления на любых более высоких уровнях.

С использованием методов сборочного программирования разработана прототипная версия ИСУ, обеспечивающая управление процессами функционирования обучающей веб-ИЭС на всех уровнях в контексте реализации типовых задач интеллектуального обучения.

Заключение

В настоящее время проводятся комплексные экспериментальные исследования прототипной версии ИСУ на основе серии имитационных экспериментов с применением разработанных и базовых средств и технологий интеллектуальной программной среды инструментального комплекса АТ-ТЕХНОЛОГИЯ.

Список литературы

- [Алешенко и др., 2016] Алешенко А.С., Трембач В.М. Интеллектуальная обучающая система кафедры ВУЗа // Открытое образование. – 2016. – № 5. – С. 47-52.
- [Башлыков и др., 2025] Башлыков А.А., Еремеев А.П. Основы конструирования интеллектуальных систем поддержки принятия решений в энергетике: учебник. – М.: ИНФРА-М, 2025. – 351 с.
- [Болотова, 2012] Болотова Л.С. Системы искусственного интеллекта: модели и технологии, основанные на знаниях: учебник. – М.: Финансы и статистика, 2012. – 664 с.
- [Брусиловский, 1990] Брусиловский П.Л. Интеллектуальные обучающие системы // Информатика. Информационные технологии. Средства и системы. – 1990. – № 2. – С. 3-22.
- [Гаврилова, 2016] Гаврилова Т.А. Инженерия знаний. Модели и методы: учебник. – СПб.: Изд-во «Лань», 2016. – 324 с.
- [Клыков, 1974] Клыков Ю.И. Ситуационное управление большими системами. – М.: Энергия, 1974. – 136 с.
- [Лаврищева, 2025а] Лаврищева Е.М. Программная инженерия. Парадигмы, технологии и CASE-средства: учебник для вузов. 2-е изд., испр. – М.: Изд-во Юрайт, 2025. – 280 с.
- [Макаров и др., 2006] Макаров И.М., Лохин В.М., Манько С.В., Романов М.П. Искусственный интеллект и интеллектуальные системы управления. – М.: Наука, 2006. – 333 с.
- [Поспелов, 1986] Поспелов Д.А. Ситуационное управление. Теория и практика. – М.: Наука, 1986. – 288 с.
- [Поспелов, 2020] Поспелов Д.А. Ситуационное управление: новый виток развития // Мягкие измерения и вычисления. – 2020. – Т. 29, № 4. – С. 20-27.
- [Рыбина, 2008] Рыбина Г.В. Теория и технология построения интегрированных экспертных систем: монография. – М.: Изд-во «Научтехлитиздат», 2008. – 482 с.
- [Рыбина, 2014] Рыбина Г.В. Интеллектуальные системы: от А до Я: Серия монографий в трех книгах. Кн. 1. Системы, основанные на знаниях. Интегрированные экспертные системы. – М.: Научтехлитиздат, 2014. – 224 с.
- [Рыбина, 2023а] Рыбина Г.В. Интеллектуальные обучающие системы: анализ базовых архитектур и особенностей применения онтологического подхода (на примере опыта разработки и практического использования обучающих интегрированных экспертных систем) // Приборы и системы. Управление, контроль, диагностика. – 2023. – № 2. – С. 23-43.

- [Рыбина, 2023b] Рыбина Г.В. Интеллектуальные обучающие системы на основе интегрированных экспертных систем: учебное пособие. – М.: Директ-Медиа, 2023. – 132 с.
- [Сычев и др., 2025] Сычев О.А., Пенской Н.А., Терехов Г.В. Метод разработки интеллектуальных тренажеров на основе онтологии предметной области // Онтология проектирования. – 2025. – Т. 15, № 1(55). – С. 67-81.
- [Фролов, 2024] Фролов И.А. Онтологический подход в управлении адаптивной подготовкой групп специалистов // Онтология проектирования. – 2024. – Т. 14, № 2(52). – С. 205-216.
- [Цветков, 2023] Цветков В.Я. Ситуационное управление // Современные технологии управления. – 2023. – № 2(102). – С. 4-15.
- [Bonner et al., 2015] Bonner D., Walton J., Dorneich M.C., Gilbert S.B., Winer E., Sottolare R.A. The development of a testbed to assess an intelligent tutoring system for teams // Workshops at the 17th International Conference on Artificial Intelligence in Education, AIED-WS 2015; CEUR Workshop Proceedings. 2015.
- [Carbonell, 1970] Carbonell J.R. AI in CAI: An artificial intelligence approach to computer-assisted instruction // IEEE Transactions on Man-Machine Systems. – 1970. – V. MMS-11, No. 4. – P. 190-202.
- [Gribova et al., 2018] Gribova V.V., Ostrovskii G.E. An intelligent cloud service for training diagnostics skills // Biomedical Engineering. – 2018. – Vol. 51, No. 6. – P. 416-421.
- [Nye, 2015] Nye B.D. Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context // International Journal of Artificial Intelligence in Education. – 2015. – No. 25. – P. 177-203.
- [Rahman et al., 2016] Rahman A.A., Abdullah M., Alias S.H. The architecture of agent-based intelligent tutoring system for the learning of software engineering function point metrics // 2nd International Symposium on Agent, Multi-Agent Systems and Robotics, ISAMSR 2016. – 2016. – P. 139-144.
- [Rybina et al., 2019] Rybina G.V., Blokhin Y.M. Methods and Software Implementation of Intelligent Planning for Integrated Expert System Design // Scientific and Technical Information Processing. – 2019. – Vol. 46, No. 6. – P. 434-445.
- [Rybina et al., 2022] Rybina G.V., Nikiforov A.Y., Slinkov A.A., Grigoryev A.A. Automated Formation of the Unified Ontological Space of Students' Knowledge and Skills to Implement Intellectual Tutoring Tasks Based on Tutoring Integrated Expert Systems // 2022 VI International Conference on Information Technologies in Engineering Education (Inforino). – 2022. – P. 1-6.
- [Rybina et al., 2023] Rybina G.V., Grigoryev A.A. Modern architectures of intelligent tutoring systems based on integrated expert systems: features of the approach to the automated formation of the ontological space of knowledge and skills of students // Pattern Recognition and Image Analysis. – 2023. – Vol. 33, No. 3. – P. 491-497.
- [Rybina et al., 2024] Rybina G.V., Grigoryev A.A. Educational Design of Dynamic Intelligent System Prototypes Using Tutoring Integrated Expert Systems // 2024 7th International Conference on Information Technologies in Engineering Education (Inforino). – 2024. – P. 1-6.
- [Sosnovsky et al., 2008] Sosnovsky S., Mitrovic A., Lee D., Brusilovsky P., Yudelson M. Ontology-based integration of adaptive educational systems // 16th International Conference on Computers in Education (ICCE 2008). – 2008. – P. 11-18.

Секция 10 | **ПРИКЛАДНЫЕ
ИНТЕЛЛЕКТУАЛЬНЫЕ
СИСТЕМЫ**

УДК 004.8

doi: 10.15622/rcai.2025.087

**КОНЦЕПЦИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ
АНАЛИТИКИ ЗДОРОВЬЯ ПАЦИЕНТОВ
С ХРОНИЧЕСКИМИ ЗАБОЛЕВАНИЯМИ
С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКИХ МОДЕЛЕЙ¹**

Т.В. Афанасьева (*afanaseva.tv@rea.ru*)

**Российский экономический университет им. Г.В. Плеханова,
Москва**

В работе представлена концепция пространственно-темпоральной аналитики многомерных числовых данных пациентов с хроническими неинфекционными заболеваниями (ХНЗ), основанная на применении нечетких моделей. В связи с увеличением продолжительности жизни количество пациентов с ХНЗ будет увеличиваться, что приведет к увеличению нагрузки на медицинский персонал. Для поддержки принятия врачебных решений при ведении таких пациентов актуальным является вычисление оценок не только состояния, но и изменений в показателях здоровья пациентов. В статье формулируется задача и приводится модель интеллектуальной системы оценивания здоровья пациентов с ХНЗ в пространственно-временном аспекте. Рассматриваются варианты применения предложенной модели.

Ключевые слова: многомерные показатели здоровья, ХНЗ, модель нечеткого логического вывода, нечеткий временной ряд.

¹ Исследование выполнено за счет гранта Российского научного фонда № 25-21-20116, <https://rscf.ru/project/25-21-20116>.

Введение

Почти половина взрослого населения РФ страдает от хронических заболеваний согласно итогам диспансеризации 2019 года. В 2020 в Европе более 60% всех случаев смерти обусловлены хроническими заболеваниями, причем среди них значительная часть трудоспособного населения в возрасте от 30 до 69 лет. К основным типам хронических неинфекционных заболеваний (ХНЗ) относятся сердечно-сосудистые и онкологические заболевания, хронические респираторные заболевания (хроническая обструктивная болезнь легких и астма) и сахарный диабет [Профилактика хронических неинфекционных заболеваний в Российской Федерации, 2022]. Пациенты с ХНЗ как правило находятся под наблюдением врача, который реализует мониторинг показателей здоровья на основе результатов анамнеза и регулярных клинических исследований. Как результат по каждому такому пациенту накапливаются значительные массивы данных показателей, изменяемых во времени. Со стороны медицинских работников отмечается временные ограничения медицинских специалистов на всесторонний анализ данных пациента, в том числе предиктивный, что приводит к снижению эффективности врачебных решений и врачебным ошибкам.

Это актуализирует проблему поддержки врачебных решений в направлении автоматизации обработки пространственно-темпоральных данных отдельного пациента.

Для повышения эффективности врачебных решений и снижении временных трудозатрат на их выработку активно развиваются системы поддержки врачебных решений, среди которых системы, использующие нечеткий логический вывод занимают достойное место. Это объясняется следующими факторами: анализируемые значения клинических показателей носят интервальный и нечеткий характер, использование которых в системах нечеткого логического вывода позволяет адекватно моделировать когнитивные процессы и знания медицинских специалистов, обеспечивая лингвистическую интерпретируемость медицинских данных и результатов их анализа [Кобринский, 2016].

Анализ публикаций наряду с растущим интересом к системам на основе моделей нечеткой логики для анализа числовых данных пациентов с хроническими заболеваниями выявил следующие подходы оценивания здоровья пациентов с ХНЗ: (1) оценивание здоровья по совокупности текущих показателей; (2) диагностика хронического заболевания или его стадии на основе методов классификации; (3) прогнозирование отдельных показателей здоровья с использованием нечетких временных рядов.

В то же время направление, дополняющее эти три подхода на основе извлечения информации о типе изменений в показателях здоровья, с целью формирования пространственно-темпоральной оценки здоровья па-

пациентов с хроническими заболеваниями с единых методологических позиций, не получило достаточного развития. В статье совокупность показателей, характеризующих здоровье пациентов с хроническими заболеваниями, рассматривается как динамическая система, в которой изменения состояния и тенденций в дополнение к оцениванию текущего и прогнозного состояния позволит повысить информативность их анализа и, как следствие, эффективность лечебно-диагностического процесса пациентов с ХНЗ.

Целью работы является разработка концептуальных основ комплексного оценивания здоровья пациентов с хроническими заболеваниями с учетом пространственно-темпоральной аналитики их числовых показателей на основе моделей нечеткого логического вывода.

1. Обзор применения нечетких моделей для аналитики числовых показателей здоровья пациентов с ХНЗ

Система на основе нечетких логических правил, в которой для раннего выявления диабета использовались показатели самомониторинга: уровень глюкозы в крови, индекс массы тела, генетическая предрасположенность, возраст представлена в работе [Aamir et al., 2021]. Авторы пришли к выводу, что предложенная система является интерпретируемым, высокоточным и эффективным инструментом для ранней диагностики диабета 2 типа. В отличие от моделей машинного обучения, их система позволяет проследить процесс принятия решений и использовать знания в виде понятных правил, и при этом модель не требует большого количества вычислительных ресурсов.

Нечеткая шкала для оценки риска смертности, ориентированная на пациентов с хроническими заболеваниями рассмотрена в работе [Jatobá et al., 2018]. Модель этой шкалы охватывает кроме физиологических показателей дополнительные аспекты жизни пациентов, такие как состав семьи, условия проживания и другие аспекты жизни пациентов. Результаты оценивания показали, что во всех случаях модель нечеткого вывода дала оценки, близкие к мнению врачей.

В исследовании [Hussain et al., 2016] предлагается нечеткая экспертная система, обрабатывающая данные, полученные от семи сенсоров для амбулаторного наблюдения и мониторинга пациентов с хроническими заболеваниями сердца (в стабильных условиях). Входные данные пациентов включают температуру, частоту сердечных сокращений, ЭКГ, систолическое давление, насыщение кислородом крови, частоту дыхания, состояние кожных покровов. Выходные данные показывают наличие или отсутствие риска развития сердечно-сосудистых заболеваний. Если обнаруживаются отклонения в показателях, специальное устройство предупреждает врачей, которые находятся удаленно.

Система нечеткого логического вывода Мамдани, которая сфокусирована на ранжировании пациентов с хронической болезнью почек, находящихся в списке ожидания трансплантации почки, описана в статье [Taherkhani et al., 2022]. В этой системе использованы 8 лингвистических переменных, а продукционные правила были построены с помощью дерева решений по базе примеров. Авторы утверждают, что разработанная система нечеткого логического вывода демонстрирует более высокую эффективность и справедливость в распределении донорских почек по сравнению с традиционными методами. Система успешно имитирует интуитивное мышление экспертов и будет способствовать улучшению выживаемости после трансплантации донорских почек.

Две входные лингвистические переменные с пятью нечеткими термами (очень низкий, низкий, средний, высокий, очень высокий) для оценки риска рака ротовой полости были использованы авторами [Scrobotă et al., 2017] в разработанной системе поддержки принятия решений на основе нечеткой логики. В работе отмечается, что экспертно оцененные результаты позволяют сделать вывод: модель разработанной системы может стать инструментом для неинвазивного скрининга и поддержки клинических решений.

Вопросу разработки метода, объединяющего основы теорий нечетких множеств и формального концептуального анализа для автоматизированного оценивания состояния сердечно-сосудистого здоровья мужчин и представления результатов в форме предложений на естественном языке посвящена статья [Afanasieva et al., 2020]. В этой статье используются 6 групп показателей здоровья пациентов: клинические показатели, показатели образа жизни, психологические показатели, показатели физического здоровья и факторы риска, а также симптомы сердечно-сосудистых событий. Для каждого показателя были сформированы лингвистические переменные, позволяющих определять степень принадлежности к двум лингвистическим термам «норма» и «вне нормы» согласно клиническим рекомендациям Минздрава РФ.

Целью исследования [Afanasieva et al., 2021] является разработка подхода к представлению пространственно-темпоральной оценки сердечно-сосудистого здоровья пациентов в виде протоформ и пропозиций, то есть предложений на естественном языке. В рамках предложенного подхода рассматриваются нечеткие лингвистические оценки каждого показателя здоровья и их тенденции изменения в заданный момент времени, которые затем резюмируются с использованием нечетких квантификаторов, основанных на частотности полученных лингвистических оценок здоровья пациента.

Применение нечеткого логического вывода Мамдани для вычисления обобщающей оценки состояния здоровья и ее изменения, по клиническим показателям крови рассмотрено в статье [Wilinski et al., 2023]. Авторы

представляют подход, в рамках которого анализируют изменения показателей и визуализируют их в виде графиков. В качестве примера рассмотрены показатели анализа крови онкологического пациента, которые регистрировались в течение трех лет с ежемесячной периодичностью.

Учитывая, что для пациентов с хроническими заболеваниями важно регулярно отслеживать изменения в показателях в статье [Abbasi et al., 2014] отмечается, что результаты прогнозирования временных рядов среднего артериального давления может помочь врачам выбрать более правильное лечение на основе их анализа. Для этой цели авторы представили метод прогнозирования временных рядов на основе нечетких функций в многомодельном режиме.

Использование нечетких временных рядов совместно с оптимизацией для выбора количества нечетких множеств с целью прогнозирования уровня глюкозы в крови для пациентов с диабетом предложено в исследовании [Ozogur et al., 2021].

На основании проведенного обзора применения нечеткой логики в анализе числовых показателей здоровья пациентов с ХНЗ можно сделать следующий вывод: модели нечеткого логического вывода представляет собой эффективный, интерпретируемый и гибкий инструмент, способный решать задачи оценивания здоровья пациентов. Такие модели позволяют формализовать экспертное знание, получать дополнительную информацию, способствующую улучшению точности врачебных решений и сделать их более прозрачными как для врача, так и для пациента.

2. Концептуальные основы пространственно-темпорального анализа показателей здоровья пациентов с ХНЗ

2.1. Постановка задачи

Пусть для оценивания состояния здоровья известны m числовых показателей здоровья, наблюдаемые на временном интервале $[1, k]$ и входные данные пациента с ХНЗ в виде числовых значений, распределенных в пространстве m показателей здоровья пациента, упорядоченных по моментам времени t :

$$X = \left\{ x_j(t), x_j(t) \geq 0, x_j(t) \in W_j, W_j \subset \mathbb{R}, W_j = \left[\min_{x_j}, \max_{x_j} \right] \right\}, \quad (2.1)$$

где $j = 1, 2, \dots, m; t = 1, 2, \dots, k$.

Требуется с использованием нечетких моделей определить общую пространственно-темпоральную оценку состояния пациента по наблюдениям X , включающую оценки текущего состояния здоровья в пространстве заданных m числовых показателей и темпоральные оценки состояния здоровья, которые рассматриваются в виде текущих к моменту времени k и прогнозируемых изменений состояния здоровья на следующий период

$k + 1$. Под оценкой состояния здоровья будем понимать общую характеристику состояния здоровья, выражаемую как в количественной, так и в качественной форме в виде лингвистически значимых терминов. Решение поставленной задачи позволит врачу составить целостное представление о состоянии здоровья пациента.

2.2. Пространственно-темпоральная модель анализа показателей здоровья пациентов с ХНЗ

Пусть $x \in X$ — некоторый числовой показатель здоровья пациента, значения которого могут изменяться на интервале значений W_j . Предположим, что на W_j существует бинарное отношение \leq удовлетворяющее свойствам рефлексивности, транзитивности и антисимметричности, а также, что:

а) W_j покрыт упорядоченными по номерам интервалами (возможно перекрывающимися) $\{p_i | i = 1, 2, \dots, r\}$, для которых существует отношение частичного порядка;

б) каждый интервал p_i является носителем нечеткого множества, обозначаемого лингвистическим термом $\tilde{Z}_i \in \tilde{Z}$, для которого определена функция принадлежности $\mu_{\tilde{Z}_i}(\cdot): W_j \rightarrow [0, 1]$, удовлетворяющая свойствам нормальности и выпуклости;

в) каждому значению $x \in W_j$ соответствует по крайней мере один лингвистический терм $\tilde{Z}_i \in \tilde{Z}$, для которого $\mu_{\tilde{Z}_i}(x) > 0$, следовательно множество функций принадлежности $\{\mu_{\tilde{Z}_i}(x) | i = 1, 2, \dots, r\}$ покрывает множество W_j .

Представленное выше описание определяет лингвистическую переменную L_j , построенную для некоторого показателя здоровья в фиксированный момент времени:

$$L_j = \langle \tilde{Z}_{ij}, r_j, W_j, P_j, Mf_j \rangle, \quad (2.2)$$

где P_j обозначает множество интервалов носителей функций принадлежности на W_j , Mf_j определяет вид функции принадлежности $\mu_{\tilde{Z}_i}(x) | i = 1, 2, \dots, r_j$.

Предположим, что для любого показателя здоровья определено одинаковое количество интервалов разбиения r и единообразный вид функций принадлежности Mf , имеющих терм-множество \tilde{Z} с одинаковыми лингвистическими терминами, применяемые в практике качественного оценивания клинических показателей пациента, такие как \tilde{Z}_1 = “в пределах нормы”, \tilde{Z}_2 = “риск”, \tilde{Z}_3 = “значимое отклонение”. Тогда лингвистическую переменную j -того показателя здоровья (2.2) можно представить в виде

$$\begin{aligned}
L_j &= \langle C, \Psi_j \rangle, \\
C &= \langle \tilde{Z}, r, Mf \rangle, \\
\Psi_j &= \{W_j, P_j\}.
\end{aligned}$$

Множество Ψ_j содержит настраиваемые на j -тый показатель здоровья пациента параметры лингвистической переменной L_j . Оно включает интервал W_j изменения j -го показателя, а каждое P_j включает r интервалов, на которых определяются функции принадлежности. Часть лингвистической переменной, обозначаемая C , не зависит от конкретного показателя, ее будем рассматривать как общую для всех показателей.

Аналогично L (2.2) определим выходную лингвистическую переменную Q для оценки общего состояния пациента, для которой термножество также будет включать три значения \tilde{E}_1 = “в пределах нормы”, \tilde{E}_2 = “риск”, \tilde{E}_3 = “патология”:

$$Q = \langle \tilde{E}, r_Q, W_Q, P_Q, Mf_Q \rangle. \quad (2.3)$$

Пространственно-темпоральную модель анализа показателей здоровья пациентов с ХНЗ определим в виде системы

$$\begin{aligned}
A &= \langle X, L, Q, \Omega, Res \rangle, \\
L &= \{L_j\}, j = 1, 2, \dots, m.
\end{aligned} \quad (2.4)$$

Здесь X – это модель пациента (2.1), L – множество лингвистических переменных, определенных для входных показателей здоровья (2.2), а Q – выходная лингвистическая переменная (2.3) для оценки текущего состояния здоровья в пространстве заданных m числовых показателей здоровья.

Множество функций Ω используются в системе A для анализа показателей X в пространственно-темпоральном контексте и получения результатов Res в виде совокупности оценок:

$$Res = \langle e_k, \Delta e_k, \hat{e}_{k+1}, \{Tend_{jk}\} \rangle. \quad (2.5)$$

Результаты e_k и Δe_k обозначают общую числовую оценку по всем показателям здоровья в момент времени k и изменение этой оценки по сравнению с предыдущей общей оценкой соответственно. Значение \hat{e}_{k+1} обозначает прогноз общей оценки состояния здоровья на следующий период наблюдения, а переменная $Tend_{jk}$ определяет одну из заранее заданных тенденций изменения каждого показателя, представленную лингвистическим термом из множества {«негативная тенденция», «позитивная тенденция», «стабильность»}. Таким образом, числовая оценка e_k характеризует текущее состояние здоровья пациента с ХНЗ в пространстве заданных m числовых показателей, оценка Δe_k описывает количественное из-

менение этой оценки, произошедшее к моменту времени k , оценка \hat{e}_{k+1} представляет возможное изменение оценки e_k в следующий момент времени. Тенденция $Tend_{jk}$ в лингвистической форме показывает изменение каждого показателя здоровья к моменту времени k , позволяя доктору принимать решения, основанные на качественном изменении конкретного показателя пациента.

Определим следующую совокупность допустимых функций Ω в системе A для получения результатов Res в виде множества:

$$\Omega = \{Fuzz, Ling, Compare, FIS, deFuzz, Predict\}. \quad (2.6)$$

Функция фазификации $Fuzz$ вычисляет степени принадлежности значения $x \in W$ нечетким множествам $\tilde{Z}_i \in \tilde{Z}$:

$$Fuzz: (x, L) \rightarrow y; y = (y_1, y_2, \dots, y_r); y_i = \mu_{\tilde{Z}_i}(x); i = 1, 2, \dots, r.$$

Функция $Ling$ определяет номер s и наименование s -того нечеткого множества (лингвистического термина) $\tilde{Z}_s \in \tilde{Z}$, которому с максимальной степенью принадлежит значение x :

$$Ling: y \rightarrow \tilde{Z}_s; \text{если } y_s \geq y_h, \forall h = 1, 2, \dots, r; y_s, y_h \in y.$$

Функция $Ling$ обеспечивает преобразование разнородных числовых показателей здоровья пациента с ХНЗ в однородные качественные оценки из терм-множества \tilde{Z} , заданные в виде лингвистических оценочных термов.

Полагая семантическую упорядоченность лингвистических термов $\tilde{Z}_i \in \tilde{Z}$ по возрастанию их номеров, определим функцию сравнения $Compare$ между двумя лингвистическими терминами с номерами $n1$ и $n2$, оценивающими тенденцию изменения показателя здоровья в моменты времени $(t - 1)$ и t :

$$Compare: (\tilde{Z}_{n1,t-1}, \tilde{Z}_{n1,t} \rightarrow Tend_t; n1, n2 \in [1, r];$$

$$Tend_t = \begin{cases} \tilde{T}_1, & n1 > n2 & (\text{падение}) \\ \tilde{T}_2, & n1 < n2 & (\text{рост}) \\ \tilde{T}_3, & n1 = n2 & (\text{стабильность}). \end{cases}$$

В зависимости от клинических рекомендаций по конкретному показателю здоровья в контексте конкретного заболевания термины «падение», «стабильность» и «рост», определяющие нечеткую тенденцию $\tilde{T}_u, u = 1, 2, 3$ могут иметь противоположную интерпретацию в терминах «положительная тенденция» и «отрицательная тенденция».

Функция FIS используется для реализации нечеткого логического вывода по базе продукционных правил $Rule$, сформированной медицинскими экспертами или моделями машинного обучения для вывода лингвистического термина \tilde{E} и его функции принадлежности

$$FIS: (y, Rule, Q) \rightarrow \langle \tilde{E}, \mu_{\tilde{E}}(W_Q), Alg \rangle,$$

где Alg определяет алгоритм реализации нечёткого логического вывода.

Модель нечеткого логического вывода рассматривается в виде системы правил *Rule* вида «ЕСЛИ [предпосылка] – ТО [заключение]».

Примером продукционного правила может служить правило

«ЕСЛИ [уровень глюкозы в крови «Вне нормы» И систолическое артериальное давление «Вне нормы»] ТО [общая оценка «Патология»].

Модель нечеткого логического вывода может включать операции агрегации степени истинности предпосылок правил с использованием *T*-норм, активизации заключений и агрегации заключений на основе *S*-конорм. Применительно к поставленной задаче модель нечеткого логического вывода используется для вычисления оценки общего состояния здоровья пациента с ХНЗ по множеству его числовых показателей.

Функция дефаззификации *DeFuzz* с использованием выбранного метода *Meth* вычисляет приближенное числовое значение по нечеткому представлению показателя здоровья пациента:

$$deFuzz: (\tilde{E}, \mu_{\tilde{E}}(W_Q), Meth) \rightarrow e.$$

Функция *Predict* вычисляет прогнозное значение общей оценки здоровья пациента \hat{e}_{k+1} на следующий временной интервал и может быть реализована на основе модели нечеткого временного ряда *FTS*:

$$Predict: (\{e_t, t = 1, 2, \dots, k\}, FTS) \rightarrow \hat{e}_{k+1}.$$

В качестве модели нечеткого временного ряда может быть использована модель, предложенная в статье [Song, 2003] или модель временного ряда нечетких тенденций, рассмотренная в работе [Afanasyeva, 2015]. В результате применения функции *Predict* и последующей дефаззификации вычисляется числовая оценка \hat{e}_{k+1} . Применение модели нечеткого временного ряда *FTS* обосновано в случае небольшого количества наблюдений за пациентами вне стационара, когда показатели здоровья пациентов фиксируются с интервалом раз месяц или реже. В условиях наблюдения за состоянием здоровья пациентов в условиях стационара для прогнозирования могут быть использованы другие модели временных рядов.

2.3. Варианты применения пространственно-темпоральная аналитики показателей здоровья пациентов с ХНЗ

Рассмотрим последовательности применения функций модели (2.4).

Так для получения общей оценки состояния здоровья пациентов с ХНЗ в конкретный момент времени по множеству показателей *X* применима следующая схема алгоритма в виде последовательность функций из множества Ω модели (2.4):

$$\langle \{x_j(t), j = 1, 2, \dots, m; t = k\}, L, Q, Rule \rangle \rightarrow Fuzz \rightarrow FIS \rightarrow deFuzz \rightarrow e_k.$$

Используя эту последовательность для моментов времени $t = (k - 1)$ и $t = k$, числовое изменение общей оценки вычисляется по формуле

$$\Delta e_k = e_k - e_{k-1}.$$

Схема алгоритма для прогнозирования общей оценки здоровья представима следующей последовательностью функций:

$$< \{e_1, e_2, \dots, e_k\}, p > \rightarrow Fuzz \rightarrow Predict \rightarrow deFuzz \rightarrow \hat{e}_{k+1}.$$

Для получения лингвистической оценки тенденции в изменении показателя здоровья к моменту времени необходимо предварительно получить их нечеткие значения в момент времени $t = (k - 1)$ и $t = k$.

Тогда следующие функции образуют последовательность для оценки тенденций показателя здоровья

$$\begin{aligned} < \{x_j(t), j = const; t = (k - 1), L\} > \rightarrow Fuzz \rightarrow Ling \rightarrow \tilde{Z}_{j,k-1} \\ < \{x_j(t), j = const; t = k, L\} > \rightarrow Fuzz \rightarrow Ling \rightarrow \tilde{Z}_{j,k} \\ < \tilde{Z}_{n1,t-1}, \tilde{Z}_{n1,t} > \rightarrow Compare \rightarrow Tend_{jk}. \end{aligned}$$

Таким образом показано, что для реализации пространственно-темпорального анализа показателей пациентов с ХНЗ в рамках поставленной задачи достаточно функций (2.6), определенных в модели (2.4), которые позволяют получить результаты (2.5). Отметим, что конкретная реализация этих функций будет определяться выбором способа их реализации, используемыми параметрами лингвистических переменных и семантикой экспертных правил.

Заключение

В связи с актуальностью автоматизации обработки постоянно растущих объемов данных пациентов с ХНЗ в статье рассмотрены концептуальные основы пространственно-темпоральной аналитики числовых показателей здоровья пациентов с ХНЗ с использованием нечетких моделей и экспертных знаний. Экспертные знания выражены в виде параметров лингвистических переменных и продукционных правил, необходимых для нечеткого логического вывода. Была сформулирована задача и модель комплексного анализа многомерных показателей здоровья для получения пространственно-темпоральных оценок, включающие оценки состояния здоровья и его изменений. Приведены схемы алгоритмов в виде последовательности функций, описанных в составе предложенной пространственно-темпоральной модели аналитики показателей пациентов в ХНЗ.

В отличие от пространственно-темпоральной модели, рассмотренной в статье [Afanasieva et al., 2021], в настоящем исследовании используется модель нечеткого логического вывода для получения общей оценки состояния здоровья пациента, а не статистика качественных оценок по множеству показателей здоровья и тенденций их изменения. Другое отличие

предложенной в статье модели анализа показателей здоровья пациента состоит в спектре функций, включающих функцию прогнозирования, и, соответственно, в содержании получаемых результатов.

Разработанные концептуальные основы аналитики многомерных числовых показателей пациентов с ХНЗ будут способствовать развитию программных систем поддержки в лечебно-диагностическом процессе любых хронических неинфекционных заболеваний пациентов в стабильных условиях, так как обеспечивают гибкую настройку правил и параметров функций принадлежности. Результаты, полученные при реализации модели пространственно-темпоральной аналитики могут быть интегрированы в ЭМК пациентов, что, как представляется, сократит временные затраты и повысит качество врачебных решений. В дальнейшем планируется расширить предложенную модель и получить на ее основе конкретные методы аналитики, ориентированные на конкретные ХНЗ.

Список литературы

- [Профилактика хронических неинфекционных заболеваний в Российской Федерации, 2022] Драпкина О.М. [и др.]. Профилактика хронических неинфекционных заболеваний в Российской Федерации // Кардиоваскулярная терапия и профилактика. – 2022. – 21(4):3235. – doi:10.15829/1728-8800-2022-3235.
- [Кобринский, 2016] Кобринский Б.А. Нечеткость в клинической медицине и необходимость ее отражения в экспертных системах // Врач и ИТ. – 2016. – № 5. – С. 6-15.
- [Aamir et al., 2021] Aamir KM, Sarfraz L, Ramzan M, Bilal M, Shafi J, Attique M. A Fuzzy Rule-Based System for Classification of Diabetes // Sensors. – 2021. – Vol. 21(23). – 8095. – doi.org/10.3390/s21238095.
- [Abbasi et al., 2014] Abbasi R., Moradi M.H. and Molaezadeh S.F. Long-term prediction of blood pressure time series using multiple fuzzy functions // In: Proc. 21th Iranian Conference on Biomedical Engineering (ICBME), Tehran, Iran, 2014. – P. 124-127. – doi: 10.1109/ICBME.2014.7043906.
- [Afanasyeva et al., 2015] Afanasyeva T., Yarushkina N., Toneryan M., Zavarzin D., Sapunkov A., Sibirev I. Time series forecasting using fuzzy techniques // In: Proc. 2015 Conference of the International Fuzzy Systems Association and the European Society for Fuzzy Logic and Technology. Gijón, Asturias. Spain. – doi.org/10.2991/IFSA-EUSFLAT-15.2015.151.
- [Afanasyeva et al., 2020] Afanasyeva T.V. and Rodionova T.E. Methodology of patient-oriented assessment of cardiovascular health of men using fuzzy sets and formal conceptual analysis // World Scientific Proceedings Series on Computer Engineering and Information Science Developments of Artificial Intelligence Technologies in Computation and Robotics. – 2020. – P. 857-865. – https://doi.org/10.1142/9789811223334_0103.
- [Afanasyeva et al., 2021] Afanasyeva T., Perfilieva I., Kozhevnikov V. Approach to Patient Assessment Based on a Spatial-Temporal Model for Decision Support Systems in Cardiology // In: Proc. Fifth International Scientific Conference “Intelligent Information Technologies for Industry” (ITI’21). Lecture Notes in Networks and Systems. Vol. 330. – Springer, Cham, 2021. – P. 461-471. – https://doi.org/10.1007/978-3-030-87178-9_46.

- [**Hussain et al., 2016**] Hussain Aamir et al. Personal Home Healthcare System for the Cardiac Patient of Smart City Using Fuzzy Logic // Journal of Advances in Information Technology. – 2016. – Vol. 7. – P. 58-64. – doi: 10.12720/jait.7.1.58-64.
- [**Jatobá et al., 2018**] Jatobá A., et al. Supporting decision-making in patient risk assessment using a hierarchical fuzzy model // Cogn Tech Work. – 2018. – Vol. 20. – P. 477-488. – doi.org/10.1007/s10111-018-0475-1.
- [**Ozogur et al., 2021**] Hatice Nizam Ozogur, Gokhan Ozogur, Zeynep Orman. Blood glucose level prediction for diabetes based on modified fuzzy time series and particle swarm optimization // Computational Intelligence. – 2021. – Vol. 37. – P. 155-175.
- [**Scrobotă et al., 2017**] Scrobotă I, Băciuț G, Filip AG, Todor B, Blaga F, Băciuț MF. Application of Fuzzy Logic in Oral Cancer Risk Assessment // Iran J Public Health. – 2017. – Vol.46(5). – P. 612-619.
- [**Song, 2003**] Song, Q. A note on fuzzy time series model relation with sample autocorrelation functions // Cybernetics and Systems: An International Journal. – 2003. – Vol. 34. – P. 93-107.
- [**Taherkhani et al., 2022**] Taherkhani N, Sepehri MM, Khasha R, Shafaghi S. Ranking patients on the kidney transplant waiting list based on fuzzy inference system // BMC Nephrol. – 2022. – Vol. 23(1). – doi: 10.1186/s12882-022-02662-5.
- [**Wilinski et al., 2023**] Wilinski A., Tadeusiewicz R., Piegat A., Bocewicz G., Skorzak A., Dabkowski K., Smereczynski A., Starzynska T. A fuzzy interval model for assessing patient status and treatment effectiveness using blood morphology // Healthc. Anal. – 2023. – Vol. 4. – DOI:10.1016/j.health.2023.100234.

УДК 004.056

doi: 10.15622/rcai.2025.088

СНИЖЕНИЕ СЛОЖНОСТИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ НА ОСНОВЕ МЕТОДА АВТОМАТИЧЕСКОЙ ЛОКАЛИЗАЦИИ АНОМАЛИЙ В МНОГОМЕРНЫХ ПРОСТРАНСТВАХ

А.Я. Бучаев (*abdulhamid0055@yandex.ru*)

И.И. Комаров (*i_krov@mail.ru*)

Университет ИТМО, Санкт-Петербург

В работе представлен способ снижения сложности интеллектуального анализа, основанный на автоматической локализации и группировке однородных и аномальных объектов в многомерных пространствах. Предложенный метод направлен на динамическую работу в условиях изменяющейся входной выборки, такой результат достигается использованием статистического анализа и формального метода принятия решения, что также позволяет легко интерпретировать результаты группировки и локализации. Таким образом, интеллектуальный анализ фокусируется на статистически неустойчивых объектах выборки и отделяет их друг от друга, что обеспечивает снижение риска атаки уклонения с использованием видимых вредоносных вставок. В качестве примера практического применения полученных результатов рассматривается динамическая маска матрицы внимания визуального трансформера. Применение предложенного решения позволяет исключить квадратичную сложность механизма внимания без предварительного регулирования маски и в режиме реального времени.

Ключевые слова: кибербезопасность, определение порога аномальности, вектор признаков, статистическая устойчивость, интеллектуальный анализ данных, механизм принятия решений, визуальный трансформер.

Введение

При интеллектуальном анализе среди множества обрабатываемых объектов и их признаков встречаются однородные выборки объектов, представленные в различных видах от скользящих окон при анализе временных рядов до фрагментов изображений. При этом существующие меха-

низмы, предназначенные для сопоставления фактов и выделения информации для дальнейшего интеллектуального анализа, используют всю подаваемую на вход информацию, описывающую объект выборки.

Побочным эффектом такого подхода являются повышение вычислительной сложности процесса выделения ключевой информации, снижение фокусировки интеллектуального анализа, а также возникновение потенциального риска атаки уклонения на модель интеллектуального анализа.

Предполагается, что обобщение однородных объектов позволит снизить вычислительную сложность при обработке входных данных интеллектуальной моделью. Такой подход позволяет обратить внимание на статистически нерегулярные объекты, которые несут в себе больше информации, и проводить их интеллектуальный анализ независимо.

Процесс предобработки входной выборки объектов для интеллектуального анализа должен сопровождаться выделением области интересов, основанном на формализованных, доказуемых и главное – интерпретируемых методах предобработки и принятия решения. Таким образом, прикладная система интеллектуального анализа снизит время обработки данных и потребляемые ресурсы, улучшит фокусировку на нетривиальных объектах и разграничит объекты анализа, сохраняя интерпретируемость выходных результатов, так как степень декомпозиции и возможность качественной интерпретации результатов влияют на конечное восприятие и уровень доверия пользователя в отношении используемой системы.

В работе предлагается способ снижения сложности интеллектуального анализа данных за счет автоматической локализации аномалий в многомерных пространствах на основе формального метода определения порога принятия решения.

Целью данной работы является снижение ресурсоемкости при интеллектуальной обработке данных и обеспечение безопасности путем локализации объектов и аномалий в многомерном пространстве.

Объект исследования – процесс интеллектуального анализа данных, представленный набором объектов.

Предмет исследования – формальные методы определения порога принятия решения в задачах кластеризации.

1. Научно-методический аппарат

Одним из актуальных механизмов интеллектуального анализа является механизм внимания, определяющий взаимную «важность» между входными объектами. Существенным ограничением механизма внимания является квадратичная сложность. Для решения этой проблемы авторы работы [Sun et al., 2025] предлагают ряд подходов, включая различные типы масок механизма внимания, сведение самовнимания к линейной сложности, в том числе гибридные вариации моделей, комбинирующие линейные

слои и слои с полным вниманием, подчеркивая, что снижение вычислительных затрат является важным этапом для будущих исследований долгосрочных последовательностей, мультимодальности и сложного (многоэтапного) мышления.

Фиксированные маски для матрицы внимания выполняют разделение связей объектов друг с другом на две группы: связь объектов есть; нулевая связь, но такой разделение не являются универсальным решением, необходимо подбирать предложенные маски в соответствии с предполагаемым графом связей объектов друг с другом. Такой подход не учитывает физическую сущность и динамику процессов, поэтому необходимо проводить разделение связей объектов друг с другом на основе формального и динамического метода определения порога разделения или принятия решения.

Методы определения порога принятия решения разделяются на две большие группы: формальные и эвристические, в рамках данной работы необходимым требованием является автономность модуля принятия решения, то есть рассматривается полностью автоматическое решение. Исходя из этого, интересующей группой являются формальные методы определения порога, которые подразделяются на:

- статистические методы на основе Монте-Карло, которые учитывают ретроспективную статистику. Такие методы являются статичными и требуют периодического обновления информации для соответствия актуальным объектам информационной системы [Shen et al., 2021], [Xu et al., 2022];
- методы с применением теории игр используют экспертную аналитику для оценки влияния тех или иных факторов на отклик целевой функции. В дальнейшем полученные оценки входят в основу математической модели [Ni et al., 2024], [Li et al., 2021];
- методы математического программирования (методы оптимизации) применяются, когда задача механизма принятия решения описывается, например, системой линейных уравнений, решаемых методами линейного программирования. В случаях, когда зависимости механизма принятия решения не сводятся к линейным, применяют методы нелинейного программирования, которые в общем случае решают задачу поиска локальных оптимумов.

Задачу разделения совокупности элементов на типичные и нетипичные (аномальные) элементы можно рассматривать как задачу разделения совокупности объектов на несколько групп, то есть поиск границ или порогов разделения совокупности. В таком случае, имея минимум два класса – типичные и нетипичные элементы, можно использовать метод бинаризации Оцу [Otsu et al., 1979], который основан на принципе минимизации внутриклассовой дисперсии, влекущей максимизацию межклассовой дисперсии. Данный метод имеет вычислительную сложность $O(n)$, где

n – число возможных значений. Изначально метод использовался для би-наризации изображений градаций серого, однако Авторы работы [Liao et al., 2001] представили метод разделения совокупности элементов на множество классов по тому же принципу максимизации межклассовой дисперсии Оцу. Рассмотренные методы сегментации изображений применимы и в другой области определения – вещественных числах. Главным компонентом при использовании методов разделения выборки на классы на основе подхода Оцу является гистограмма значений, формируемая из области определения объектов выборки.

В качестве исходной выборки объектов будет использоваться совокупность фрагментов размером 100 на 100 пикселей изображения (рис. 1.а), обработка которого в рамках задачи выявления нетипичных объектов выполняется в соответствии с предыдущей научной работой [Бучасев и др., 2024]. Разделение изображения на фрагменты является одним из этапов обработки данных визуальным трансформером в части кодирования. Масштабируемость используемых и предложенных методов позволяет обрабатывать входные совокупности различной природы, например, в работе [Ляховенко и др., 2024] Авторы описывают модель построения многомерного пространства автономных транспортных средств. Однако для простоты визуализации и обработки данная работа сосредоточена на обработке изображений визуальным трансформером.

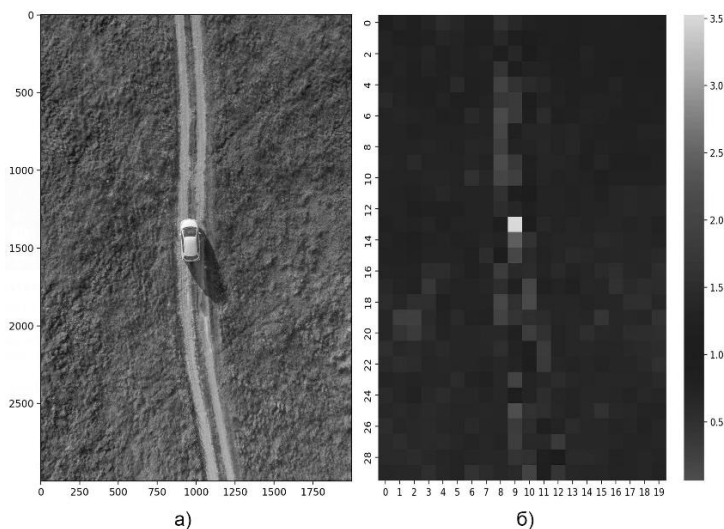


Рис. 1. а) исходное изображение; б) тепловая карта в соответствии с показателем отклонения дисперсии

В таком случае разделяемой на классы совокупностью будет являться набор показателей отклонений дисперсии фрагментов изображения (рис. 2). При использовании классического метода Оцу было получено следующее разделение на группы типичных (значения меньше порога) и нетипичных (значения больше порога) фрагментов изображения, порог принятия решения в данном случае равен 1,127.

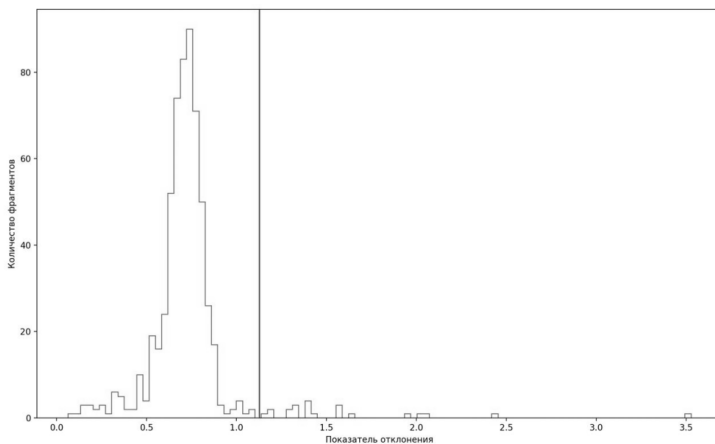


Рис. 2. Гистограмма плотности распределения значений показателя отклонения дисперсии фрагментов изображения с разделением на классы

2. Определение строгости порога аномальности относительно степени однородности изображения

Определение порога аномальности необходимо проводить с учётом динамики информационной системы, в том числе – физической сущности и динамики процессов. Предполагается, что для однородной выборки автоматическое вычисление порога должно быть более строгим, чем вычисление порога в разнородной выборке, где в рамках общей статистики выбросом является однородный фрагмент, то есть деление на классы происходит наоборот. Следовательно, в случае однородной выборки степень допустимого отклонения дисперсии текущего фрагмента от общей дисперсии выборки должна быть меньше.

Гипотеза А: абсолютная разница среднего и медианы объекта, имеющего равномерное распределение вероятностей значений каждого компонента вектора описателя объекта по отдельности и представляющего однородную структуру (количественное преимущество имеет только одно значение среди наборов значений по каждой компоненте, то есть низкая энтропия), стремится к нулю.

Обоснование А (для изображения): в случае, когда структура отдельной компоненты фрагмента X представлена преимущественно одним значением $s = \{0, 1, 2, 3, \dots, 255\}$, и распределение вероятностей значений структуры отдельной компоненты фрагмента стремится к равномерному, $\bar{X} \rightarrow s$; $M_e(X) \rightarrow s$, следовательно, $|\bar{X} - M_e(X)| \rightarrow 0$.

Согласно Гипотезе А, чем меньше абсолютная разница среднего и медианы структуры отдельной компоненты фрагмента, тем более однородное и равномерное распределение значений компоненты он имеет. Для вычисления «степени» однородности [Arafeh et al., 2022] предполагается использовать расстояние Хеллингера, которое оценивает степень разности двух дискретных распределений вероятности $P = (p_1, p_2, p_3, \dots, p_k)$ и $Q = (q_1, q_2, q_3, \dots, q_k)$ [Kumari et. al., 2017] (2.1).

$$H(P, Q) = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^k (\sqrt{p_i} - \sqrt{q_i})^2}, \quad (2.1)$$

где k – количество значений дискретного распределения; p_i – вероятность k -го элемента.

Предлагается сравнивать распределение вероятностей текущего фрагмента и распределение вероятностей сгенерированного набора значений (для изображений от 0 до 255), имеющего распределение близкое к равномерному. Результаты, полученные в ходе проверки гипотезы, показаны на рис. 3.

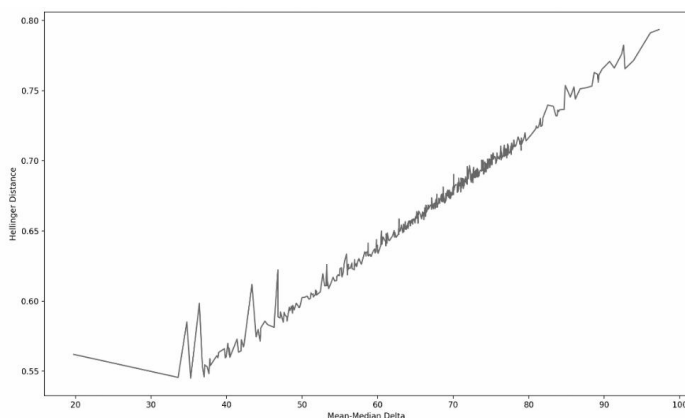


Рис. 3. Усредненная по трем компонентам изображения зависимость расстояния Хеллингера от разности среднего и медианы

Предложенная гипотеза о разности среднего и медианы имеет ряд ограничений, связанных со случаями, когда фрагмент данных содержит в себе перетекающие градиенты. В таком случае $|\bar{X} - Me(X)| \rightarrow 0$, однако структура отдельной компоненты фрагмента не имеет равномерного рас-

пределения вероятностей и однородности в совокупности значений компоненты. Подобные случаи можно наблюдать в первом, третьем и пятом наблюдениях графика, исходные части которых изображены на рис. 4.

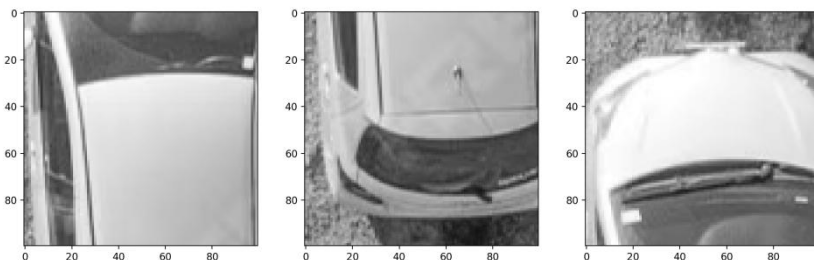


Рис. 4. Фрагменты изображения, не соответствующие предложенной гипотезе о разнице среднего и медианы

3. Результат работы и преимущества автоматической локализации

На исходном изображении (рис. 1,а) можно наблюдать 2 однородные среды – зеленое поле и элементы дороги (включая боковые насыпи). Используя модификацию метода Оцу для многоклассового разделения, были получены результаты, представленные на рис. 5,а. Для данного случая были получены два значения, разделяющие совокупность на три среды – 0,543 и 1,161.

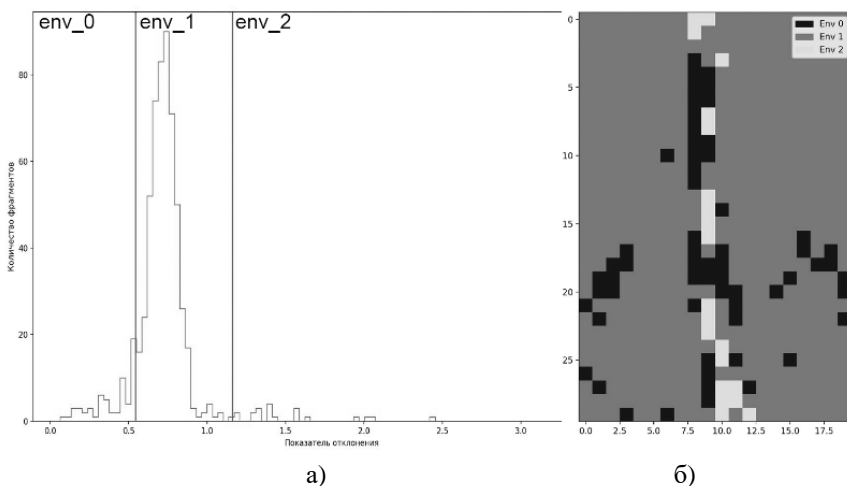


Рис. 5. а) разделение совокупности значений показателя отклонения дисперсии фрагментов на три класса; б) группировка однородных фрагментов изображения

Определяя строгость порога аномальности в соответствии с абсолютной разницей медианы и среднего, были получены результаты сегментирования исходного изображения на три класса. На рис. 5,б можно наблюдать разделение фрагментов на классы в соответствии со средами, определяемым формальными статистическими методами.

Результат работы предложенного метода позволяет сгруппировать однородные фрагменты изображения, динамически формируя маску для механизма внимания, то есть при векторизации похожие фрагменты изображения будут иметь большую связь в матрице внимания, а объекты, принадлежащие разным группам (кластерам), будут иметь слабую связь. Таким образом, динамически формируется маска для матрицы внимания с учетом физической сущности обрабатываемой выборки.

Заключение

В работе представлен способ снижения сложности интеллектуального анализа данных, основанный на методе автоматической локализации и группировки однородных объектов и аномальных объектов входной совокупности обрабатываемых данных. Предложенный метод масштабируется на пространства большей размерности, является интерпретируемым и динамическим, что соответствует требованиям технологий, в которых предполагается его применение. Вместе с этим осуществляется снижение риска атаки уклонения, так как группировка объектов разграничивает объекты, содержащие вредоносные вставки, от легитимных объектов из-за нарушения статистической устойчивости.

В дальнейшей планируется проведение экспериментов по преобразовке входных данных и последующей обработке визуальным трансформером с использованием динамической маски матрицы внимания.

Список литературы

- [Бучаев и др., 2024] Бучаев А.Я., Бегаев А.Н., Комаров И.И. Метод автоматического обнаружения аномалий в пространстве событий информационной безопасности // Промышленные АСУ и контроллеры. – 2024. – № 2. – С. 31-41.
- [Ляховенко и др., 2024] Ляховенко Ю.А., Мухамеджанов С., Комаров И.И. Метод Построения Модели Пространства Автономных Транспортных Средств // International Journal of Open Information Technologies. – 2024. – Т. 12, № 7. – С. 38-42.
- [Arafteh et al., 2022] Arafteh, Mohamad & Hammoud, Ahmad & Otrok, Hadi & Mourad, Azzam & Talhi, Chamseddine & Dziong, Zbigniew. Independent and Identically Distributed (IID) Data Assessment in Federated Learning. – 2022. – P. 293-298. – 10.1109/GLOBECOM48099.2022.10001718.
- [Kumari et al., 2017] A. Kumari and U. Thakar, "Hellinger distance based over-sampling method to solve multi-class imbalance problem," 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), Nagpur, India, 2017. – P. 137-141. – doi: 10.1109/CSNT.2017.8418525.

- [**Li et al., 2021**]. Li Y, Fan Z. and Zhao X. Game Decision of a Closed-loop Supply Chain with components remanufacturing // 2021 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE), Shanghai, China, 2021. – P. 1-6. – doi: 10.1109/RASSE53195.2021.9686926.
- [**Liao et al., 2001**] Liao, Ping-Sung, Tse-Sheng Chen and P. C. Chung. A Fast Algorithm for Multilevel Thresholding // J. Inf. Sci. Eng. – 2001. – 17. – P. 713-727.
- [**Ni et al., 2024**] Ni W. and Zhang Y. Threshold Detection of Cloud Computing System with Impatient Users Based on Game Theory // 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Tianjin, China, 2024. – P. 3128-3133. – doi: 10.1109/CSCWD61410.2024.10580422.
- [**Otsu et al., 1979**] Otsu N. A threshold selection method from gray-level histogram // IEEE Transactions on System Man Cybernetics. – 1979. – Vol. SMC-9, No. 1. – P. 62-66.
- [**Shen et al., 2021**] Shen L. and Wang S. Monte Carlo Tree Search for Network Planning for Next Generation Mobile Communication Networks // 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021. – P. 1-6. – doi: 10.1109/GLOBECOM46510.2021.9685526.
- [**Sun et. al., 2025**] Sun W., Hu J., Zhou Y. et. al. Speed always wins: A survey on efficient architectures for large language models [Электронный ресурс] // arXiv. – 2025. – Режим доступа: <https://arxiv.org/abs/2508.09834> (дата обращения: 16.08.2025).
- [**Xu et. al., 2022**] Xu L., Hurtado-Grueso J., Jeurissen D., Liebana D.P. and Dockhorn A. Elastic Monte Carlo Tree Search with State Abstraction for Strategy Game Playing // 2022 IEEE Conference on Games (CoG), Beijing, China, 2022. – P. 369-376. – doi: 10.1109/CoG51982.2022.9893587.

УДК 004 891.3

doi: 10.15622/rcai.2025.089

ОБЪЕДИНЕНИЕ МУЛЬТИСЕНСОРНЫХ ДАННЫХ МНОГОСТАДИЙНОГО ПРОИЗВОДСТВА ДЛЯ ПРОГНОЗИРОВАНИЯ ОТКАЗОВ ТЕХНОЛОГИЧЕСКОГО ОБОРУДОВАНИЯ

А.Н. Ветров (*vetrov_48@mail.ru*)

Б.В. Палюх (*pboris@tstu.tver.ru*)

Тверской государственный технический университет, Тверь

В этой работе мы предлагаем новый подход, объединяющий мультисенсорные данные для обнаружения бифуркации многомерного случайного процесса, порождаемого работой технологического оборудования непрерывного многостадийного производства. В статье описана параметрическая многомерная модель обнаружения аномалий временного ряда, методика и процедура определения точки бифуркации. Рассмотрен метод скользящего окна для объединения мультисенсорных данных и их агрегация для регистрации аномалий в режиме реального времени.

Ключевые слова: объединение мультисенсорных данных, динамические экспертные системы, диагностика, точка бифуркации.

Введение

Мультисенсорное объединение данных (МСОД) – это технология, позволяющая объединять информацию из нескольких источников для формирования единой картины [Ciuonzo, 2014]. Технология МСОД отбирает полезную информацию от различных датчиков, использует определенные правила для объединения этой взаимной информации, чтобы обеспечить наиболее надежную и точную информацию для повышения эффективности системы. Объединение на уровне данных подразумевает интеграцию необработанных данных, получаемых от датчиков для получения агрегированной информации, необходимой для принятия решений. Объединение данных на уровне принятия решений, позволяет получать не только конкретную информацию, получаемую от датчиков. Этот способ обеспечивает лучшую интеграцию разнородных данных. Объединение данных для принятия решений хорошо сочетается с использованием байесовской

теории логического вывода, теории свидетельств Демпстера - Шафера, искусственных нейронных сетей и системы нечеткого логического вывода. В данной статье мы рассматриваем применение технологии МСОД для обнаружения аномалий во временных рядах, порождаемых потоками данных от сенсоров технологического оборудования [Paliukh, 2024], [Ветров, 2023].

Обнаружение аномалий во временных рядах связано с выявлением неожиданных изменений поведения системы для получения важной информации, сигнализирующей о дефектах в ней. Во многих промышленных приложениях обнаружение аномалий используется для мониторинга сбоев в работе технологического оборудования, оповещения пользователей о внешних атаках, обнаружения потенциальных катастрофических событий на ранней стадии. Несмотря на все прилагаемые усилия, разработка хорошего универсального алгоритма обнаружения аномалий является чрезвычайно сложной задачей. В последнее время для поиска аномалий во временных рядах успешно используются нейронные сети [Yangdong et al., 2019], [Su et al., 2019], [Zhao et al., 2020] [Zekai et al., 2021]. Одна из них VAE – Variational Autoencode [Dohyung et al, 2018]. VAE – это тип генеративной вероятностной модели, способной семплировать новые объекты, не принадлежащие исходным данным. Модель VAE была использована для обнаружения аномалий в различных областях с хорошими результатами [Lin et al, 2020]. Другая модель – LSTM. Это рекуррентные нейронные сети (RNN), способные эффективно работать с последовательными данными. Примеры использования моделей RNN для обнаружения аномалий приведены в [Yangdong et al., 2019], [Dan et al., 2019]. В работе [Malhotra, 2016] предложена гибридная модель, объединяющая модели VAE и LSTM.

В нашей работе мы использовали параметрическую модель векторной авторегрессии для автоматического обнаружения аномалий. Это обусловлено требованиями, которые предъявляются к системе обнаружения.

Основная часть этой статьи организована следующим образом: в разделе 1 обсуждается задача объединения мультисенсорных данных для обнаружения бифуркаций и требования к системе обнаружения. Основные определения, параметрическая многомерная модель обнаружения аномалий временного ряда, процедура объединения мультисенсорных данных для поиска точки бифуркации описаны в разделе 2. В разделе 3 рассматриваются метод скользящего окна для объединения мультисенсорных данных и их агрегация для регистрации аномалий в режиме реального времени. В заключении представлены полученные основные результаты.

1. Задача объединения мультисенсорных данных для обнаружения бифуркаций

Для многих промышленных предприятий обнаружение потенциально катастрофических событий на ранней стадии является актуальной проблемой. Для ее решения на базе информационных технологий и методов искусственного интеллекта нами была разработана динамическая экспертная система (ЭС) ExpSystem [Shabanov et al., 2019], [Palyukh et al., 2021], [Ветров, 2023]. Интеграционная схема программного комплекса представлена на рис. 1.

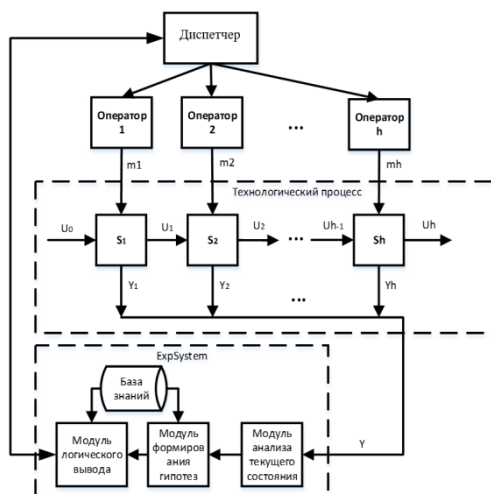


Рис. 1. Схема взаимодействия ЭС ExpSystem с системой управления многостадийным производством

Здесь использованы следующие обозначения: S_i – i ая стадия многостадийного производства; U_i , Y_i – технологические параметры; m_i – управляющие воздействия. Система ориентирована на широкий класс технологических процессов, а именно на непрерывные многостадийные процессы. Особенность таких процессов заключается в том, что весь производственный процесс разбит на отдельные элементы – технологические стадии, которые связаны между собой непрерывными материальными потоками. Например, технологический процесс производства неконцентрированной азотной кислоты состоит из следующих основных стадий:

1. Подготовка и компримирование воздуха.
2. Подготовка газообразного аммиака.
3. Подготовка аммиачно-воздушной смеси.
4. Окисление аммиака и охлаждение нитрозных газов.

5. Абсорбция оксидов азота.

6. Селективная низкотемпературная каталитическая очистка хвостовых газов от оксидов азота и рекуперация энергии хвостовых газов в газовой турбине.

7. Хранение и выдача продукционной кислоты.

На каждой стадии используется свое, специфическое оборудование, включающее множество различных агрегатов. Для того, чтобы контролировать состояние такого сложного процесса используются десятки показателей, которые называются технологическими параметрами. Обеспечение оперативного контроля, управление и защита процесса от аварийных ситуаций, сигнализация отклонений от номинальных режимов осуществляется с помощью автоматизированной системы управления технологическим процессом. В ее функции входит контроль режимов технологического процесса, измерение технологических параметров с помощью сенсоров, сбор данных, представление полученной информации о состоянии технологического объекта оперативно-технологическому персоналу.

Состояние технологического процесса оценивается на основе значений технологических параметров, измеряемых с определенной периодичностью. Анализ данных о состоянии процесса позволяет определить момент его перехода в критический режим. Индикатором перехода технологического процесса в критический режим служит аномальное динамическое поведение значений технологических параметров. Такое поведение есть следствие влияния на технологический процесс внешних и внутренних факторов. Например, неисправность или выход из строя оборудования, изменение характеристик поступающего в производство сырья и материалов. В экспертной системе ExpSystem заложена функция мониторинга значений технологических параметров. Анализ их значений является основой для раннего предупреждения дефектов, отказов и аварий.

При работе технологического оборудования в штатном режиме порождаемый им случайный процесс изменения технологических параметров является стационарным с математическим ожиданием не зависящим от времени. Его колебания обусловлены шумами и имеют ограниченный характер.

Причиной появления аномалий во временном ряду является переходный процесс, который возникает в системе как следствие влияния внешних и внутренних факторов на работу технологического оборудования. Таким образом стационарный процесс переходит в нестационарный. Наличие тренда в нестационарном процессе является причиной перехода системы в предаварийный режим. Определение момента начала переходного процесса (точки бифуркации) и выявление тренда по отдельным временным рядам с помощью метода двойного скользящего окна описано в нашей работе [Paliukh, 2024]. На самом деле наблюдаемый случайный процесс является многомерным и отдельные его компоненты коррелированы между собой.

Учет взаимосвязи компонент многомерного случайного процесса делает соответствующие оценки в определении точки бифуркации более достоверными и снижают вероятность возникновения ложной тревоги.

Кроме того, анализ многомерных временных рядов поможет выявить кластеры связанных технологических показателей. В дальнейшем это позволит уточнить причину возникшей аномалии.

В данной работе решается задача объединения мультисенсорных данных для прогнозирования бифуркаций и определения взаимосвязи между технологическими показателями.

Исходя из вышеизложенного можно сформулировать минимальные требования к выбору метода решения поставленной задачи:

- 1) необходимо учитывать тот факт, что процесс нестационарный;
- 2) необходимо обрабатывать данные в режиме реального времени без предварительного обучения;
- 3) необходимо учитывать объединенную информацию от всех сенсоров;
- 4) необходима получать результат в ограниченные сроки.

2. Параметрическая многомерная модель обнаружения аномалий временного ряда

Контроль состояния производственного процесса осуществляется на основании дискретной последовательности измеренных значений множества технологических параметров. Эти последовательности образуют бесконечные многомерные временные ряды (МВР). Все измеренные значения являются отражением (реализацией) многомерного случайного процесса (МСП).

Будем обозначать его как вектор $Y(t) = (Y_1(t), \dots, Y_i(t), \dots, Y_m(t))^T$, где m соответствует числу сенсоров технологического оборудования. Каждому сенсору соответствует один технологический параметр. Будем обозначать временной ряд, порождаемый случайным процессом $Y_i(t)$ как бесконечная последовательность $y_i(t_1), y_i(t_2), \dots, y_i(t_k), \dots$ значений технологического параметра, поступающих от i -го сенсора в дискретные моменты времени t_k . Обозначим $y_i(t) = (y_i(t_1), y_i(t_2), \dots, y_i(t_N))^T$ вектор, содержащий конечный набор значений технологического параметра, зарегистрированный i -м сенсором. Для описания многомерных временных рядов будем использовать вектор $Y_t = (y_{1t}, \dots, y_{it}, \dots, y_{mt})^T$. Математическое ожидание стационарного МСП $M[Y] = (\mu_1, \dots, \mu_i, \dots, \mu_m)^T$ не зависит от времени, в то время как математическое ожидание нестационарного МСП $M[Y(t) = (\mu_1(t), \dots, \mu_i(t), \dots, \mu_m(t))^T]$ есть функция времени. Для анализа взаимосвязи многомерных случайных процессов используются автокорреляционная функция (АФ) $R_{Y_i}(t_1, t_2) = M[\dot{Y}_i(t_1)\dot{Y}_i(t_2)]$ и взаимная корреляционная функция (ВКФ)

$R_{Y_i Y_j}(t_1, t_2) = M[\dot{Y}_i(t_1)\dot{Y}_j(t_2)]$. Здесь $\dot{Y} = Y - M[Y]$. Если процесс стационарный, значения АФ и ВКФ зависят не от моментов времени t_1 и t_2 , а от их разницы $t_2 - t_1 = \tau$. Тогда АФ $R_{Y_i}(t_1, t_2) = R_{Y_i}(\tau) = M[\dot{Y}_i(t_1)\dot{Y}_i(t_1 + \tau)]$ и ВКФ $R_{Y_i Y_j}(t_1, t_2) = R_{Y_i Y_j}(\tau) = M[\dot{Y}_i(t_1)\dot{Y}_j(t_1 + \tau)]$. Как уже было отмечено, при работе технологического оборудования в штатном режиме, порождаемый им МСП является стационарным, что подтверждено проведенными исследованиями. Кроме того, МСП является эргодическим. Это позволяет вычислять оценки характеристик стационарного МСП на основе одной единственной реализации путем усреднения значений временных рядов по времени.

В отличие от стационарного процесса соответствующие характеристики нестационарного МСП являются функциями времени. Для вычисления оценок характеристик нестационарного МСП требуется усреднение значений временного ряда по ансамблю, поскольку многомерный закон распределения случайной величины в каждом временном сечении разный. Полученные по единственной реализации оценки в случае нестационарного случайного процесса будут недостоверными (смещенными, несостоятельными и неэффективными).

В такой ситуации параметрическое представление МВР является эффективной альтернативой классическому корреляционному описанию. Для обнаружения аномалий в МВР и определения точки бифуркации мы используем модель векторной авторегрессии.

Для того, чтобы учесть взаимную зависимость компонент случайного процесса выразим каждое значение временного ряда $y_i(t)$ в виде линейной комбинации остальных, т.е. $y_1(t), \dots, y_{i-1}(t), y_{i+1}(t), \dots, y_m(t)$, и их значений, сдвинутых на величину τ . Сдвиг τ принимает целочисленные значения от 1 до p . Таким образом мы имеем систему n линейных уравнений. Модель векторной авторегрессии имеет вид:

$$Y_t = B_1 Y_{t-1} + B_2 Y_{t-2} + \dots + B_p Y_{t-p} + \varepsilon_t. \quad (1)$$

Здесь использованы следующие обозначения. Вектор текущих значений временного ряда $Y_t = (y_{1t}, \dots, y_{it}, \dots, y_{mt})^T$. Векторы значений МВР, сдвинутых на величину $\tau - Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$. Соответственно имеем $Y_{t-1} = (y_{1-t-1}, \dots, y_{i-t-1}, \dots, y_{m-t-1})^T$, $Y_{t-2} = (y_{1-t-2}, \dots, y_{i-t-2}, \dots, y_{m-t-2})^T, \dots, Y_{t-p} = (y_{1-t-p}, \dots, y_{i-t-p}, \dots, y_{m-t-p})^T$.

B_1, B_2, \dots, B_p – матрицы коэффициентов векторов значений МВР, сдвинутых на величину τ . Произвольная матрица B_i содержит компоненты

$$B_l = \begin{pmatrix} b_{1l}^1 & b_{2l}^1 & \dots & b_{ml}^1 \\ b_{1l}^2 & b_{2l}^2 & \dots & b_{ml}^2 \\ \dots & \dots & \dots & \dots \\ b_{1l}^m & b_{2l}^m & \dots & b_{ml}^m \end{pmatrix}. \quad (2)$$

Вектор случайных компонент ε_t имеет вид $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{mt})^T$. Компоненты вектора, независимы имеют нулевое математическое ожидание и постоянную дисперсию. Используя параметрическую многомерную модель случайного процесса (1) сформулируем условия для определения точки бифуркации. Предварительно, используя оператор сдвига $Hu_t = y_{t-1}$, преобразуем выражение (1). В результате получим тождественное (1) выражение $(I - B_1H + B_2H^2 + \dots + B_mH^m)y_t = \varepsilon_t$.

Здесь I – единичная матрица $m \times m$, вида

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & \dots & \dots & 1 \end{pmatrix}$$

Рассмотрим характеристическое уравнение

$$\Phi(H) = I - B_1H + B_2H^2 + \dots + B_mH^m = 0, \quad (3)$$

где $\Phi(H)$ – квадратная матрица $m \times m$.

Корни характеристического уравнения, действительные и комплексно-сопряженные, можно вычислить как собственные числа матрицы $\Phi(H)$. Обозначим их $\lambda_1, \lambda_2, \dots, \lambda_i, \dots, \lambda_m$. Если выполняются условия

$$\lambda_1 \neq \lambda_2 \neq \dots \neq \lambda_i \neq \dots \neq \lambda_m \text{ и } |\lambda_i| > 0 \quad \forall i, \quad (4)$$

то рассматриваемый многомерный случайный процесс является стационарным. Определим точку бифуркации t_0 , как момент времени, в котором зафиксирован переход случайного процесса из стационарного состояния в нестационарное. Определим $\delta_Y(t)$ индикаторную функцию, указывающую на смену характера случайного процесса в момент времени t_0 , следующим образом

$$\delta_Y(t) = \begin{cases} 0 & \text{если } t < t_0 \text{ } |\lambda_i| > 0 \quad \forall i \\ 1 & \text{если } t < t_0 \text{ } |\lambda_i| \leq 0 \quad \exists i \end{cases}. \quad (5)$$

Значение $\delta_Y(t) = 0$ указывает на то, что процесс стационарный – выполняются условия стационарности (4). Значение $\delta_Y(t)$ равное 1, предполагает, что не для всех компонент МВР выполняются условия стационар-

ности (4). Таким образом, значение $\delta_Y(t)=1$ указывает на то, что случайный процесс носит нестационарный характер. Те значения собственных чисел λ_i , для которых условия стационарности нарушены, указывают на технологические параметры, значения которых выходят за пределы нормативных значений. Обозначим $K(t_0)=\{i \in \overline{1, m}: |\lambda_i| < 0\}$ множество номеров сенсоров и связанных с ними технологических параметров, которые указывают на нарушение штатного режима функционирования технологического оборудования.

После того, как был установлен факт нарушения технологического режима и выявлены технологические параметры, указывающие на нарушение штатного режима работы производственного оборудования, необходимо определить связи между этими параметрами. Для этого будем использовать временные ряды, которые формируются сенсорами с номерами $i \in K(t_0)$. Поскольку процесс нестационарный, использование традиционного корреляционного или спектрального анализа мало информативно. Для анализа связи между временными рядами будем проверять наличие коинтегрированности между ними. Для этого используем параметрическую многомерную модель, подобную модели (1), исключив из рассмотрения стационарные компоненты. Преобразуем каждую из оставшихся компонент к стационарному виду путем взятия конечных разностей. Обозначим Δ – конечная разность порядка d для временного ряда y_{it} , $i \in K(t_0)$. Определим коинтеграцию между временными рядами с индексами i и j , $i, j \in K(t_0)$. Параметрическая модель имеет вид

$$\begin{aligned}\Delta y_{it} &= a_{i1}^1 \Delta y_{it-1} + \dots + a_{i1p}^1 \Delta y_{it-p} + a_{i21}^1 \Delta y_{it-1} + \dots + a_{i2p}^1 \Delta y_{it-p} + \varepsilon_{it}, \\ \Delta y_{jt} &= a_{j1}^2 \Delta y_{jt-1} + \dots + a_{j1p}^2 \Delta y_{jt-p} + a_{j21}^2 \Delta y_{jt-1} + \dots + a_{j2p}^2 \Delta y_{jt-p} + \varepsilon_{jt},\end{aligned}$$

Представим эту модель в общем виде

$$\Delta Y_t = A_1 \Delta Y_{t-1} + A_2 \Delta Y_{t-2} + \dots + A_p \Delta Y_{t-p} + \varepsilon_{1t}.$$

Проверим модель на стационарность так, как было описано выше. Если корни матрицы $\Phi(H)$ $|\lambda_1| > 1$ и $|\lambda_2| > 1$, то коинтеграция между временными рядами y_{it} и y_{jt} отсутствует. В противном случае эти ряды коинтегрированы. Это означает, что между ними существует некая зависимость, которая определяет их общее поведение.

3. Скользящее окно для объединения мультисенсорных данных

Для проверки условий стационарности на основе конечного набора данных, поступающих непрерывно от сенсоров технологического оборудования применяется потоковая обработка, включающая агрегацию ин-

формации с использованием скользящих окон. Данные в бесконечном потоке фиксируются скользящим окном и агрегируются по мере того, как элементы данных вставляются и удаляются из скользящего окна. Скользящее окно строится на основе кортежей над бесконечным потоком $S - S_H[H_L, H_R]$, где $H_L > H_R$ любые два неотрицательных целых числа, определяющие левую и правую границы окна. Когда текущая длина потока равняется H , скользящее окно $S_H[H_L, H_R]$ содержит набор кортежей $\{y_j: \{|(H - H_L + 1)| \leq j \leq (H + H_R - 1)\}\}$. С течением времени состояние скользящего окна меняется. Обозначим k – номер состояния, целое положительное число. Значения левой и правой границ скользящего окна меняются при изменении состояния окна. Состояние скользящего окна определяется набором кортежей, входящих в него. Обозначим O_1, \dots, O_n набор операторов агрегации. Для входного потока S $O_j = f_A(S_H[H_L, H_R])$, где f – функция агрегации, а A – атрибут агрегации. Каждое значение атрибута агрегации a_i представляет собой элемент потока. В процессе обработки потоковых данных соответствующие значения элементов потока записываются в циклический буфер и обрабатываются операторами агрегации O_j , а результат агрегации сохраняется в памяти. Рассматриваемый метод двойного скользящего окна предполагает использование системы агрегации, состоящей из двух скользящих окон для каждого сенсора технологического оборудования. Первое скользящее окно W_1^i используется как низкочастотный фильтр, подавляющий кратковременные случайные выбросы значений технологических параметров. Здесь i – номер сенсора, $i = \overline{1; m}$. Второе скользящее окно W_2^i аккумулирует информацию для анализа состояния технологического процесса и определение точки бифуркации. Текущее состояние скользящего окна W_1^i определяется набором кортежей, хранящихся в циклическом буфере окна W_1^i в данный момент. Текущее состояние скользящего окна W_2^i определяется набором агрегированных значений, вычисленных в окне W_1^i .

Укрупненный алгоритм обработки потоковых данных содержит следующие шаги.

1. Считываем текущее состояние скользящего окна W_1^i из циклического буфера. $I_k^i = \{y_j \in S_H^i(H_{Lk}^i, H_{Rk}^i)\}$.
2. Вычисляем N значений агрегата $f_A(I_k^i)$, для всех элементов потока, соответствующих текущему состоянию k скользящего окна W_1^i .

3. Добавляем все вычисленные значения $f_A(I_k^i)$ в окно W_2^i . Обозначим эти значения $f_1^i, f_2^i, \dots, f_N^i$. Они определяют текущее состояние скользящего окна W_2^i .
4. Формируем объединенную матрицу исходных данных для расчета коэффициентов модели (1). Для этого считываем данные текущего состояния всех скользящих окон W_2^i из циклических буферов. Вычисляем коэффициенты векторной авторегрессии (1), соответствующие текущему состоянию всех скользящих окон W_2^i .
5. Решаем характеристическое уравнение (3). Если корни характеристического уравнения не удовлетворяют условию (4), индикаторная функция (4) принимает значение 1. Точка бифуркации соответствует моменту получения индикаторной функцией значения, равного единице, в реальном времени. Процесс поиска аномалий заканчивается. Если корни характеристического уравнения соответствуют условию (4), индикаторная функция принимает значение 0. Процесс поиска аномалий продолжается.
6. Продолжаем проверку условия (4). Удаляем значения “самого старого” интервала из окна W_1^i .
7. Добавляем значения следующего интервала в окно W_1^i .
8. Вычисляем значения $f_A(I_k)$ для последнего добавленного интервала.
9. Добавляем последнее вычисленное значения $f_A(I_k)$ в окно W_2^i .
10. Удаляем значения “самого старого” значения агрегата из окна W_2^i .
11. Возвращаемся на шаг 1.

Процедура обработки данных останавливается внешним прерыванием.

4. Результат апробации методики

Предложенная методика была апробирована на производстве неконцентрированной азотной кислоты. В течении 2 лет был проведен детальный анализ работы агрегата УКЛ-7, включающий сбор данных для дальнейшего исследования и проверки эффективности программного комплекса и разработанной методики, направленной на повышение эксплуатационной надежности оборудования. На основе этих данных была рассчитана вероятность безотказной работы многостадийного производства (табл. 1).

Таблица 1

Изменение показателя вероятности безотказной работы многостадийного производства

	Вероятность безотказной работы многостадийного производства	Вероятность отказа многостадийного производства
До внедрения	0,9868	0,0132
После внедрения	0,9882	0,0118
Улучшение показателя в %		~10,6%

Таким образом, применение предложенной методики позволило повысить эксплуатационную надежность многостадийного производства за рассматриваемый период на ~10,6%.

Заключение

В статье представлен новый подход объединяющий мультисенсорные данные для обнаружения бифуркации многомерного случайного процесса, порождаемого работой технологического оборудования непрерывного многостадийного производства. Достоинство этого подхода заключается в том, что в отличие от существующих разработок предлагаемая система детектирования бифуркаций не требует предварительного обучения. Обнаружение аномалий в технологическом процессе происходит on-line на основе анализа бесконечного многомерного потока технологических данных, получаемых от сенсоров оборудования. Проведенные испытания подтвердили эффективность предложенной методики детектирования бифуркаций.

Полученные результаты могут служить методологической основой для разработки динамической экспертной системы, работающей в режиме реального времени для прогнозирования отказов технологического оборудования непрерывного многостадийного производства.

Список литературы

- [Ветров, 2023] Ветров А.Н., Мальков, А.А., Палюх Б.В. Детектирование бифуркаций многомерных случайных процессов в динамических экспертных системах // Двадцать первая Национальная конференция по искусственному интеллекту с международным участием, КИИ-2023 (Смоленск, 15-20 октября 2023 г.): Труды конференции. В 2-х т. Т. 2. – С. 173-182.
- [Ciunzo, 2014] Ciunzo D., Salvo Rossi P. Decision Fusion with Unknown Sensor Detection Probability // IEEE Signal Processing Letters. – 21 (2). – P. 208-212. – doi: 10.1109/LSP.2013.2295054.

- [**Dan et al, 2019**] Dan Li, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks // In Artificial Neural Networks and Machine Learning: Text and Time Series. – 2019. – P. 703-716.
- [**Dohyung et al, 2018**] Dohyung Kim, Hyochang Yang, Minki Chung, Sungzoon Cho, Huijung Kim, Minhee Kim, Kyungwon Kim, and Eunseok Kim. Squeezed convolutional variational autoencoder for unsupervised anomaly detection in edge device industrial internet of things // In International Conference on Information and Computer Technologies. – 2018. – P. 67-71.
- [**Malhotra, 2016**] Malhotra P., Ramakrishnan A. LSTM-based encoder-decoder for multi-sensor anomaly detection. Available: arXiv:1607.00148 [cs.AI]. – 2016.
- [**Palyukh et al, 2021**] Palyukh B., Merkuriev S., Vetrov A., Shabanov B. and Sotnikov A. Methods for Forecasting Critical States of the Technological Process in the Evolutionary Management of Continuous Multi-Stage Production // 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). – 2021. – DOI: 10.1109/ElConRus51938.2021.9396111.
- [**Paliukh, 2024**] Paliukh B.V., Malkov A.A., Vetrov A.N. Bifurcation Detections of Multidimensional Random Processes in Dynamic Expert Systems // PATTERN RECOGNITION AND IMAGE ANALYSIS. – Vol. 34, Issue 3. – P. 751-756. – DOI: 10.1134/S1054661824700627.
- [**Shabanov et al, 2019**] Shabanov B., Sotnikov A., Palyukh B., Vetrov A., Alexandrova D. Expert System for Managing Policy of Technological Security in Uncertainty Conditions: Architectural, Algorithmic, and Computing Aspects // Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus) January 28-30, 2019.
- [**Lin et al, 2020**] Lin Shuyu, Clark Ronald, Birke Robert, Schonborn Sandro, Trigoni Niki, Roberts Stephen. Anomaly detection for time series using vae-lstm hybrid model. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – DOI: 10.1109/ICASSP40776.2020.9053558.
- [**Su et al, 2019**] Su Y., Zhao Y., Niu C., Liu R., Sun W., and Pei D. Robust anomaly detection for multivariate time series through stochastic recurrent neural network // In ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. – 2019. – P. 2828-2837.
- [**Yangdong et al, 2019**] Yangdong He and Jiabao Zhao. Temporal convolutional networks for anomaly detection in time series // Journal of Physics: Conference Series. – 2019. – 1213(4).
- [**Zekai et al, 2021**] Zekai Chen, Dingshuo Chen, Zixuan Yuan, Xiuzhen Cheng, and Xiao Zhang. Learning graph structures with transformer for multivariate time series anomaly detection in IoT. ArXiv, abs/2104.03466, 2021.
- [**Zhao et al, 2020**] Zhao H., Wang Y., Duan J., Huang C., Cao D., Tong Y., Xu B., Bai J., Tong J., and Zhang Q. Multivariate time-series anomaly detection via graph attention network // In 20th IEEE International Conference on Data Mining. – 2020. – P. 841-850.

УДК 004.77

doi: 10.15622/rcai.2025.090

ИНТЕРНЕТ БОЕВЫХ ВЕЩЕЙ В ВООРУЖЁННЫХ СИЛАХ

Г.П. Виноградов (*wgp272ng@mail.ru*)

НИИ «Центрпрограммсистем», Тверь

Статья посвящена актуальной проблеме – созданию Интернета боевых вещей. Рассмотрены предпосылки, обуславливающие актуальность. Рассмотрено состояние вопроса и результаты применения технологии «умного города» в вооруженных силах США. Показано, что реагирующие сенсорные сети определяют эффективность подобных технологий. Предложен вариант архитектуры реагирующей сенсорной сети. Представлен вариант построения элементов архитектуры и наиболее важных алгоритмов: позиционирования и отслеживания мобильности целей.

Ключевые слова: Интернет боевых вещей, реагирующая сенсорная сеть, управление, отслеживание, модель измерения.

Введение

Современные возможности обнаружения противника, высокоточное вооружение и высокая мобильность боевых систем требуют быстрого принятия решений. Одним из способов решения этой проблемы стало применение решений на базе Интернета вещей, названным **Интернетом боевых вещей** (англ. **Internet of Battle Things, IoBT**) [Winkler et al., 2008]. Это концепция сети передачи данных между физическими объектами различных родов войск («вещами»), оснащёнными встроенными средствами сбора, обработки данных и информационными технологиями для взаимодействия друг с другом с целью информирования командиров всех уровней о ситуационной обстановке и выработки эффективных решений. IoBT позволяет реализовать совершенно новый способ проведения военных операций, при котором все участники (техника, живая сила, штабы и т.д.) связаны единой информационной сетью для выполнения всего спектра основных и вспомогательных боевых задач.

1. Предпосылки использования Интернета вещей (IoT) в военной сфере

Роботизированная автоматизация процессов решения задач, основанных на правилах в вооруженных силах.

Датафикация – процесс, позволяющий сделать данные простыми для понимания и использования путем объединения процессов сбора, обработки, агрегирования и представления данных.

Цифровые двойники. Технология цифровых двойников позволяет создать виртуальные модели физических систем или процессов, *которые можно использовать для моделирования, анализа и оптимизации.*

Периферийные вычисления позволяют приблизить вычислительную мощность к источнику данных, когда критически важно время, необходимое для передачи и обработки данных.

Сети 5G работающие в диапазоне частот миллиметровых волн, что позволяет передавать данные с высокой скоростью.

Достижения в области микроэлектроники и сенсорных сетей и создание реагирующих беспроводных сенсорных сетей (РБСС), используемых для контроля физических параметров некоторой области и выработки воздействия на нее для получения желаемых состояний [Бородин и др., 2021].

2. Состояние вопроса с IoT в ВС

Лидерство по применению IoT в военных целях принадлежит США. С августа 2019 года в США начато изучение возможности применения инфраструктуры "умного" города на поле боя. Тестируется глобальная сеть дальнего радиуса действия (LoRaWAN) при создании и модернизации «умных» военных баз [Chen et al., 2001].

Разработано мобильное приложение АТАК (Android Tactical Assault Kit), позволяющее накапливать данные в режиме real-time и накладывать их на Google Maps. В зоне боевых действий это решение используется для передачи данных от наводчика на цель пилоту самолета или оператору БПЛА.

Попытки американских военных развить глобальную сеть наталкиваются на проблемы уязвимости IoT-сетей, когда они становятся объектом атаки [Rippin et al., 2012].

3. Кибер-физические системы – основа военного Интернета вещей

В военном применении Интернета вещей одной из ключевых стала тема кибер-физических систем (КБФ). Суть их состоит в том, что они соединяют физические процессы при выполнении боевых задач, требующие управления в режиме реального времени, с программно-электронными системами и информационными технологиями. Тактические кибер-

физические системы, обеспечивающие ситуационную осведомленность во всем стратегическом спектре боевых операций и принятие решений в условиях жестких временных ограничений, основа для превосходства над противником. Обработка максимально возможной совокупности мультимедийных и мультиспектральных данных в реальном времени позволяет реализовать адаптивное планирование боевых задач, вести целеуказание, оценку боевых потерь, оценивать надежность распределенных мобильных средств связи и т.п. на уровне зоны проведения боевых операций. Для этого мобильное боевое информационное оборудование, тактические сетевые средства связи и спешенные оперативные средства управления, интегрируются с регулируемыми, устойчивыми и прозрачными для пользователя глобальными сетями. Основой КБФ являются реагирующие беспроводные сенсорные сети (РБСС), обеспечивающие сбор, обработку исходной информации для военных приложений и реализацию решений.

4. Примерная архитектура РБСС военного назначения

В настоящее время РБСС рассматривается как одна из технологических основ окружающего интеллекта [Бородин и др., 2021], [Shit et al., 2018]. Вариант архитектуры РБСС для военных приложений может быть построен с использованием сенсорных узлов с радиосвязью ближнего действия и беспроводных шлюзов с беспроводной связью на большие расстояния, что обеспечивает большую гибкость и расширяемость в возможных видах операций от небольшого одиночного кластера сенсорных узлов до множества соединений на площади до 20 км² [Виноградов и др., 2021]. Сенсорные узлы находятся на первом уровне иерархии, где они выполняют основные операции мониторинга и воздействия на среду. Они оснащены встроенными преобразователями, такими как акустические, сейсмические, пассивные инфракрасные (PIR), магнитные и пьезоэлектрические и др. для обнаружения событий, представляющих интерес. Каждый узел датчика в сети действует как маршрутизатор, пересылая пакеты данных на соседние узлы. Они формируют сеть «на лету» и поддерживают единый радиоинтерфейс для двунаправленной связи между узлами с датчиками и узлом слияния.

На втором уровне иерархии узлы слияния обеспечивают более сложные функции, такие как синхронизация баз данных, формирование кластеров, логики работы приложения и управление. Узлы слияния получают информационные запросы от пользователей, отслеживают команды, отвечают на запросы, формируют задачи узлам датчиков, собирают информацию и сохраняют историю событий, произошедших в зоне ответственности, охватываемой узлом слияния. Узлы слияния могут выступать в качестве исполнительных механизмов в сети, например, для запуска ракет, открытие огня и пр. Узлы слияния на втором уровне иерархии также об-

разуют специальную. Они могут быть оснащены несколькими радио интерфейсами для связи с сенсорными узлами и другими узлами слияния, а также для передачи данных на большие расстояния на узел управления. Сенсорные узлы и узлы слияния образуют кластеры, которые соединяются между собой через головные узлы для создания автоматической сенсорной системы.

Узел управления на третьем уровне обеспечивает оперативный контроль и управление системой. Узел слияния на этом уровне использует канал дальней связи для связи с узлом управления. Авторизованные пользователи могут в соответствии с правом доступа получать доступ к системе с узлов слияния данных на уровне 2 или с узла управления на уровне 3.

5. Сенсорный узел

Сенсорный узел (модуль) используется как для мониторинга среды, так и для локального или в составе группы воздействия на среду. В последнем случае он называется реагирующим. Его архитектура содержит пять основных компонент [Виноградов и др., 2021]: 1) сенсорная подсистема, включающая датчики и радар, для контроля состояния внешней среды с соответствующими преобразователями; 2) подсистема обработки, включающая микроконтроллеры и память для хранения данных; 3) радиопередатчик для беспроводной передачи данных; 4) устройство электропитания; 5) подсистема исполнения решений. Сенсорные узлы образуют сенсорное поле, как правило, они находятся в спящем состоянии. Когда происходит событие, или по расписанию в определенное время, узлы «просыпаются», самоорганизуются в сеть и передают информацию в центр сбора данных. Область применения РБСС – это, как правило, неопределенные и плохо формализуемые среды, поэтому достичь желаемой эффективности подобных комплексов возможно, главным образом, путем совершенствования интеллектуальной составляющей их системы управления в целом и отдельным узлом в частности.

6. Система управления узлом, использующая паттерны

Жесткие требования в ряде предметных областей к времени реагирования, объему памяти, быстродействию микрочипа и уровню энергопотребления предполагают использование сравнительно простых алгоритмов, построенных путем описания лучшего опыта человека при выполнении функциональных задач. В работах [Виноградов и др., 2021], [Brooks et al., 2002] показано, что такое описание можно реализовать, если выявить классы типовых ситуаций и успешные способы действия в реальных условиях для формирования моделей поведения (паттернов). В работах [Xu et al., 2004], [Виноградов и др., 2020] дана следующая обобщенная логическая схема описания паттерна

$$\begin{aligned}
& \text{Имя паттерна:} \\
& \text{так как [мотивы } M] \\
& \text{поскольку [цели } G] \\
& \text{если [предусловия } U'] \\
& \text{то способ действия } r_q(t) \\
& \text{из-за чего [постусловие } U''] \\
& \dots \\
& \text{есть альтернатива [} r_p(t) \text{]}
\end{aligned} \tag{1}$$

В этой модели все составляющие, кроме логических связей, могут представлять собой языковые конструкции на естественном или естественно-профессиональном языке. Модель (1) – типовая естественно-языковая модель паттерна. Формальная модель паттерна поведения в типовой ситуации приведена в [Toor et al., 2017], [Vinogradov, 2020]. Там же показано, что модель предметной области и модель принятия решений в условиях дефицита времени можно формализовать с помощью нечетких продукционных сетей. Мотивы в модели (1) – это причины, определяемые потребностями и/или обязательствами. Они побуждают агента к активности, в рамках которой потребовался паттерн. Описание мотивов и целей в модели (1) представляет контекст повторного применения паттерна с позиции необходимости и возможности его применения (потребности → мотивы → цели). Контекст определяется «ожидаемыми» постусловиями. Постусловия связаны с целями, затребовавшими паттерн, но отражают их в форме возможного состояния, к которому приведет применение паттерна. Способ действия представляет собой код как естественно-языковое описание схемы действия, реализация которой приводит к выполнению паттерна. Такое описание имеет вид методик, реализованных на языке программирования. Набор моделей или паттернов поведения, образуют опыт или базу знаний его носителя. Модель индивидуального поведения автономного узла предполагает формирование базы паттернов на основе экспериментального опыта, что позволяет реализовать эволюцию «кооперативного интеллекта» благодаря использованию искусственного когнитивного процесса, аналогичного тому, что имеет место у естественных существей. Отметим, что эта возможность отсутствует в системах на основе знаний.

Для этого предложено состояние ситуации описывать ситуационным вектором $x_i, i = \overline{1, n}$, каждая координата которого лингвистическая переменная x_i с множеством термов $A_i = \{a_i^k, k = \overline{1, K^i}\}$. Пусть определен набор классов реализаций ситуационного вектора, при котором есть паттерн поведения с успешным разрешением любой ситуации класса (кластера). Пусть накоплено множество $d_j, j = \overline{1, p}$ паттернов, каждый из которых связан с множеством кластеров ситуаций, для разрешения которых он был

выбран. Предложено составить матрицу соответствия кластеров ситуаций и располагаемым множеством паттернов поведения. В соответствии с (1) ядром системы управления становится машина нечеткого вывода с набором нечетких продукционных правил (база знаний) и алгоритмами перевода численных значений в лингвистическую форму и обратно. Как результат можно реализовать любую зависимость между входными и выходными переменными и, основное, организовать пересылку как новых шаблонов правил, так и параметров функций принадлежности. То есть решить задачу самоорганизации с участием внешнего наблюдателя. В докладе обсуждаются соответствующие алгоритмы структурной и параметрической адаптации нечеткой базы правил.

Архитектура интеллектуальной системы управления РБСС имеет иерархическую структуру. Верхний уровень соответствует таким свойствам, как выживание, безопасность, выполнение обязательств согласно миссии, накопление и корректировка базы знаний в виде эффективных паттернов поведения. Его основные функции: 1) расчет текущих показателей удельной ценности по результатам и эффективности в момент t [Vinogradov, 2020]; 2) расчет и реализацию способа действия (поведения) в момент t согласно заданному паттерну поведения; 3) мониторинг результатов реализации паттерна поведения. Оператор отрабатывает паттерны поведения при выполнении миссии и выполняет анализ выполнения миссии. Узел рассчитывает последовательности состояний $y(t)$, реализует задачи миссии и выполняет расчет оценки удельной ценности и эффективности отработки фактических ситуаций [Виноградов и др., 2021]. Такой подход к моделированию поведения узлов в сети на основе паттернов аналогичен подходу в прагматической эпистемологии, согласно которому знания рассматриваются как множество моделей, каждая из которых представляет собой описание поведения при решении определенного класса задач. Критерием их выбора становится ожидаемая удельная ценность возможного результата [Виноградов и др., 2021].

7. Система локализации, классификации, отслеживания и поражения объектов вторжения в зону РБСС

Практическая реализация задач локализации, классификации и отслеживания объектов вторжения в зону РБСС предполагает организацию взаимодействия алгоритмов отслеживания с традиционными алгоритмами инициализации/маршрутизации в РБСС. Алгоритмы обнаружения формируют сегменты данных на основе измерений сенсоров. Предлагается перемещение объектов в беспроводных сенсорных сетях записывать в системные журналы узлов. В предлагаемой сетевой модели каждый узел может записывать событие появление объекта вместе со временем прибытия на каждый узел. Для сбора журнала перемещений несколько мощных сен-

сорных узлов, оснащенных устройствами хранения данных, предложено развернуть на внешней стороне сети для получения журнала каждого объекта, выходящего из сети. Сенсоры отслеживают амплитуду (мощность) для обнаружения событий в каждый момент времени в течение определенного временного интервала. События фиксируются тогда, когда амплитуда превышает пороговое значение. Пороговое значение динамически обновляется на основе статистики фоновых шумов для снижения частоты ложных срабатываний. Как только узел обнаруживает событие (например, присутствие движущегося транспортного средства), в нем формируется временной ряд, соответствующий событию. Сегмент временного ряда формируется на интервале, когда энергия сначала превышает пороговое значение (начало события), а затем падает ниже него (завершение события). На рис. 1 приведена блок-схема потока данных на узле согласно описанному подходу к распределенному отслеживанию.

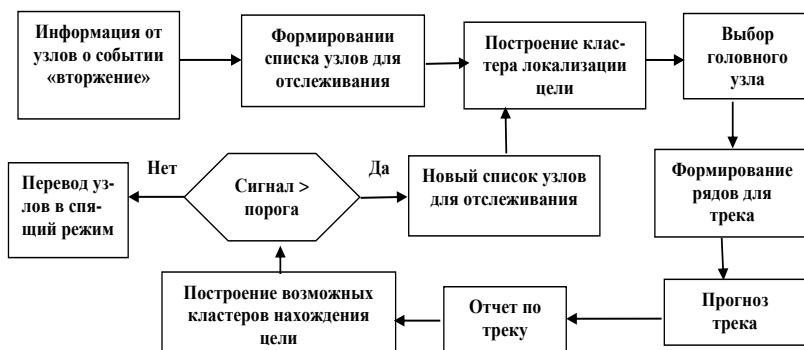


Рис. 1. Блок-схема потока данных

Несколько потоков выполняются одновременно, а система слежения представляет собой одноранговую сеть меньшей размерности. Все узлы выполняют одну и ту же логику. Это подход позволяет значительно повысить энергоэффективность сети.

1) На первом этапе активизируются узлы, фиксирующие событие «вторжение». Порог доверия устанавливается таким образом, чтобы предотвратить ложное срабатывание. Создается для каждого узла новая запись трека. Предварительно атрибуты узла сети ориентированы на местоположение.

2) Информация о маршруте цели от узла-кандидата непрерывно принимается и сохраняется. Определяется тип цели каждым узлом.

3) Формируется кластер узлов, локализующих цель.

4) Выполняется выбор головного узла кластера. Оценки параметров являются входными данными для алгоритма отслеживания. Информация агрегируется с треком, который наилучшим образом соответствует текущим данным. Очереди, в которых местоположение цели и узла близки, рассматриваются в первую очередь. Это соответствует пику сигнала временного ряда от узла кластера. Атрибуты цели и данные временного ряда прошлых измерений используются для прогнозирования.

5) Оценивается прогнозный трек на основе последней информации и обновляется список узлов, участвующих в отслеживании.

6) Формируется отчет по результатам отслеживания цели.

7) Передается отчет в пункт управления. Узлы с низким уровнем сигнала переводятся в спящий режим.

Оценка локальных параметров выполняется с использованием подхода, ориентированного на местоположение. Алгоритм позволяет формировать кластер слежения динамически в пределах ограниченного пространственно-временного окна. Головной узел выбирается из списка узлов-кандидатов с самым сильным сигналом в кластере. Линейная регрессия с использованием тригонометрии расположения узлов используется для оценки положения цели, скорости и курса.

Подход предполагает решение ряда проблем, среди которых следует выделить две основные. Первая – это разработка эффективных методов обмена информацией между локальными узлами в зоне вторжения, вторая – организация совместной обработки сигналов группой узлов на основе собранной информации о состоянии среды в зоне их ответственности в результате наступления событий. Рассмотрены подходы, лежащие в основе алгоритмов обнаружения, локализации и отслеживания, а также основные аспекты их реализации. Предлагаемые решения учитывают ограничения, связанные с возможностями локальных узлов, сетью в целом и маршрутизацией. Источником данных для предлагаемых алгоритмов являются сигналы от сенсоров, у которых мощность сигнала имеет выраженный максимум в зависимости от расстояния от цели и до узла сети. Полученные результаты распространяются на проблему отслеживания множества объектов, что предполагает оценку применимости методов идентификации и классификации в условиях, когда наблюдается наложение воспринимаемых сигналов сенсорами разными объектами. Обсуждаются алгоритмы для решения таких задач.

Важнейшей характеристикой приложений на базе РБСС является длительность жизненного цикла [Бородин и др., 2021], [Mikusz et al., 2014], [Xu et al., 2004] определяемая возможностями энергетической системы сенсорной сети [Shit et al., 2018]. Поток вторжений быстро движущихся целей в сенсорную сеть оказывает наиболее сильное воздействие на длительность ее жизненного цикла [Shit et al., 2018]. Это ставит перед разра-

ботчиками ряд проблем. Возможные подходы и варианты их решения рассматривались в [Chen et al., 2001]. Повышение живучести системы защиты предложено реализовать путем наделения ее набором следующих функций: 1) *Распределенная обработка (CSP)*. 2) *Обработка данных по событию*. 3) *Агрегирование информации*. 4) *Динамическая кластеризация* [Виноградов и др., 2021].

8. Входные данные для обнаружения и локализация объектов РБСС

Объект проникновения в процессе движения генерирует сигналы. Величина уровня мощности воспринимаемого сенсорами сигнала (англ. *RSSI* – *Received Signal Strength Indicator*) определяется расстоянием узел-цель. Спектр сигналов будет иметь максимум при прохождении цели над узлом или в непосредственной близости, и он может рассматриваться как индивидуальная характеристика цели или сигнатура, которую можно использовать при определении типа цели [Suryadevara et al., 2015] [Dan et al., 2015]. Событие «обнаружение цели» будет возникать тогда, когда выход сенсора узла превышает некоторый порог, величина которого регулируется так, чтобы величина частоты ложной тревоги не превышала некоторую настраиваемую норму. Значение сигнала при перемещении объекта воспринимается широкоэвентально, то есть все узлы, в радиусе восприятия которых находится цель, ее «слышат». Узлы, «слышащие» цель осуществляют считывание значений спектра сигнала цели в определенные моменты времени при ее движении. Тем самым формируется пространственная и временная выборка поля фактической сигнатуры цели. Характер изменения поля пространственно-временной сигнатуры определяет требуемую частоту дискретизации в пространстве-времени и количество активизируемых узлов. В работах [Brooks et al., 2002] [Dan et al., 2015] для обеспечения отслеживания цели и эффективной локальной обработки данных в сенсорной сети предложено разделить область проникновения на пространственно-временные ячейки (окна). Их размер зависит от скорости, направления движения цели и от показателя затухания прохождения сигнала в среде. Размер ячейки должен приблизительно соответствовать области, внутри которой в течение некоторого времени спектр считываемой сигнатуры остается практически постоянным, а его падение на границах меньше некоторого порогового значения. Это предполагает введение пространственно-временных координат в анализ и динамическую коррекцию размера пространственно-временных ячеек на основе прогнозируемых местоположений, типов целей и показателей их движения.

Заключение

Для повышения возможностей применения РБСС в военных приложениях исследования и разработки должны быть направлены на создание: отечественных одно кристаллических процессоров, разработку алгоритмов идентификации нескольких одновременных событий; классификации объектов и событий в задачах обнаружения; миниатюризации и интеграции различных типов датчиков, повышение их надежности; форматов и стандартов для выходов датчиков и коммуникаций. Востребованными проблемами являются: разработка алгоритмов Sensor Data Mining на основе методов искусственного интеллекта и интеллектуального анализа данных, синтеза и генерации знаний на основе бортовой осведомленности и онтологий. Для повышения жизненного цикла РБСС требуется разработка новых источников питания, и энергоэффективных протоколов коммуникации.

Список литературы

- [Бородин и др., 2021] Бородин А.С., Волков А.Н., Мутханна А.С.А., Кучерявый А.Е. Искусственный интеллект в сетях связи пятого и последующих поколений. Электросвязь. – 2021. – № 1. – С. 17-22.
- [Виноградов и др., 2021] Виноградов Г.П., Емцев А.С., Федотов И.С. Беспроводные сенсорные сети в защищаемых зонах // Известия ЮФУ. Технические науки. – 2021. – № 1. – С. 19-30.
- [Виноградов и др., 2021] Виноградов Г.П., Конюхов И.А., Шепелев Г.А. Подход к проектированию программного обеспечения систем управления искусственными сущностями // Программные продукты и системы. – 2021. – Т. 34, № 1. – С. 005-018. – DOI: 10.15827/0236-235X.133.005-018.
- [Виноградов и др., 2020] Виноградов Г.П., Прохоров А.А., Шепелев Г.А. Паттерны в системах управления автономными робототехническими комплексами // Виноградов Г.П., Прохоров А.А., Шепелев Г.А. Паттерны в системах управления автономными системами // Мягкие измерения и вычисления. – 2020. – № 12.
- [Brooks et al., 2002] Brooks R. and Griffin C. Traffic model evaluation of ad hoc target tracking algorithms // to be published in the International Journal of High-Performance Computing Applications, Special Issue on Sensor Networks, 2002.
- [Chen et al., 2001] Chen J.C., Hudson R.E., and Yao K. A Maximum likelihood parametric approach to source localization // Proc. ICASSP'2001, Salt Lake City, UT, 2001. – P. 1043-1046.
- [Dan et al., 2015] Dan Li, Kerry D. Wong, Yu H. Hu, Akbar M. Sayeed. Detection, Classification and Tracking of Targets in Distributed Sensor Networks // Источник: <http://www.ecce.wisc.edu/~sensit>. Дата обращения: 09.01.2021.
- [Hsua et al., 2012] Hsua J.M., Chenb C.C., Li C.C. POOT: An efficient object tracking strategy based on short-term optimistic predictions for face-structured sensor networks, Elsevier International Journal of Computers and Mathematics with Applications. – 2012. – Vol. 63, Issue 2. – P. 391-406.

- [**Mikusz et al., 2014**] Martin Mikusz. Towards an understanding of cyber-physical systems as industrial software-product-service systems // Conference paper in Procedia CIRP. – 2014.
- [**Rippin et al., 2012**] Rippin B. Pearls of wisdom: wireless networks of miniaturized unattended ground sensors // SPIE, Orlando, USA, 2012. – 8388-17
- [**Shit et al., 2018**] Shit R.C., Sharma S., Puthal D., Zomaya A.Y. Location of things (lot): a review and taxonomy of sensors localization in Iot infrastructure // IEEE Communications Surveys & Tutorials. – 2018. – Vol. 20, No. 3. – P. 2028-2061. – Look at: Look at: Google Scholar.
- [**Shit et al., 2019**] Shit R.C., Sharma S., Puthal D. et al. Ubiquitous localization (UbiLoc): a survey and taxonomy on device free localization for smart world // IEEE Communications Surveys & Tutorials. – 2019. – Vol. 21, No. 4. – P. 3532-3564. – Look at: Google Scholar.
- [**Suryadevara et al., 2015**] Suryadevara N.K., Mukhopadhyay S.C., Kelly S.D.T. and Gill S.P.S. WSN-based smart sensors and actuator for power management in intelligent buildings // IEEE/ASME Transactions on Mechatronics. – 2015. – Vol. 20, No. 2. – P. 564-571. – Look at: Google Scholar.
- [**Toor et al., 2017**] Toor A.S. and Jain A.K. A survey on wireless network simulators // Bulletin of Electrical Engineering and Informatics. – 2017. – Vol. 6, No. 1. – P. 62-69. – Look at: Google Scholar.
- [**Xu et al., 2004**] Xu Yingqi, Winter Julian, Lee Wang-Chien. Prediction-based strategies for energy saving in object tracking sensor networks // Proceedings - 2004 IEEE International Conference on Mobile Data Management. – 2004. – P. 346-357. – 10.1109/MDM.2004.1263084.
- [**Winkler et al., 2008**] Winkler M., Tuchs K.-D., Hughes K., and Barclay G. Theoretical and practical aspects of military wireless sensor networks // in Journal of Telecommunications and Information Technology. – 2/2008. – P. 37-45.
- [**Vinogradov, 2020**] Vinogradov G.P. Patterns in Intelligent Systems. Russian Advances in Fuzzy Systems and Soft Computing: selected contributions to the 8-th International Conference on Fuzzy Systems, Soft Computing and Intelligent Technologies (FSSCIT-2020), June 29 – July 1, 2020, Smolensk, Russia. CEUR Workshop Proceedings, 2020, 2782. – P. 208-216.

ГИБРИДНЫЙ МЕТОД ИЗВЛЕЧЕНИЯ ПРЕЦЕДЕНТОВ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ЛАЗЕРНОМ АДДИТИВНОМ ПРОИЗВОДСТВЕ¹

В.В. Грибова (*gribova@iacp.dvo.ru*)
А.И. Никитин (*anikitin@iacp.dvo.ru*)
В.А. Тимченко (*vadim@iacp.dvo.ru*)

Институт автоматики и процессов управления ДВО РАН,
Владивосток

В рамках реализации прецедентного подхода для поддержки принятия решений в лазерном аддитивном производстве (ЛАП) предложен гибридный метод извлечения прецедентов, сочетающий предметные экспертные знания и алгоритм k-ближайших соседей. Ключевая особенность метода заключается в разработке мер оценки сходства разнородных параметров технических заданий на выполнение технологических операций ЛАП, максимально учитывающих особенности предметной области. Предлагаемый метод может быть использован при разработке интеллектуальных ассистентов для технологов в сфере ЛАП по технологии направленного энерговклада.

Ключевые слова: поддержка принятия решений, рассуждения по прецедентам, извлечение прецедентов на основе знаний, меры сходства случаев, лазерная обработка материалов.

Введение

Одной из наиболее сложных и нерешенных технических проблем внедрения лазерных технологий в производственные процессы, как в России, так и за рубежом, является сложность управления качеством синтезируемого металлического материала. В связи с этим, разработка средств для оказания инженерам-технологам интеллектуальной помощи по настройке режимов работы технологического оборудования, реализующего процессы лазерной аддитивной обработки металлических деталей является крайне актуальной задачей [Ethiraj et al., 2025].

¹ Работа выполнена в рамках государственного задания ИАПУ ДВО РАН (тема № FFWF-2025-0004).

В настоящее время поддержка принятия решений на основе прецедентов (рассуждений по аналогии) активно применяется во многих предметных областях [Varshavskii et al., 2010], [Варшавский и др., 2013], [Lang et al., 2021]. Актуальность применения методов на основе прецедентов обусловлена, в том числе, и тем, что они эффективны при работе с небольшими наборами данных, в то время как использование других методов (на основе, например, знаний и/или машинного обучения) может привести к неточным и даже ошибочным результатам или вовсе не позволяет получить решение ввиду недостатка знаний и/или данных [Малых, 2019], [Полкова и др., 2022].

Методы рассуждения по прецедентам положены в основу ряда разработанных систем в области медицины, используемых для помощи врачам при назначении лечения [Choudhury et al., 2016], [Юдин и др., 2017], в производстве [Трофимов, 2020], в частности, для поддержки принятия решений в выборе подходящих технологических процессов [Mabkhot et al., 2019] и в области лазерной обработки материалов [Майоров и др., 2009] и др.

Наиболее известными методами рассуждения по аналогии (извлечения прецедентов) являются методы k -ближайших соседей (k -NN), извлечения прецедентов на основе деревьев решений и извлечения прецедентов на основе знаний [Крылов, 2018].

Метод k -NN считается наиболее простым и универсальным. Он используется во многих системах поддержки принятия решений на основе прецедентного подхода. К недостаткам данного метода можно отнести сложность выбора метрики для определения степени сходства, неэффективность при работе с неполными данными, а также ощутимая зависимость от размера базы прецедентов (БП) [Крылов, 2018].

Метод извлечения прецедентов на основе деревьев решений обычно используется для больших баз прецедентов, так как основная часть работы по извлечению прецедентов выполняется заранее на этапе построения дерева решений, что значительно сокращает время поиска решения. Данный подход в сочетании с методом k -NN применяется в системе оценки безопасности программного обеспечения, используемого в сфере железнодорожных перевозок [Hadj-Mabrouk, 2020]. Существенным недостатком подхода на основе деревьев решений является то, что деревья дают полезные результаты только в случае независимых признаков. Имеет место и проблема значимости: дерево дробит данные на большое количество частных случаев. Чем их больше, тем меньше обучающих прецедентов попадает в каждый такой частный случай [Полин и др., 2020].

Метод извлечения прецедентов на основе знаний позволяет учитывать знания предметных экспертов и может успешно применяться совместно с другими методами извлечения прецедентов, особенно когда БП имеют большой объем.

Одной из главных проблем прецедентного подхода является выбор соответствующей метрики (меры подобия) для сравнения прецедентов. В каждом конкретном случае этот выбор производится по-разному в зависимости от целей лица, принимающего решения, а также самих данных. Также анализ существующих решений на основе прецедентов позволил сделать вывод, что зачастую описание прецедентов и получаемое решение выполнены в математических терминах, которые непонятны предметным специалистам. Авторы считают наиболее перспективным применение гибридного подхода, сочетающего различные методы извлечения прецедентов [Khan et al., 2019].

Настоящая работа ориентирована на поиск похожих прецедентов при проектировании инженером-технологом пригодных для конкретного практического применения режимов работы робототехнического оборудования – на этапе планирования технологических операций лазерной аддитивной обработки материалов (лазерного аддитивного производства, ЛАП) и подготовки управляющих оборудованием программ для их реализации.

Предлагается оригинальный комбинированный (гибридный) метод рассуждения по аналогии (поиска релевантных прецедентов). Гибридность метода состоит в том, что степень похожести определяется на основе предметных экспертных знаний о мерах близости элементов (параметров) формализованных производственных технических заданий на выполнение технологических операций в сочетании с используемым для интеллектуального анализа данных классическим методом k -NN. В методе используется информация из постоянно пополняемой базы структурированных формализованных протоколов технологических операций ЛАП, а также обновляемых нормативно-справочных баз данных (мастер данных), содержащих информацию о ключевых параметрах обрабатываемых (расходных) материалов ЛАП. Экспертные знания и справочные базы данных необходимы для определения т.н. «локальных» мер подобия/близости (*LocalSim*) разнородных элементов технического задания.

1. Общие сведения и постановка задачи

В рамках решаемой задачи прецедент представляет собой часть протокола выполненной технологической операции (ТО), куда не входят: общие сведения о ТО (название ТО, номер протокола, а также срок, цель и место выполнения ТО), комментарии (примечания) технолога, условия окружающей среды при выполнении ТО. Описанием ситуации является техническое задание (ТЗ) на выполнение ТО, решением – выбранные технологом средства и параметры выполнения ТО. Результатом применения решения является результат выполнения ТО: его описание, а также оценка – «соответствует требованиям ТЗ», «не соответствует требованиям ТЗ, но оценивается как положительный», «не соответствует требованиям ТЗ и оценивается как отрицательный».

Извлечение прецедентов основано на знаниях специалистов по материаловедению и лазерной роботизированной обработке материалов о мерах близости элементов формализованных производственных ТЗ на выполнение ТО, пополняемых нормативно-справочных баз данных, содержащих информацию о ключевых параметрах обрабатываемых (расходных) материалов ЛАП, а также несколько модифицированной формуле L_1 -расстояния (манхэттенского расстояния).

Экспертные знания включают в себя информацию о пороговых значениях степени сходства элементов (параметров) ТЗ и, наряду с нормативно-справочными базами данных, используются для вычисления меры их подобия – $LocalSim_i$ ($i = 1, \dots, n$; n – количество параметров ТЗ). К ключевым параметрам ТЗ относятся: материал объекта обработки (металлической детали или подложки), масса объекта обработки, его геометрические характеристики, материал для выполнения ТО (металлический порошок или металлическая проволока), технологический газ, используемый для создания защитной среды, требования к геометрическим характеристикам, пористости, элементному составу, микроструктуре результата ТО.

К обрабатываемым и расходным материалам ЛАП относятся металл-порошковые и металлопроволочные материалы на основе различных сплавов, собственно металлические сплавы, представленные своим элементным (химическим) составом, микроструктурой, набором теплофизических и механических свойств, а также технологические газы, используемые в качестве защитного (защитной среды), транспортирующего и обжимного. Также в базах представлена информация о характерных для металлов (металлических покрытий) дефектах. Описания объектов в соответствующих нормативно-справочных базах включают описания их ключевых характеристик и возможных значений (качественных, количественных и интервальных). Кроме того, в этих базах данных имеет место использование синонимии, что позволяет учитывать различные названия одних и тех же характеристик (а также их значений) элементов ТЗ [Грибова и др., 2024]. Отметим, что все базы данных формируются по своим онтологиям, четко задающим их структуру, ограничения на возможное содержание и правила интерпретации. Более детальное рассмотрение данного аспекта также представлено в работе [Грибова и др., 2024].

Формой представления протокола ТО является также описанный в терминах онтологии предметной области оргграф. Соответственно, ТЗ представляется подграфом в оргграфе протокола ТО. Таким образом, описание прецедентов и получаемое решение представлены в онтологических терминах и понятны предметным специалистам.

На вход метода поступает новое техническое задание на выполнение технологической операции. Требуется найти похожие технические задания из базы прецедентов (структурированного архива протоколов техно-

логических операций лазерной обработки), ранжировать их по степени сходства, а также обеспечить пользователю демонстрацию сходных элементов из текущего (нового) технического задания и ранжированного списка.

При этом поиск прецедентов осуществляется не во всей БП, а среди протоколов ТО, относящихся, во-первых, к тому же типу выполняемой ТО, что и новое ТЗ, а во-вторых, в которых обрабатываемые и/или расходные материалы совпадают с таковыми в новом ТЗ.

Поскольку в данном случае мы имеем дело с предметной областью, в которой структура ТЗ в значительной степени неоднородна (состоит из разнотипных перечисленных выше параметров), а количество анализируемых параметров в нем фиксировано и невелико, правила вычисления мер сходства (близости) для отдельных параметров ТЗ – $LocalSim_i$ – определяются индивидуально, в зависимости от конкретного параметра ТЗ.

Таким образом, сначала выполняется расчет значений $LocalSim_i$ отдельно для каждой пары соответствующих параметров ТЗ, а затем – расчет интегральной меры сходства ($GlobalSim$) для ТЗ_{new} (описание нового ТЗ/нового случая) и ТЗ_{case} (описание ТЗ в рассматриваемом прецеденте из БП). Значение $GlobalSim$ рассчитывается как усредненное значение (среднее арифметическое) для всех $LocalSim_i$ ($i = 1, \dots, n$). Такой подход позволяет максимально гибко и корректно (с точки зрения семантики предметной области) оценивать похожесть технических заданий с разнородными элементами (параметрами).

2. Гибридный метод извлечения прецедентов

2.1. Описание метода

ТЗ на выполнение ТО представлено следующим набором параметров: (*Объект обработки_Материал*, *Объект обработки_Масса*, *Объект обработки_Геометрические характеристики*, *Материал для выполнения ТО*, *Технологический газ*, *Требования к результату операции_Геометрические характеристики*, *Требования к результату операции_Дефекты наплавленного материала*, *Пористость*, *Требования к результату операции_Элементный состав*). Соответственно, каждый из этих параметров представляется подграфом в иерархическом онтологическом орграфе ТЗ. $LocalSim_i$ для пары соответствующих (однотипных) параметров ТЗ из набора – параметр ТЗ _{i} ($i = 1, \dots, 8$) – определяется как отображение (функция):

$LocalSim_i$: параметр ТЗ _{i} \times параметр ТЗ _{i} $\rightarrow \{0, 5, 10\}$.

Таким образом, область возможных значений для $LocalSim_i$ представляет собой множество, состоящее из трех чисел, кодирующих следующую шкалу качественных значений: 10, означающее максимальное расстояние между параметрами («полное различие», цвет – «красный»); 0 – минимальное расстояние («полное сходство», цвет – «зеленый»); 5 – усреднен-

ное расстояние («частичное сходство», цвет – «оранжевый»). Данная шкала была согласована и признана (в первом приближении) адекватной специалистами Центра инженерных разработок (ЦИР) ИАПУ ДВО РАН.

Вход:

- TZ_{new} – ссылка на начальную вершину орграфа, представляющего собой описание нового ТЗ (новой ситуации).
- Ссылка на вершину орграфа (представляющего структурированную БП), под которой находятся протоколы ТО, среди ТЗ которых нужно искать похожие на TZ_{new} , т.е., под этой вершиной протоколы хранятся как список ($TO_{case\ 1}, \dots, TO_{case\ m}$), элементы которого нужно последовательно перебрать.

Выход:

- Множество SC (возможно, пустое), включающее не более k протоколов ТО из ($TO_{case\ 1}, \dots, TO_{case\ m}$), где $0 \leq k \leq m$, технические задания которых преодолели заданный порог похожести $H\%$ на TZ_{new} .

Описание работы.

$SC' = \emptyset$.

В цикле по элементам в списке протоколов технологических операций – $TO_{case\ 1}, \dots, TO_{case\ m}$ из БП выполнять следующие шаги 1–3.

Шаг 1.

Из начальной вершины орграфа, представляющего описание протокола $TO_{case\ j}$ ($j = 1, \dots, m$), перейти к вершине, являющейся начальной вершиной подграфа орграфа протокола, описывающего ТЗ в этом протоколе – $TZ_{case\ j}$.

Шаг 2.

Обход (одновременно/параллельно) двух орграфов, начиная с вершин TZ_{new} и $TZ_{case\ j}$. В процессе обхода:

Для соответствующих пар параметров ТЗ – параметр $TZ_{new\ i}$ и параметр $TZ_{case\ j\ i}$ вычисляется значение их $LocalSim_i$ (параметр $TZ_{new\ i}$, параметр $TZ_{case\ j\ i}$) ($i = 1, \dots, 8$).

По отсутствующим значениям параметров:

- если значение параметра присутствует в $TZ_{case\ j}$, но отсутствует в TZ_{new} , то по ним выставляется частичное сходство: $LocalSim_i$ ($null$, параметр $TZ_{case\ j\ i}$) = 5;
- если значение параметра отсутствует в $TZ_{case\ j}$, но присутствует в TZ_{new} , то по ним выставляется максимальное отличие: $LocalSim_i$ (параметр $TZ_{new\ i}$, $null$) = 10;
- если значение параметра отсутствует как в $TZ_{case\ j}$, так и в TZ_{new} , то по ним не выставляется никакого значения: $LocalSim_i$ ($null$, $null$) не будет учитываться в качестве слагаемого в формуле Шага 3.

Далее, в качестве примера опишем правила вычисления *LocalSim* для такого параметра ТЗ как «Объект обработки_Масса». Здесь на основе знаний и опыта экспертов предметной области были выделены следующие пять интервалов, в пределах которых массы считаются полностью похожими (между ними расстояние 0):

- 1 – [0.01 ... 0.1] кг
- 2 – (0.1 ... 1) кг
- 3 – [1 ... 10] кг
- 4 – (10 ... 100] кг
- 5 – (100 ... 400] кг

Объекты обработки массой свыше 400 кг не рассматривались. Далее сначала каждый из этих интервалов делится пополам, а затем на каждом $(i+1)$ -ом интервале от его левой границы откладывается отрезок, равный половине i -го интервала ($i = 1, \dots, 4$), образуя два подинтервала (рис. 1).

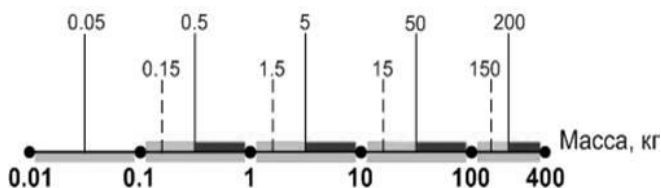


Рис. 1. Пороговые значения (интервалы) для определения меры сходства масс объектов обработки (подложки или детали)

В соответствии с выполненным разбиением имеем следующие случаи:

1. если значение массы m_1 (масса объекта обработки из нового случая) и значение массы m_2 (масса объекта обработки из прецедента) принадлежат одному и тому же интервалу (из пяти исходных), то $LocalSim(m_1, m_2) = 0$;

2. если значение массы m_1 и значение массы m_2 принадлежат несмежным интервалам, то $LocalSim(m_1, m_2) = 10$;

3. если значение массы m_1 и значение массы m_2 принадлежат смежным интервалам, то возможны три ситуации (без ограничения общности считаем, что значение m_1 принадлежит i -му интервалу, а значение m_2 – $(i+1)$ -му, который разбит на три подинтервала так, как описано выше):

- 3.1. если значение m_2 попадает в 1-ый подинтервал $(i+1)$ -го интервала, то $LocalSim(m_1, m_2) = 0$;
- 3.2. если значение m_2 попадает во 2-ой подинтервал $(i+1)$ -го интервала, то $LocalSim(m_1, m_2) = 5$;
- 3.3. если значение m_2 попадает во 3-ий подинтервал $(i+1)$ -го интервала, то $LocalSim(m_1, m_2) = 10$.

Схожим образом, по своим правилам определяются (вычисляются) значения $LocalSim_i$ для пар остальных параметров TZ_{new} и $TZ_{case j}$.

Шаг 3.

По формуле

$$GlobalSim(T3_{new}, T3_{case\ j}) = (1 - \frac{\sum_{i=1}^n LocalSim_i}{n*10}) * 100,$$

где $n*10$ – это максимально возможное расстояние между новым случаем – $T3_{new}$ и прецедентом – $T3_{case\ j}$, вычисляется значение (в процентах) $GlobalSim$ для $T3_{new}$ и $T3_{case\ j}$.

Если $GlobalSim(T3_{new}, T3_{case\ j}) \geq H$, то $SC' = SC' \cup \{TO_{case\ j}\}$.

Конец цикла.

Если множество $SC' \neq \emptyset$, то его элементы сортируются по убыванию значения процента схожести и в качестве результата возвращается множество SC , элементами которого являются первые k элементов множества SC' , где $1 \leq k \leq |SC'|$, ($SC \subseteq SC'$), в противном случае возвращается \emptyset .

2.2. Апробация и направления дальнейшего улучшения метода

Предложенный метод проходит апробацию в рамках реализации системы поддержки принятия решений (СППР) для инженеров-технологов в сфере лазерной аддитивной обработки металлических материалов по технологии направленного энерговклада.

СППР реализована как композитный сервис на облачной платформе IASPaas [Gribova et. al., 2023]. Система выводит первые пять протоколов ТО, сходство технических заданий которых с новым ТЗ превышает заданный порог (см. рис. 2). Протоколы сортируются по убыванию сходства, а процент схожести отображается рядом с названием. Параметры ТЗ выделены соответствующим цветом (красным, оранжевым или зеленым в зависимости от степени схожести), под ними указаны их значения. Для каждого протокола ТО наряду с описанием ТЗ на выполнение ТО также отображаются выбранные в этом случае технологом средства и параметры выполнения ТО, а также описание результата выполнения ТО (не показано на рисунке).

В настоящее время ведется апробация СППР специалистами Центра инженерных разработок (ЦИР) ИАПУ ДВО РАН. Стоит отметить, что в текущей версии реализации СППР все коэффициенты важности (веса) параметров ТЗ считаются равными 1. В настоящее время ведется реализация возможности гибкой экспертной настройки значений весов для всех параметров ТЗ. Также, по результатам первой апробации, метод может быть улучшен в следующих аспектах. Прежде всего, необходимо доопределить правила вычисления $LocalSim$ для такого параметра ТЗ как «Микроструктура материала».

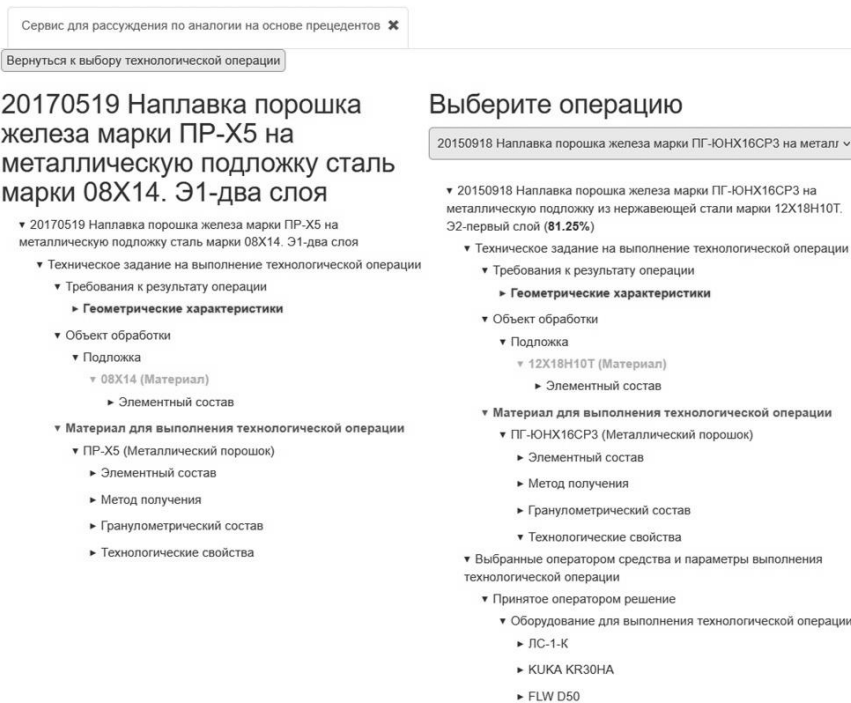


Рис. 2. Визуализация протоколов технологических операций (правая колонка), технические задания которых преодолели заданный порог схожести с новым техническим заданием (левая колонка)

Также следует расширить правила вычисления меры подобия металлических порошков на случай использования нескольких (более одного) порошков при выполнении ТО. Поскольку рассматривается «аддитивный процесс», необходимо будет расширить метод возможностями обработки электронных геометрических моделей деталей, а также стратегии построения и обработки деталей в соответствии с их электронными моделями. Кроме того, при сравнении ситуаций, вероятно, имеет смысл анализировать и учитывать результат выполнения ТО.

Заключение

Принципиальная особенность предложенного метода заключается в обеспечении максимальной гибкости и корректности разработанных мер оценки сходства разнородных элементов ТЗ на выполнение ТО лазерной аддитивной обработки металлических материалов. Используемые при этом базы нормативно-справочной информации формируются по своим

онтологиям. Это позволяет описать метод на основе онтологических структур и сделать его, таким образом, инвариантным по отношению к конкретному наполнению (содержимому) всех баз данных. Последний аспект является принципиальным: базы данных и справочники постоянно пополняются новыми описаниями (данная работа проводится различными предметными специалистами) и появление новой информации в них не должно приводить к модификации разработанного метода.

Предлагаемое решение особенно эффективно в условиях дефицита знаний, стандартизованных решений и рекомендаций (что, в целом, характерно для сферы лазерного аддитивного производства), а также когда другие подходы, в частности, основанные только на знаниях или методах машинного обучения, не могут предложить корректное решение.

Список литературы

- [Варшавский и др., 2013] Варшавский П., Алехин Р. Метод поиска решений в интеллектуальных системах поддержки принятия решений на основе прецедентов // *Information Models and Analyses*. – 2013. – Vol. 2(4). – Р. 385-392.
- [Грибова и др., 2024] Грибова В.В., Кульчин Ю.Н., Никитин А.И., Тимченко В.А. Ансамбль онтологических моделей для обеспечения интеллектуальной поддержки лазерных аддитивных технологических процессов // *Онтология проектирования*. – 2024. – Т. 14, № 2. – С. 279-300. – doi: 10.18287/2223-9537-2024-14-2-279-300.
- [Крылов, 2018] Крылов А.В. Проблема извлечения знаний с использованием рассуждений на основе прецедентов // *Известия высших учебных заведений. Приборостроение*. – 2018. – Т. 61, № 11. – С. 956-962. – doi: 10.17586/0021-3454-2018-61-11-956-962.
- [Майоров и др., 2009] Майоров В.С., Майоров С.В., Стернин М.Ю. Компьютерные системы поддержки принятия решений для лазерных технологических процессов обработки материалов // *Лазерные технологии обработки материалов: современные проблемы фундаментальных исследований и прикладных разработок* / под ред. Панченко В.М. – М.: Физматлит, 2009. – С. 494-506.
- [Малых, 2019] Малых В.Л. Системы поддержки принятия решений в медицине // *Программные системы: теория и приложения*. – 2019. – Т. 10, № 2. – С. 155-184. – doi: 10.25209/2079-3316-2019-10-2-155-184.
- [Полин и др., 2020] Полин Я.А., Зудилова Т.В., Ананченко И.В., Войтюк Т.Е. Деревья решений в задачах классификации: особенности применения и методы повышения качества классификации // *Современные наукоемкие технологии*. – 2020. – № 9. – С. 59-63. – doi: 10.17513/snt.38215.
- [Полкова и др., 2022] Полкова Е.В., Кузяков О.Н. Прецедентный подход в процессах принятия решений // *Инновации. Интеллект. Культура* (Тобольск, 22 апреля 2022 г.). Труды конференции – Тюмень: Тюменский промышленный университет, 2022. – С. 170-173.
- [Юдин и др., 2017] Юдин В.Н., Карпов Л.Е. Неполностью описанные объекты в системах поддержки принятия решений // *Программирование*. – 2017. – № 5. – С. 24-31.

- [Choudhury et al., 2016] Choudhury N., Begum S.A. A survey on case-based reasoning in medicine // International Journal of Advanced Computer Science and Applications. – 2016. – Vol. 7(8). – P. 136-144. – doi: 10.14569/IJACSA.2016.070820.
- [Ethiraj et al., 2025] Ethiraj N., Sivabalan T., Sofia J., Harika D., Nikolova M.P. A comprehensive review on application of machine intelligence in additive manufacturing // Turkish Journal of Engineering. – 2025. – Vol. 9(1). – P. 37-46. – doi: 10.31127/tuje.1502587.
- [Gribova et al., 2023] Gribova V.V., Moskalenko P.M., Timchenko V.A., Shalfeeva E.A. The IACPaaS Platform for Developing Systems Based on Ontologies: A Decade of Use // Scientific and Technical Information Processing. – 2023. – Vol. 50(5). – P. 406-413. – doi: 10.3103/S0147688223050064.
- [Hadj-Mabrouk, 2020] Hadj-Mabrouk H. Application of case-based reasoning to the safety assessment of critical software used in rail transport // Safety Science. – 2020. – Vol. 131. – 104928. – doi: 10.1016/j.ssci.2020.104928.
- [Khan et al., 2019] Khan M.J., Hayat H., Awan I. Hybrid case-base maintenance approach for modeling large scale case-based reasoning systems // Hum. Cent. Comput. Inf. Sci. – 2019. – Vol. 9. 9. – doi: 10.1186/s13673-019-0171-z.
- [Lang et al., 2021] Lang S., Plenk V., Schmid U. A Case-Based Reasoning Approach for a Decision Support System in Manufacturing / In: Fujita H., Selamat A., Lin J.C.W., Ali M. (eds) Advances and Trends in Artificial Intelligence. From Theory to Practice (IEA/AIE 2021). Vol. 12799. – Springer, Cham, 2021. – P. 265-271. doi:10.1007/978-3-030-79463-7_22.
- [Mabkhot et al., 2019] Mabkhot M.M., Al-Samhan A.M., Hidri L. An Ontology-Enabled Case-Based Reasoning Decision Support System for Manufacturing Process Selection // Advances in Materials Science and Engineering. – 2019. – Vol. 2019(1). – 2505183. – doi: 10.1155/2019/2505183.
- [Trofimov, 2020] Trofimov V. An Approach to Intelligent Control of Complex Industrial Processes: An Example of Ferrous Metal Industry // Autom Remote Control. – 2020. – Vol. 81(10). – P. 1856-1864. – doi: 10.1134/S0005117920100057.
- [Varshavskii et al., 2010] Varshavskii P.R., Ereemeev A.P. Modeling of case-based reasoning in intelligent decision support systems // Scientific and Technical Information Processing. – 2010. – Vol. 37(5). – P. 336-345. – doi: 10.3103/S0147688210050096.

ВИРТУАЛЬНЫЙ ТЬЮТОР НА ОСНОВЕ ИНТЕГРАЦИИ БОЛЬШОЙ ЯЗЫКОВОЙ МОДЕЛИ С КОГНИТИВНОЙ АРХИТЕКТУРОЙ

А.А. Долгих (*tolick.dolgi2013@yandex.ru*)
Д.Л. Хабаров (*vip.dima020210@mail.ru*)
А.В. Самсонович (*avsamsonovich@mephi.ru*)

Национальный исследовательский ядерный университет «МИФИ»,
Москва

Биологически инспирированные когнитивные архитектуры (BICA) способны выполнять сложные когнитивные функции на уровне человека. Но для эффективного взаимодействия с социальным окружением им необходим языковой интерфейс, поддерживающий обработку как основного смысла изречений, так и их коннотаций, которые используются, например, в архитектуре eBICA. Большие языковые модели (LLM) могут эффективно выполнять роль связующего звена между внутренними представлениями eBICA и коннотациями речевых актов. Отсюда следует идея нейросимвольной интеграции нового типа. Для оценки работоспособности концепции нами был реализован прототип Виртуального Тьютора, обучающий студентов написанию эссе. Он объединяет когнитивный модуль eBICA, реализующий принципы саморегулируемого обучения через моральные схемы, и интерфейсный модуль на базе LLM GPT-4o, предоставляемый OpenAI. Эксперимент с 48 студентами показал значительное и устойчивое улучшение качества эссе, оцененное по 16 критериям. Ожидается, что подход будет эффективен и в других областях образования: математика, программирование, иностранные языки, и более широко – в интерактивных приложениях, требующих тонкого понимания аспектов социального взаимодействия.

Ключевые слова: интеллектуальные обучающие системы, человеко-машинное взаимодействие, когнитивные архитектуры, большие языковые модели, многомодальный интерфейс, виртуальные интеллектуальные агенты, модели социальных взаимодействий.

Введение

Сегодня все большее внимание уделяется поиску форм интеграции когнитивного (символьного) и статистического нейросетевого подходов в искусственном интеллекте. В частности, интерес вызывает интеграция биологически инспирированных когнитивных архитектур (BICA) и больших языковых моделей (LLM). В настоящей работе исследуется одна возможность такой интеграции, при которой для LLM отводится роль «переводчика» между естественным языком и языком внутренних представлений когнитивной архитектуры eBICA [Samsonovich, 2013, 2020]. Последняя выбрана благодаря использованию в ней моральных схем, определенных в терминах **интенсий**: то есть, коннотационных характеристик речевых актов (не путать с интенцией, то есть намерением). Это качество отличает eBICA от других BICA и ставит новую задачу перед LLM: распознавание и выражение интенсий в речевых актах. Нами было ранее установлено [Nguyen et al., 2025], что современные LLM, такие как DistilBERT и GPT, способны эффективно решать данную задачу. Основной исследовательский вопрос данной работы – это вопрос о работоспособности BICA-LLM интеграции данного типа в решении практической задачи, в качестве которой выбрана задача обучения студента на основе теории саморегулируемого обучения (Self-Regulated Learning, SRL) [Zimmerman et al., 2005], воплощенной в моральных схемах eBICA. Целью не есть достижение современного уровня эффективности интеллектуальных обучающих систем (Intelligent Tutoring Systems, ITS).

Разработка ITS в последнее время стала одной из наиболее быстро развивающихся областей применения как BICA, так и LLM (ограничиваясь инжинирингом исходных промптов) [Chowdhury et al., 2024], [Scarlatos et al., 2025]. Исследуются возможности моделирования человекоподобного обучающего поведения у виртуальных агентов [García-Méndez et al., 2024], [Niu et al., 2024], [Sarker, 2022], [Schmucker et al., 2024]. Ключевая задача здесь – в создании систем, которые эффективно владеют педагогическими стратегиями. При этом важна способность таких систем распознавать и выражать эмоции, прежде всего вербально. Классические когнитивные архитектуры, такие как ACT-R [Anderson et al., 2004] или Soar, не обладают языковыми возможностями, необходимыми для ведения естественного диалога. В то же время большие языковые модели (LLM), такие как GPT-3, LLama и GPT-4, способны к пониманию и генерации произвольного человекоподобного текста, но испытывают трудности с поддержанием структурированного, целенаправленного рассуждения на протяжении длительного времени [Dong, 2023].

Современные подходы к интеграции LLM с когнитивными архитектурами стремятся использовать сильные стороны обеих технологий. Тем не менее, гибридные системы сталкиваются с рядом сложностей. Во-первых,

согласование выходных данных LLM с когнитивно-обоснованным поведением нетривиально: модель может выдавать грамматически правильные, но с точки зрения логики – неподходящие ответы. Во-вторых, координация в реальном времени между вероятностной природой LLM и логикой, основанной на правилах когнитивных моделей, порождает как технические, так и концептуальные трудности [Crompton et al., 2023], [Steinert et al., 2024].

Для преодоления этих ограничений в развиваемом здесь подходе когнитивной архитектуре отводится роль контроллера, который направляет и ограничивает поведение LLM. Это можно представить как рациональное управление «красноречивым говорящим»: когнитивная архитектура отслеживает и направляет диалог, генерируемый LLM, обеспечивая педагогически обоснованное взаимодействие со студентом.

В рамках настоящего исследования мы реализуем предложенный подход, разработав модульную систему, направленную на развитие навыков написания эссе. В данной системе GPT-4o выполняет функции разговорного «движка», в то время как когнитивная архитектура eBICA, воплощающая механизмы SRL, отвечает за рассуждение, извлечение информации из памяти и принятие педагогически обоснованных решений.

В отличие от классических ITS с жесткими правилами, современные подходы обеспечивают динамичную и глубоко персонализированную адаптацию. Примерами таких систем может служить Carnegie Learning Systems (включая MATHia) [Ritter et al., 2016], TECH8 [Dolenc et al., 2015] и SCOPE-IT [Nye et al., 2018]. Предлагаемая система на основе когнитивной архитектуры eBICA и GPT-4o расширяет возможности ITS за счет синтеза когнитивно-эмоционального моделирования и передовых языковых технологий.

Данная работа исследует следующий вопрос: может ли система, основанная на указанном типе интеграции когнитивной архитектуры eBICA и GPT-4o, способствовать улучшению навыков написания эссе?

Представленные результаты демонстрируют положительный эффект. Согласно оценке по 16 критериям, экспериментальные данные указывают на значительное улучшение навыков написания эссе у студентов, взаимодействующих с системой, сохраняющееся в последующих сессиях.

1. Материалы и методы

1.1. Когнитивная архитектура eBICA

Отличительной особенностью данной когнитивной архитектуры является использование моральных схем. Понятие моральной схемы [Samsonovich, 2013], [Samsonovich, 2018], [Samsonovich, 2020] моделирует высший уровень социально-эмоционального рассуждения у когнитивных агентов. Моральная схема представляет собой структурированное пред-

ставление аффективных отношений, охватывающее нормативные состояния и эмоциональные реакции в социальных контекстах (например, доверие, предательство, сожаление). Отклонения от нормативного состояния инициируют когнитивную или поведенческую коррекцию.

Каждая схема включает три компонента: Интерфейс, Фабула и Агентность. Формальная структура выглядит следующим образом:

- **Простая схема:** *Интерфейс* = <Акторы, Домен, Ограничения>, *Фабула* = <Перспектива, Условия>, *Агентность* = <Перспектива, Цель, Средства, Настроение>.
- **Сложная схема:** *Интерфейс* = <Символы, Домен, Условия>, *Фабула* = <Перспективы, Связи, План>, *Агентность* = <Перспектива*, Цель, Средства, Настроение>.

В нарративных терминах фабула задает граф временно- или причинно-связанных перспектив. В то время как простые схемы предполагают наличие единой перспективы (единой точки зрения), сложные схемы поддерживают множественные перспективы агентов, разветвленные сюжетные линии и противоречивые цели.

Эмоциональные оценки агента в рамках схемы кодируются в семантическом пространстве в виде вектора $X_n = (x_1, \dots, x_n, y_1, \dots, y_n)$, где x_i обозначает положительное намерение, а y_i – его антоним.

Архитектура eBICA выделяется тем, что в отличие от классических когнитивных архитектур, которые заточены на моделирование мыслительного процесса, фокусируется на коннотационных социально-эмоциональных аспектах взаимодействия, что позволяет повысить когнитивный уровень взаимодействия с обучаемым.

1.2. Саморегулируемое обучение (SRL)

SRL представляет собой циклический процесс, включающий три основные фазы: (1) предварительное планирование (forethought), (2) выполнение задания (performance), и (3) самоанализ (self-reflection) [Panadero, 2017], [Zimmerman et al., 2005].

Первая фаза охватывает подготовительные действия, предшествующие учебной задаче. К ним относятся постановка целей, стратегическое планирование и анализ задачи. Во второй фазе учащиеся применяют запланированные стратегии, одновременно осуществляя контроль и наблюдение за собственными действиями. В этой фазе может осуществляться контроль времени выполнения, а также организацию учебной среды. Третья фаза включает анализ результатов, эффективности примененных стратегий, а также принятие решений о будущих действиях.

Подобные практики позволяют учащимся вносить осознанные корректировки в свои подходы, способствуя непрерывному улучшению в последующих учебных заданиях.

1.3. Эксперимент

1.3.1 Участники эксперимента. Ими были 48 студентов бакалавриата Национального исследовательского ядерного университета «МИФИ», обучающихся по программе «Программная инженерия». В пул испытуемых вошли 7 женщин и 41 мужчина. Все участники являются носителями русского языка в возрасте от 21 до 28 лет. Студенты-участники были набраны из трех учебных групп четвертого курса обучения на Кафедре кибернетики. Участники не имели предыдущего опыта участия в подобных экспериментах с использованием тьюторов или написанием эссе. Все участники предоставили согласие на участие в эксперименте, обработку результатов и их последующее использование.

1.3.2. Организация эксперимента. Эксперимент проводился в аудиториях университета под наблюдением преподавателей. Студенты использовали свои личные ноутбуки, подключенные к университетской Wi-Fi сети, и могли пользоваться только текстовым редактором и браузером, в котором по предоставленной ссылке открывался диалог с тьютором, развернутым на удаленном сервере (Debian 12 Bookworm, 3.3 ГГц, 2 ГБ). Использование других приложений, интернет-ресурсов, а также взаимодействие с другими участниками было строго запрещено.

1.3.3. Процедура эксперимента. Участники были случайным образом разделены на четыре равные по величине экспериментальные группы, каждая из которых получила свою уникальную конфигурацию задания, причем условия проведения эксперимента, подлежащие сравнению, были сбалансированы. Каждый студент написал два эссе на следующие темы: Тема 1, «Роль open-source проектов в развитии программных экосистем», и Тема 2, «Важность документирования кода и проектов». Каждое эссе писалось в одном из двух условий: с использованием Виртуального Тьютора или самостоятельно. Порядок выбора тем эссе и использования Виртуального Тьютора четырьмя группами показан в табл. 1.

Таблица 1

Группа	Задание	Тема эссе	Тьютор
1	1	Тема 1	С тьютором
	2	Тема 2	Без тьютора
2	1	Тема 1	Без тьютора
	2	Тема 2	С тьютором
3	1	Тема 2	С тьютором
	2	Тема 1	Без тьютора
4	1	Тема 2	Без тьютора
	2	Тема 1	С тьютором

1.4. Детали взаимодействия модулей системы

Реализованная система состоит из четырех модулей (рис. 1). **Модуль визуализации** обеспечивает отображение интерактивной, учебной среды с помощью специально разработанного веб-приложения, созданного на базе стандартных веб-технологий (HTML, CSS, JavaScript). **Модуль взаимодействия** координирует обмен данными между основными компонентами системы и внешними источниками посредством RESTful API. **Модуль генерации** выполняет две ключевые функции: формирование текстовых ответов, соответствующих контексту, и выполнение вычислительных задач в рамках когнитивной модели. **Модуль рассуждений**, реализованный в виде автономного Python-приложения на выделенном сервере, выступает ядром тьютора, объединяя архитектуру eBICA и механизмы саморегулируемого обучения (SRL).

Взаимодействие между студентом и виртуальным тьютором структурировано вокруг фабулы сложной моральной схемы eBICA, включающей четыре уровня. Каждый уровень содержит несколько этапов.

Уровень 1 определяет управление взаимодействием на макроуровне. Система регулирует общее взаимодействие между студентом и тьютором, акцентируя внимание на ключевых аспектах отношений, таких как уважение, иерархия, лидерство и соблюдение правил.

Уровень 2 задает фазы совместной работы студента и тьютора. Этот уровень описывает этапы выполнения учебного задания.

Уровень 3 отвечает за этапы написания эссе. Здесь представлены фазы написания эссе, включая «План», «Историю» и «Детализацию». Данный уровень может рассматриваться как уточнение второго уровня.

Уровень 4 отражает саморегулируемое обучение (SRL). Стандартный цикл SRL, включающий три фазы, может быть применен как явно, так и неявно к любому из этапов трех вышеперечисленных уровней. Он концептуализируется как внутренний цикл рассуждения, выполняемый виртуальным тьютором, невидимый для студента и организованный в виде набора процедурных инструкций.

1.5. Сбор и обработка данных системой

В процессе эксперимента пользователь отправляет реплику через специальное текстовое поле в браузере. Модуль визуализации передает это сообщение на сервер посредством POST-запроса. На сервере модуль взаимодействия принимает запрос, возвращает модулю визуализации статус, а затем пересылает сообщение в модуль рассуждений. В модуле рассуждений сообщение записывается в историю диалога, после чего формируется запрос к модулю генерации – основанному на GPT-4o – с целью получения оценочного вектора. Каждое сообщение оценивается по нескольким аттитудным шкалам, определенным моральной схемой, что

приводит к формированию оценочного вектора, который, в свою очередь, влияет на процесс генерации ответа. Оценки производятся посредством GPT-4o после каждого обмена и сохраняются для последующего анализа.

В России обращение к средствам OpenAI через API ограничено. Поэтому для быстрого доступа к возможностям GPT-4o было реализовано собственное проху-приложение, которое было размещено на хостинге DreamHost.com за пределами Российской Федерации.

После получения вектора оценок архитектура eBICA определяет текущее состояние диалога и формирует индивидуализированный запрос для генерации ответа тьютора с учетом механизмов SRL. Завершающий API-запрос отправляется в OpenAI для генерации ответа тьютора, который затем передается обратно пользователю через модули взаимодействия и визуализации. Весь алгоритм представлен диаграммой на рис. 1.

Валидация оценок по аттитудным шкалам проводилась на 140 высказываниях путем сравнения оценок, полученных с помощью GPT-4o, и оценок, выданных десятью экспертами (в качестве которых были выбраны студенты старших курсов НИЯУ МИФИ). Результаты показали высокий уровень согласованности: 95%-ный доверительный интервал для коэффициента корреляции Пирсона r составил от 0,84 до 0,87. Этот факт подтверждает приемлемую надежность оценок, выданных моделью GPT.

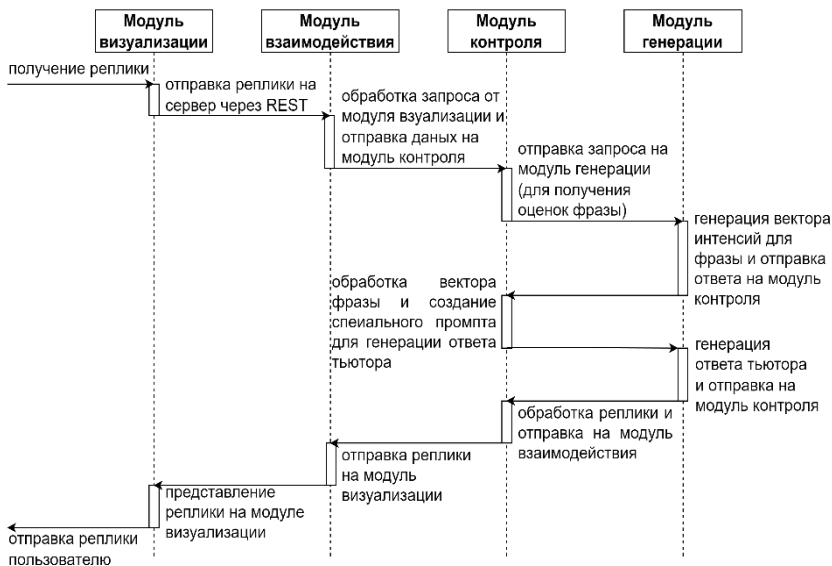


Рис. 1. Диаграмма последовательности виртуального тьютора

Критерии оценок эссе. В данном исследовании для оценки качества написанных студентами эссе были использованы две группы критериев. Критерии первой группы оценивали глобальные характеристики эссе; они определялись следующим образом.

- a) Содержание эссе соответствует его заголовку.
- b) Тема раскрыта в достаточной степени.
- c) Основные положения подобраны удачно.
- d) Аргументация обоснована.
- e) Работа завершена.
- f) Структура эссе адекватна.
- g) Текст логичный, связный и последовательный.
- h) Собственная точка зрения автора выражена.

Вторая группа критериев проверяла наличие в эссе конкретно следующих элементов.

- i) Определение предмета с интуитивными пояснениями.
- j) Обсуждение предыстории и смежных тем.
- k) Рассмотрение вопросов или проблем и существующих решений.
- l) Характеристика современного состояния области.
- m) Изложение методик или технических деталей.
- n) Обсуждение широкого влияния на технологии и общество.
- o) Обсуждение будущих тенденций и перспектив.
- p) Выводы, ключевой итог или «take-home message».

2. Результаты, анализ и обсуждение

Основной целью эксперимента, описанного в секции 1.3, было проверить гипотезу о том, что использование виртуального тьютора способствует улучшению качества эссе, написанных студентами. Для проверки гипотезы все эссе были оценены по 101-бальной шкале Лайкерта (от 0 до 100) с использованием модели GPT-4o на основе специально разработанного промпта, в котором были определены 16 критериев оценки. Для каждого эссе модель возвращала 16 отдельных целых чисел, по одному для каждого критерия. Этот метод автоматизированной оценки был валидирован тремя экспертами и показал высокую согласованность с независимыми оценками экспертов, что также соответствует результатам других недавних работ [Jauhainen et al., 2024], [Quah et al., 2024]. Эссе оценивались в случайном порядке. Число сессий взаимодействия с GPT для оценки одного эссе в разных типах тестов варьировалось от 1 до 100.

Для анализа результатов была проведена трехфакторная дисперсионная оценка (ANOVA). Исследовалось влияние трех независимых факторов: **порядка** написания эссе, **темы** эссе и наличие виртуального **тьютора**. Только фактор тьютора показал значимое влияние на качество эссе

($p < 0,011$). Ни тема эссе, ни порядок их написания не продемонстрировали значимого эффекта. Данные результаты продемонстрированы в таблице ниже (табл. 2).

Для сравнения оценок эссе между условиями **с тьютором** и **без тьютора** был также отдельно проведен двухвыборочный независимый t-тест. Анализ выявил значимую разницу в средних значениях ($M_{\text{Тьютор}}=79,48$; $M_{\text{NoТьютор}}=76,13$; $t(df)=2,63$; $p < 0,011$), при этом эссе, написанные с тьютором, показали более высокие оценки: зафиксирована разница в 3,35 балла.

Таблица 2

Источник вариации	Сумма квадратов	Степени свободы	Средний квадрат	F-score	p-value
Тьютор	273.75	1	273.753	6.76	0.0109*
Тема	31.98	1	31.978	0.79	0.3766
Порядок	48.61	1	48.611	1.20	0.2762
Тьютор:Тема	22.31	1	22.310	0.55	0.4599
Тьютор:Порядок	0.47	1	0.472	0.01	0.9142
Тема:Порядок	18.47	1	18.475	0.46	0.5012

Примечательно, что в группах, начавших писать эссе без тьютора, второе эссе (написанное с поддержкой тьютора) оказалось лучше первого ($p < 0.013$ с поправкой Бонферрони). Это указывает на то, что вмешательство тьютора оказало существенное влияние при его использовании после первоначальной самостоятельной попытки. В то же время в группах, писавших первое эссе с тьютором, второе эссе (написанное самостоятельно) оказалось по качеству не хуже первого. Это указывает на то, что эффект тьютора сохранился после его однократного использования.

Было также проанализировано влияние тьютора на оценки по индивидуальным критериям. Наибольший положительный эффект (15,2%) был получен для критерия (о): наличие в эссе выводов, основной мысли или заключительного утверждения ($p < 0.0094$). Это свидетельствует о том, что Виртуальный Тьютор играет ключевую роль в стимулировании студентов к более аналитическому и ориентированному на вывод подходу при написании эссе. Несколько других критериев также продемонстрировали значимое улучшение в диапазоне от 7% до 10%, включая критерии (е), (i), (j), (l), (n) и (p). Данные результаты проиллюстрированы на рисунке ниже.

Полученные результаты свидетельствуют о том, что наибольшую эффективность Виртуальный Тьютор демонстрирует при повышении глубины и структурированности эссе – особенно в тех аспектах, которые требуют интеграции знаний, технической проработки и учета широкого социального или перспективного контекста.

Работа системы была стабильной. Максимальная задержка ответа Виртуального Тьютора составила не более 10 секунд, при этом в среднем на формирование ответа уходило 3.8 секунд.

Индивидуальные различия между обучаемыми учитываются путем анализа интенсий ответов студента, определения на этой основе его ати-тюдных и мотивационных характеристик, и выбора адекватной модели поведения Виртуального Тьютора на основе моральных схем eBICA. Система работает со студентом в режиме один-на-один, что делает возможной индивидуальную подстройку. При этом внесение априорной дополнительной информации об обучающемся (например, уровень исходных знаний) могло бы помочь повысить эффективность системы.

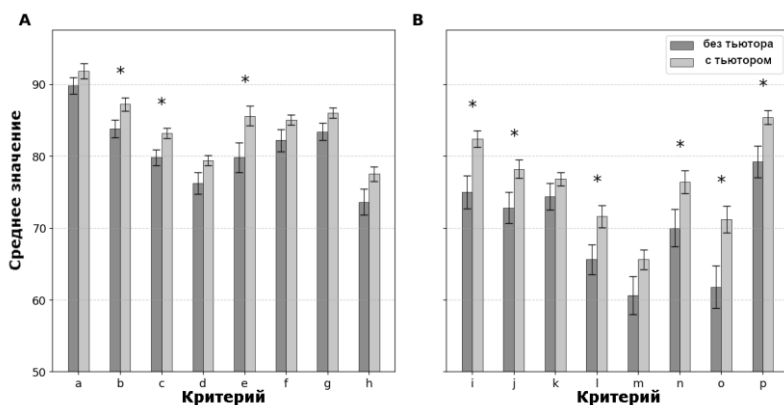


Рис. 2. Сравнение средних значений по каждому критерию

Для сравнения были проведены два контрольных эксперимента с другой группой студентов, отличавшиеся отсутствием в системе модуля рассуждений на базе eBICA: вместо этого LLM получала изначальный промпт, описывающий весь сценарий взаимодействия, и далее действовала самостоятельно. Причем в одном из этих экспериментов вместо GPT-4o использовался DeepSeek. Результаты, которые в силу ограничений объема не могут быть представлены здесь, не показали значимого улучшения качества эссе по выбранным критериям.

Целью данной работы было не сравнение эффективности системы с уровнем развития ITS технологий, а подтверждение работоспособности новой концепции нейро-символьной интеграции. В то же время, перспективным видится проведение более масштабных экспериментов, включая расширение набора тем эссе и разнообразие контрольных групп.

Таким образом, детальный анализ демонстрирует, что Виртуальный Тьютор способствует развитию аналитической насыщенности, технической точности и стратегической полноты в студенческих эссе. Это полностью соответствует заложенной в систему когнитивной функции: не столько корректировать язык, сколько направлять мышление и углублять аргументацию в процессе написания.

Заключение

Результаты исследования подтверждают работоспособность прототипа предложенного варианта нейро-символьной интеграции, обеспечившей в данном случае эффективное применение SRL в образовательном процессе. В то же время ограниченность проведенного эксперимента не позволяет делать обобщения на основе данного результата.

Ограниченный объем статьи не позволил раскрыть все детали воплощенной системы и контрольных экспериментов. Это будет сделано в более обстоятельной журнальной публикации.

Выбор когнитивной архитектуры eBICA обусловлен тем, что данное исследование продолжает предыдущие работы авторов на эту тему. Отличительные черты eBICA – это наличие в ней моральных схем, благодаря чему предложенная интеграция становится нетривиальной.

Преимущество предложенного общего подхода в том, что он позволяет установить связь когнитивной архитектуры с реальным социальным окружением (в данном случае, со студентом) на уровне тонких коннотационных характеристик социального взаимодействия, что позволяет использовать более продвинутые модели вроде eBICA для реализации полноценного социального контакта агента с пользователем.

Среди недостатков следует назвать ограниченность созданного прототипа и проведенного с ним исследования. Целью работы была лишь проверка работоспособности предложенной концепции. В то же время бесспорно, что представленная система имеет существенные ограничения.

В дальнейших исследованиях планируется внедрение нейросетевой системы для более точного и независимого определения интенсий высказываний, что может повысить социально-эмоциональный интеллект Виртуального Тьютора и улучшить его взаимодействие с обучающимися.

Список литературы

- [Anderson et al., 2004] Anderson J.R. et al. An integrated theory of the mind // *Psychological Review*. – 2004. – Vol. 111(4). – P. 1036. – doi: 10.1037/0033-295X.111.4.1036.
- [Chowdhury et al., 2024] Chowdhury S.P., Zouhar V., and Sachan M. AutoTutor meets large language models: A language model tutor with rich pedagogy and guardrails // *arXiv preprint*. – 2024. – doi: 10.48550/arXiv.2402.09216.

- [Crompton et al., 2023] Crompton H., Burke D. Artificial intelligence in higher education: The state of the field // *International Journal of Educational Technology in Higher Education*. – 2023. – Vol. 20(1). – P. 22. – doi: 10.1186/s41239-023-00392-8.
- [Dolenc et al., 2015] Dolenc K., Aberšek B. TECH8 intelligent and adaptive e-learning system: Integration into Technology and Science classrooms in lower secondary schools // *Computers & Education*. – 2015. – Vol. 82. – P. 354-365.
- [Dong, 2023] Dong C. How to build an AI tutor that can adapt to any course and provide accurate answers using large language model and retrieval-augmented generation // *arXiv preprint arXiv:2311.17696*. – 2023. – doi: 10.48550/arXiv.2311.17696.
- [García-Méndez et al., 2024] García-Méndez S., Arriba-Pérez F. de, Somoza-López M. del C. A review on the use of large language models as virtual tutors // *Science & Education*. – 2024. – P. 1-16. – doi: 10.1007/s11191-024-00530-2.
- [Jauhiainen et al., 2024] Jauhiainen J.S., Guerra A.G. Generative AI in education: ChatGPT-4 in evaluating students' written responses // *Innovations in Education and Teaching International*. – 2024. – P. 1-18. – doi: 10.1080/14703297.2024.2422337.
- [Nguyen et al., 2025] Nguyen T.M. and Samsonovich A.V. Developing a general-purpose system for intentionality detection in dialogue using neural networks // *Lecture Notes in Artificial Intelligence*. – 2025. – Vol. 16058. – P. 12-22. – doi: 10.1007/978-3-032-00800-8_2.
- [Niu et al., 2024] Niu Q. et al. Large language models and cognitive science: A comprehensive review of similarities, differences, and challenges // *arXiv preprint arXiv:2409.02387*. – 2024. – doi: 10.48550/arXiv.2409.02387.
- [Nye et al., 2018] Nye B. et al. SKOPE-IT (Shareable Knowledge Objects as Portable Intelligent Tutors): overlaying natural language tutoring on an adaptive learning system for mathematics // *International journal of STEM education*. – 2018. – Vol. 5(1). – P. 12. – doi: 10.1186/s40594-018-0109-4.
- [Panadero, 2017] Panadero E. A review of self-regulated learning: Six models and four directions for research // *Frontiers in Psychology*. – 2017. – Vol. 8. – P. 422. – doi: 10.3389/fpsyg.2017.00422.
- [Quah et al., 2024] Quah B. et al. Reliability of ChatGPT in automated essay scoring for dental undergraduate examinations // *BMC Medical Education*. – 2024. – Vol. 24 (1). – P. 962. – doi: 10.1186/s12909-024-05881-6.
- [Ritter et al., 2016] Ritter S., Fancsal S. MATHia X: The Next Generation Cognitive Tutor // *EDM*. – P. 624-625.
- [Samsonovich, 2013] Samsonovich A.V. Emotional biologically inspired cognitive architecture // *Biologically Inspired Cognitive Architectures*. – 2013. – Vol. 6. – P. 109-125. – doi: 10.1016/j.bica.2013.07.009.
- [Samsonovich, 2018] Samsonovich A.V. Schema formalism for the common model of cognition // *Biologically Inspired Cognitive Architectures*. – 2018. – Vol. 26. – P. 1-19. – doi: 10.1016/j.bica.2017.12.002.
- [Samsonovich, 2020] Samsonovich A.V. Socially emotional brain-inspired cognitive architecture framework for artificial intelligence // *Cognitive Systems Research*. – 2020. – Vol. 60. – P. 57-76. – doi: 10.1016/j.cogsys.2019.12.002.
- [Sarker, 2022] Sarker I.H. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems // *SN Computer Science*. – 2022. – Vol. 3(2). – P. 158. – doi: 10.1007/s42979-022-01043-x.

- [**Scarlatos et al., 2025**] Scarlatos A., Liu N., Lee J., Baraniuk R., and Lan A. Training llm-based tutors to improve student learning outcomes in dialogues // arXiv preprint arXiv:2503.06424. – doi: 10.48550/arXiv.2503.06424.
- [**Schmucker et al., 2024**] Schmucker R. et al. Ruffle&Riley: Insights from designing and evaluating a large language model-based conversational tutoring system. International Conference on Artificial Intelligence in Education. – Springer, 2024. – P. 75-90. – doi: 10.1007/978-3-031-64302-6_6.
- [**Steinert et al., 2024**] Steinert S. et al. Harnessing large language models to develop research-based learning assistants for formative feedback // Smart Learning Environments. – 2024. – Vol. 11(1). – P. 62. – doi: 10.1186/s40561-024-00354-1.
- [**Zimmerman et al., 2005**] Zimmerman B.J., Kitsantas A. The Hidden dimension of personal competence: Self-Regulated Learning and Practice. – 2005.

УДК 004.82

doi: 10.15622/rcai.2025.093

ПРОГНОСТИЧЕСКАЯ СИСТЕМА МОНИТОРИНГА БОЛЬНЫХ ХРОНИЧЕСКИМ ЛИМФОЛЕЙКОЗОМ

Н.А. Игнатьев (*n_ignatev@rambler.ru*)^A

Е.Н. Згуральская (*iatu@inbox.ru*)^B

М.В. Марковцева (*mmark7@yandex.ru*)^C

^A Национальный университет Узбекистана, Ташкент

^B Ульяновский государственный технический университет, Ульяновск

^C Ульяновский государственный университет, Ульяновск

Интеллектуальные системы мониторинга состояния здоровья в клинической практике оптимизируют оказание медицинской помощи пациентам. Такие системы позволяют с течением времени отслеживать изменения значений наблюдаемых параметров, вносить коррективы в прогноз заболевания. Результаты интеллектуального анализа, полученные в виде правил продукций, особенно полезны для представления экспертных знаний и автоматизации процессов принятия решений. Для формализации знаний разработана онтология, применяемая для мониторинга состояния пациентов в течение 5 лет с момента постановки диагноза хронический лимфолейкоз (ХЛЛ). Выделены основные связи понятий, достаточные для решения задач мониторинга состояния больных ХЛЛ в процессе лечения. Разработанная онтология входит в основу прогностической системы, где параметры мониторинга могут изменяться с учётом стадии лечения, клинических, лабораторных данных пациентов. Описан метод извлечения знаний, основанный на теории нечётких множеств. Описана система мониторинга, связанная с базой знаний.

Ключевые слова: онтология, правила продукций, база знаний, интеллектуальный анализ данных, хронический лимфолейкоз.

Введение

Мониторинг состояния пациентов является неотъемлемой частью организации лечебно-диагностического процесса в современных условиях [Грибова и др., 2024]. На практике мониторинг позволяет своевременно прояснить клиническую ситуацию пациента и при необходимости скор-

ректировать лечение больного. Полученные данные хранятся и используются для создания медицинских информационных, экспертных или рекомендательных систем, в которых накопленную информацию объединяют в базы знаний (БЗ), формируя на их основе методы решения задач [Богданова и др., 2024]. Система обобщения и классификации, которая реализована в интеллектуальных системах (ИС), позволяет сводить множество возможных ситуаций в многопризнаковом пространстве к небольшому числу типовых ситуаций, по которым выдвигаются гипотезы и производится выбор методов решения [Кобринский и др., 2020].

Навыки и опыт врача не представлены в строго формализованном виде, они являются результатом синтеза теоретических знаний и клинических наблюдений. С учётом этих особенностей в медицине связано построение ИС, т. е. систем, основанных на работе не с данными о больных, а со знаниями высоко профессиональных специалистов-экспертов [Кобринский, 2001]. Интеллектуальный анализ данных (ИАД), предлагая методы для выявления закономерностей, проверки гипотез и извлечения ценной информации, нивелирует имеющиеся экспертные противоречия. Гипотезы, выдвинутые врачами, проверяются методами ИАД и в случае их подтверждения включается в БЗ.

Желательным свойством при поиске закономерностей в базах и хранилищах данных является инвариантность результатов анализа к масштабам их измерений. При поиске закономерностей использовались интервальные методы, одним из универсальных ограничений которых является инвариантность к масштабам измерений данных [Згуральская, 2018]. Свойство инвариантности упрощает процедуру интерпретации результатов в терминах предметных областей. Это свойство даёт возможность выбора латентных признаков при моделировании процесса интуитивного принятия решений специалистами.

В [Игнатьев и др., 2020] рассматривается поиск логических закономерностей по описаниям объектов в спрямляющем пространстве. При синтезе латентных признаков этого пространства используются правила иерархической агломеративной группировки. Выбор пары признаков для объединения в группу реализуется по максимуму критерия разбиения значений признаков на непересекающиеся интервалы. Приводится аналитический вид арифметических выражений для расчета латентных признаков, используемых для обнаружения скрытых закономерностей на данных больных ХЛЛ.

В статье предлагается подход к построению системы мониторинга больных ХЛЛ с использованием БЗ, наполнение которой произведено методами ИАД.

1. Постановка задачи

ХЛЛ – одно из наиболее распространённых онкогематологических заболеваний. Высокая распространённость ХЛЛ у пациентов европейской популяции в пожилом (60–75 лет) и старческом (более 75 лет) возрасте представляет на сегодняшний день актуальную проблему. Мониторинг пациентов с ХЛЛ является важной задачей для оптимизации тактики ведения больных. Предложенные в последнее время способы прогнозирования и мониторинга ХЛЛ требуют сложных генетических исследований, которые не всегда являются доступными: мутационный статус генов вариабельного региона иммуноглобулинов (IGHV), мутация TP53, делеция 13q14 и др. [Mayr et al., 2006]. Подобная ситуация диктует необходимость продолжения поиска общедоступного решения данного вопроса [Марковцева и др., 2023].

Анализ клинических данных методами ИАД позволяет оперировать более точной и персонализированной информацией о прогнозе заболевания, учитывая влияние на состояние пациента возможных коморбидных эффектов. На основании текущих индивидуальных определений параметров нельзя однозначно утверждать о возможном состоянии больных в последующем, т.е. сделать прогноз [Адмакин, 2016]. В задаче интеллектуального мониторинга требуется по информации о системе, состоянии которой вышло из гипотетической нормы, с помощью формализованных знаний определить показатели каких наблюдений и в какие моменты времени является индикатором, что состояние системы отличается от ожидаемого (прогнозируемого). Для этого требуется:

- собрать информацию (клинические данные) о пациентах с ХЛЛ, с момента постановки диагноза и в течении последующей пятилетней динамике заболевания;
- методами ИАД вывить наличие закономерностей в динамически изменяемых данных;
- обнаруженные закономерности представить в виде правил продукций;
- разработать блок объяснения процесса принятия решений.

2. Модель БЗ для мониторинга состояния больных

Для формализации знаний разработана онтология, применяемая для мониторинга состояния пациентов с момента постановки диагноза ХЛЛ и на протяжении всего заболевания.

Онтологию O_{HLL} можно представить в виде следующего упорядоченного набора элементов:

$$O_{HLL} = \langle G, I, SWRL \rangle,$$

где G – утверждения, соответствующие уровню схемы онтологии O_{HLL} ;

I – множество утверждений, соответствующих индивидам онтологии;

$SWRL$ – множество продукционных правил на языке SWRL.

Множество G может быть определено кортежем, состоящим из следующих взаимосвязанных элементов:

$$G = \{G^C, G^O, G^D, G^A\},$$

где G^C – множество утверждений, характеризующих классы (концепты) онтологии;

G^O – множество утверждений, описывающих объектные свойства, объявленные и используемые в онтологии для формирования связей между концептами;

G^D – множество утверждений, описывающих свойства данных;

G^A – множество аксиом, представляющих основные ограничения для элементов множеств G^C, G^O, G^D .

Онтологическая модель (ОМ) включает классы, отношения и ограничения, которые определяют предметную область, тем самым устанавливает единую терминологию для совместного использования информации о взаимодействии элементов, необходимых для мониторинга состояния пациентов в лечебно-диагностическом процессе.

В ОМ O_{HLL} представлены пять классов $G^C: G^C\{G^C_{clin}, G^C_{lab}, G^C_{stage}, G^C_{risk}, G^C_P\}$, где:

G^C_{clin} – класс «Клинические факторы», содержащий сведения о возрасте, наличие сопутствующих заболеваний, наличие симптомов (лихорадка, ночная потливость, потеря веса, усталость) и т.п.

G^C_{lab} – класс «Лабораторные показатели» содержащий сведения о лабораторных исследованиях крови АЛТ, АСТ, билирубин-общий, непрямой-билирубин, креатинин, мочевины, глюкозы и т.п.;

G^C_{stage} – класс «Стадия заболевания», определение стадии по системе Binet, которые основаны на клинических и лабораторных факторах (количество лимфоцитов, вовлечение органов, уровень гемоглобина).

G^C_{risk} – класс «Риск», отнесение к данному классу производится на основе правил продукций, полученных методами ИАД и зависит от клинических, лабораторных показателей: креатинин, пол, возраст и стадии заболевания. Это позволяет механизму логического вывода автоматически классифицировать экземпляры этих классов, если они удовлетворяют заданным условиям.

G^C_P – класс «Пациенты».

Правила продукций являются одним из методов представления и использования знаний в экспертных системах и в ИАД. Они особенно полезны для записи экспертных знаний и автоматизации процессов принятия решений.

Для формирования схемы БЗ и правил продукций на языке SWRL использован редактор онтологий Protégé [Horrocks et al., 2004]. В результате схема БЗ и множество продукций на языке SWRL представляются в виде онтологии на языке OWL. Фрагмент онтологии представлен на рис. 1.



Рис. 1. Фрагмент онтологии

Правила продукций на языке SWRL представлены на рис. 2.

	Query
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, ?srok) ^ swrlb:lessThanOrEqual(?skf, 76) -> Риски_есть(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, autogen1:_год) ^ swrlb:greaterThan(?skf, 75) -> Рисков_нет(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, autogen0:_года) ^ swrlb:greaterThan(?skf, 70) -> Рисков_нет(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, autogen4:_года) ^ swrlb:greaterThan(?skf, 65) -> Рисков_нет(?p)	
Пациенты(?p) ^ имеет_СКФ(?p, ?skf) ^ имеет_срок_заболевания(?p, постановка_диагноза) ^ swrlb:greaterThan(?skf, 76) -> Рисков_нет(?p)	

Рис. 2. Правила продукций на языке SWRL

3. Применение методов ИАД для выявления закономерностей

Для поиска закономерностей применялись методы разбиения значений признаков на непересекающиеся интервалы. На базе этих методов стало возможным как вычисление обобщённых оценок объектов (латентных признаков) в разнотипном признаковом пространстве, так и меры их устойчивости [Згуральская, 2018]. Следует отметить, что применение альтернативных подходов для выявления закономерностей существенно ограничено особенностями процедуры сбора исходных данных.

В предлагаемой системе мониторинга используется такое фундаментальное свойство размеченных данных, как устойчивость. Вычисление устойчивости со множеством допустимых значений в $(0,5;1]$ основано на теории нечётких множеств. Доказано, что свойство устойчивости присуще всем выборкам из генеральной совокупности. Свойство выполняется для количественных и номинальных (качественных) данных, является инвариантным к масштабам измерений. Инвариантность прослеживается при разбиении значений признаков на непересекающиеся интервалы с доминированием в каждом из них представителей одного из двух классов.

Количество интервалов и их состав определяются как решение задачи о минимальном покрытии по упорядоченной последовательности значений признака. Вычисление значений функции принадлежности по интервалам с учётом их оптимальных границ рассматривается как источник новых знаний о предметной области. Доказано [Игнатьев и др., 2023], что определяемый по многообразиям значений функции принадлежности показатель устойчивости признака сходится к фиксированной величине.

Было исследовано 207 пациентов (98 мужчин и 109 женщин) в возрасте от 45 до 80 лет с ХЛЛ А-С по Binet [Binet et al., 1981] и известной общей выживаемостью. Описание клинической значимости признаков, используемых в вычислительных экспериментах приводится в многочисленных публикациях в области медицины [Марковцева и др., 2023], [Pektaş et al., 2025].

Вычислительный эксперимент проводился в несколько этапов. На первом этапе исследовались показатели группы больных, прошедших начальное обследование. Последующие наблюдения проводились каждые двенадцать месяцев в течение пяти лет. За указанный промежуток времени численность исследуемой группы постепенно сокращалась вследствие смертности пациентов. Следует отметить, что проведение оценки состояния здоровья с меньшими временными интервалами было ограничено рядом факторов объективного характера.

Данные полученные на этапе постановки диагноза и в результате повторных обследований больных использовались для поиска закономерностей в выборках данных по значениям устойчивости.

Была получена закономерность, связанная со значениями латентного признака скорость клубочковой фильтрации (СКФ). Существует функциональная зависимость СКФ от значений возраста и креатинина в зависимости от гендерной принадлежности. Значение латентного признака СКФ рассчитывают по формуле [Марковцева, 2023]:

$$\text{СКФ} = 141 * \min\left(\frac{\text{Креатинин}}{\kappa}, 1\right) * \alpha * \max\left(\frac{\text{Креатинин}}{\kappa}, 1\right) * 1.209 * 0.993 * \text{Возраст} * \text{Пол} * \text{Раса}, \quad (1)$$

где: для женщин используются следующие значения: пол = 1.018, α = - 0.329; κ = 0.7; для мужчин используются следующие значения: пол = 1, α = - 0.411; κ = 0.9; для представителей негроидной расы: коэффициент «Раса» = 1.159.

Динамика заболевания по этапам анализировалась по результатам разбиения значений СКФ на непересекающиеся интервалы. Никаких предположений о плотности распределения данных не делалось. Количество интервалов, оптимальные значения их границ и состав входящих в них объектов не являлись изначально фиксированными. Использовался алгоритм метода [Згуральская, 2018] минимального покрытия значений количественного признака непересекающимися интервалами, в границах которых доминируют представители одного из двух классов.

Для мониторинга пациентов при повторных обследованиях будет достаточно задать значения измеримых показателей возраст и креатинин [Игнатьев и др., 2023]. При значении СКФ более 76 мл/ мин/1,73 м² в течение первого года заболевания, более 75 мл/мин/1,73 м² в течение второго года заболевания, более 70 мл/мин/1,73 м² в течение третьего года заболевания, более 68 мл/мин/1,73 м² в течение четвёртого и пятого годов заболевания пациент преодолает расчётные параметры медианы выживаемости для соответствующей стадии ХЛЛ по Binet. В противном случае – общая выживаемость пациента будет меньше расчётной медианы выживаемости по Binet.

В [Марковцева и др., 2023] приведены показательные клинические примеры. Выявлено, что значение СКФ у пациентов ХЛЛ планомерно снижается с течением заболевания (рис. 3) и его в динамике можно рассматривать в качестве значимого фактора мониторинга пациентов с ХЛЛ.

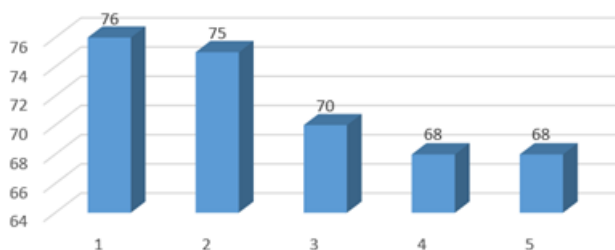


Рис. 3. Границы значения показателя СКФ при котором пациент преодолевает медиану выживаемости по классификации Binet в зависимости от этапа обследования

4. Функциональная схема использования онтологии в системе мониторинга

Формализованные клинические знания и обнаруженные скрытые закономерности методами ИАД составляют основу системы мониторинга больных с ХЛЛ. Функциональная схема использования онтологии в системе мониторинга:

1. На этапе постановки диагноза врач вносит информацию о лабораторных и клинических показателях пациента в форму ввода прогностической системы.
2. На основе полученной информации вычисляется показатель СКФ по формуле (1).
3. Данные о пациенте трансформируются в формат онтологии, создается экземпляр класса пациенты (ФИО пациента, номер обследования).

4. Путем логического вывода на основе множества продукционных правил на языке SWRL в БЗ формируются новые знания.
5. Полученная информация поступает обратно к врачу, формируя отчет для специалиста с оценкой развития неблагоприятных событий (смерти) у пациентов с ХЛЛ (стадия заболевания, группа риска и т.п.).
6. Проводятся повторные обследования, информация о которых поступает в систему чем соответствующую форму.
7. Происходит поиск информации по онтологии о данном пациенте, если информация о предыдущих обследованиях не найдена, то переходим к п.1.
8. Переходим к п.2.

Заключение

Разработанная прогностическая система, основанная на знаниях для стратификации рисков и мониторинга больных с ХЛЛ предназначена для автоматизированной оценки вероятности развития неблагоприятных событий (смерти) у пациентов и организации персонализированного мониторинга на основе этой оценки, позволит оптимизировать распределение ресурсов здравоохранения, улучшить результаты лечения.

Список литературы

- [Грибова и др., 2024] Грибова В.В., Окунь Д.Б., Шалфеева Е.А. Применение графов знаний для клинического мониторинга процесса лечения // Онтология проектирования. – 2024. – Т. 14, № 4(54). – С. 504-517. – doi: 10.18287/2223-9537-2024-14-4-504-517.
- [Богданова и др., 2024] Богданова Д.Р., Шахматова Г.Р., Ниязгулов А.М. Структура информационного хранилища системы поддержки принятия клинических решений // Онтология проектирования. – 2024. – Т. 14, № 2(52). – С. 270-278. – doi: 10.18287/2223-9537-2024-14-2-270-278.
- [Кобринский и др., 2020] Кобринский Б.А., Хавкин А.И., Волюнец Г.В. Перспективы применения систем искусственного интеллекта в гастроэнтерологии // Экспериментальная и клиническая гастроэнтерология. – 2020. – № 179(7). – С. 109-117. – doi: 10.31146/1682-8658-esg-179-7-109-117.
- [Кобринский, 2001] Кобринский Б.А. Искусственный интеллект и медицина: Возможности и перспективы систем, основанных на знаниях // Новости искусственного интеллекта. – 2001. – № 4. – С. 44-51.
- [Игнатьев и др., 2020] Игнатьев Н.А., Згуральская Е.Н., Марковцева М.В. Нелинейные преобразования признаков и поиск закономерностей на данных больных хроническим лимфолейкозом // VI Международная конференция. Информационные технологии и нанотехнологии (Самара, 26–29 мая 2020 г.): Труды конференции. В 4-х т. Т. 4. – Самара: Самарский национальный исследовательский университет имени академика С.П. Королева, 2020. – С. 123-128.

- [**A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis, 1981**] Binet J.L., Auquier A., Dighiero G., et al. A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis // *Cancer*. – 1981. – No. 48(1) – URL: [https://doi.org/10.1002/1097-0142\(19810701\)48:1<198::aid-cnrcr2820480131>3.0.co;2-v](https://doi.org/10.1002/1097-0142(19810701)48:1<198::aid-cnrcr2820480131>3.0.co;2-v) (дата обращения: 29.05.2025).
- [**Марковцева и др., 2023**] Марковцева М.В., Згуральская Е.Н. Прогнозирование общей выживаемости больных хроническим лимфолейкозом в динамике заболевания // *Качественная клиническая практика*. – 2023. – № 3. – С. 66-71.
- [**Адмакин, 2016**] Адмакин А.Л. Цепи Маркова – стохастическая модель анализа состояния тяжелообожженных // *Мед. биол. и соц. психол. пробл. безопасности в чрезв. ситуациях*. – 2016. – № 3. – С. 119-125. – doi: 10.25016/2541-7487-2016-0-3-119-125.
- [**SWRL: A Semantic Web Rule Language Combining OWL and RuleML, 2024**] Horrocks I., Patel-Schneider P.F., Boley H., Tabet S., Grosz B., Dean M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML // *Member Submissions W3C*. 2004. URL: <https://www.w3.org/Submission/SWRL/> (дата обращения: 21.04.2025).
- [**Згуральская, 2018**] Згуральская Е.Н. Устойчивость разбиения данных на интервалы в задачах распознавания и поиск скрытых закономерностей // *Известия Самарского научного центра Российской академии наук*. – 2018. – Т. 20, № 4-3(84). – С. 451-455.
- [**Игнатьев и др., 2023**] Игнатьев Н.А., Акбаров Б.Х. Оценка близости структур отношений объектов обучающей выборки на многообразиях наборов латентных признаков // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. – 2023. – № 65. – С. 69-78. – doi: 10.17223/19988605/65/7.
- [**Mayr et. al., 2006**] Mayr C., Speicher M.R., Kofler D.M. Chromosomal translocations are associated with poor prognosis in chronic lymphocytic leukemia // *Blood*. – 2006. – Vol. 107(2). – P. 742-51.
- [**Pektaş et. al., 2025**] Pektaş G., Gönül E., Öncü Ş. Chronic Lymphocytic Leukemia: Investigation of Survival and Prognostic Factors with Drug-Related Remission // *Diagnostics*. – 2025. – Vol. 15. – P. 728. – doi: 10.3390/diagnostics15060728.

УДК 004. 942+712.00
doi: 10.15622/rcai.2025.094

МАРКИРОВКА ВИДИМОСТИ ЭЛЕМЕНТОВ BIM-МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ГРАФОВОЙ НЕЙРОННОЙ СЕТИ

Д.Д. Кабайлов (*denis2000kab@gmail.com*)^{А,В}
А.М. Наместников (*am.namestnikov@gmail.com*)^А

^А Санкт-Петербургский государственный
архитектурно-строительный университет, Санкт-Петербург
^В ООО «Рэд Лайн Проект», Санкт-Петербург

В данной работе описан метод маркировки элементов цифровых информационных моделей зданий на определенных категориях видов с использованием графовой нейронной сети. В первом разделе работы приведен анализ существующих решений формирования чертежей их особенностей и недостатков. Во втором разделе описываются основные принципы построения и структура несвязного графа, представляющего совокупность информационных моделей зданий. Также приведены графические модели, отображающие полученный несвязный граф. В третьем разделе работы описаны причины выбора графовой нейронной сети Graph Attention Network (GAT) и слоев GATv2Conv. Описана структура нейросети, обучаемая на полученном графе, которая позволяет предсказывать метки видимости на видах для элементов информационных моделей зданий. Приведены результаты вычислительных экспериментов с разработанной нейросетевой моделью.

Ключевые слова: графовая нейронная сеть, сверточная нейронная сеть, GAT, ТИМ, проектирование.

Введение

Создание чертежей зданий и сооружений является основополагающим этапом проектирования, в том числе в строительной сфере. Однако, чтобы создать качественный чертеж, необходимо определить оптимальное количество видов (планов/ фасадов/ разрезов и т.п.) и то, какие элементы на этих видах должны быть отображены.

В ряде исследований применялись алгоритмы классификации элементов информационных моделей, например в [Austern et al., 2024], однако они не позволяют определить, к каким видам какой элемент относится. Графовые нейросетевые модели Graph Convolutional Network и Graph Attention Network (GCN, GAT) [Çeter, 2024] демонстрируют перспективные результаты в смежных задачах аннотирования планов, но их точность и адаптивность пока недостаточны для масштабного применения. Таким образом, остаётся нерешённой задача автоматизированного формирования проектной документации, способной учитывать взаимосвязь всех элементов и атрибутов в модели.

Данные ограничения указывают на необходимость разработки моделей, которые будут учитывать геометрию и полный набор атрибутов элементов, обеспечивая корректное отображение на всех типах видов. Целью научного исследования является создание метода формирования проектной документации на основе интеграции технологий информационного моделирования (ТИМ) и нейросетевых алгоритмов анализа информационных моделей зданий и сооружений. Научная новизна работы состоит в исследовании способов адаптации и эффективности применения нейросетевых алгоритмов в задаче формирования чертежей зданий и сооружений.

Для достижения цели, принято решение провести ее декомпозицию на подзадачи. В рамках данной статьи приведен разбор следующих задач:

- анализ существующих подходов к формированию чертежей зданий и сооружений;
- описание метода анализа информационных моделей, использующего нейросетевой алгоритм, выполняющий маркировку видимости элементов на различных типах видов;
- анализ результатов исследования и разбор дальнейших шагов исследования.

В данной работе для классификации элементов по отображению на видах в модели принято решение использовать графовые нейросетевые алгоритмы, способные учитывать большой массив связанных данных. Это позволит создавать чертежные виды и выносить их на листы строительной документации.

1. Анализ современных подходов к формированию чертежей

В настоящее время большинство задач в процессе формирования чертежей реализованы с помощью специально разработанных дополнительных надстроек для популярных систем автоматизированного проектирования (САПР). Системы на основе правил (rule-based) алгоритмов [Kim et al., 2024] позволяют формализовать отдельные этапы оформления чертежей, например простановку размеров и формирование спецификаций.

Однако такие решения, как показано в [Kim et al., 2024], не учитывают уже созданные виды и плохо масштабируются на проекты с различной структурой моделей.

В России достаточно много решений, имеющих инструменты по автоматизации формирования чертежей, одним из которых является плагин «ModPlus»¹ для ПО «Revit». Данные инструменты реализуют rule-based подход, ориентированный на дублирование и настройку видов, но не предоставляют возможность по созданию видов для информационных моделей с учетом всех атрибутов и связей элементов.

В работе [Çeter, 2024] автор представляет результаты исследования нейросетевых алгоритмов GCN и GAT, основанных на графах, в смежной задаче расстановки аннотированных обозначений на планах зданий: размеров, выносок, меток. В своей работе S. Çeter для создания информационных моделей использует программу ArchiCAD и сформированное приложение ServCAD. По итогам тренировки моделей автор выделяет следующие результаты:

- В задачах простановки марок для дверей и помещений наибольшую точность в размере 91% показала модель с архитектурой GAT.
- В задаче простановки размеров для стен обе модели показали достаточно низкую точность. Наибольшую точность в размере 41% показала модель с архитектурой GCN.

Автор указывает на высокую перспективность и репрезентативность использования графовых нейросетей в задачах, связанных с информационным моделированием.

В результате анализа можно сделать вывод, что большинство rule-based инструментов для решения конкретных задач процесса формирования чертежей представляют собой традиционные алгоритмы. Чем сложнее задача, тем специализированнее становится разрабатываемый алгоритм. Также существует алгоритм на основе машинного обучения, позволяющий проставлять аннотированные обозначения на планах зданий. Продемонстрирована достаточно успешная применимость графовых нейросетевых алгоритмов в смежных задачах ТИМ-проектирования [Çeter, 2024]. Однако работ, связанных с ТИМ и нейронными сетями, на текущий момент достаточно мало, что указывает на присутствие множества ниш и большое поле для исследований в области подготовки строительной документации с использованием ТИМ и нейросетевых алгоритмов.

¹ URL: <https://modplus.org/ru>.

2. Формализация задачи и представление информационных моделей зданий

Информационные модели зданий имеют явную графовую структуру, в которой:

- Узлы (vertices, V) – элементы BIM-модели: стены, колонны, плиты.
- Рёбра (edges, E) – связи между элементами: примыкание, расстояние и другие метрики, пригодные для описания взаимодействия элементов [Austern et. al., 2024].

Математически, каждая отдельная модель здания представляется графом:

$$G = (V, E, X),$$

где $X \in \mathbb{R}^{|V| \times d}$ – матрица признаков узлов, d – размерность векторного описания элемента.

Особенность предлагаемого нейросетевого алгоритма заключается в необходимости его обучения на едином массиве данных. Формальное описание датасета будет представлять собой несвязный граф:

$$G' = \bigcup_{i=1}^M G_i,$$

где M – количество отдельных информационных моделей, а индексы узлов дополнены префиксами для предотвращения пересечения идентификаторов.

2.1 Формальная модель нейронной сети

Для решения поставленной задачи архитектура нейросети должна включать методы учета влияния соседних узлов, чтобы лучше запоминать окружение и присваивать степень «важности», так как, например, для формирования плана здания степень «важности» между конструкциями, находящимися на одном уровне, выше, чем степень «важности» между конструкциями, расположенными на разных этажах [Лунева и др., 2024].

Чтобы реализовать учет «важности», необходимо применить технологию множественного внимания (multi-head attention). Данный способ применяется в нейросетевых алгоритмах вида GAT. Принцип работы данного алгоритма заключается в дополнительном использовании независимых «голов», формирующих индивидуальные векторы весов влияния α_{ij} от i -узла к j -узлу посредством объединения признаков (рис.). После вычислений весов, необходимо объединение полученных векторов, которое выполняется по двум схемам: вычисление вектора со средними значениями по всем векторам внимания и конкатенация всех векторов внимания [Labonne, 2023].

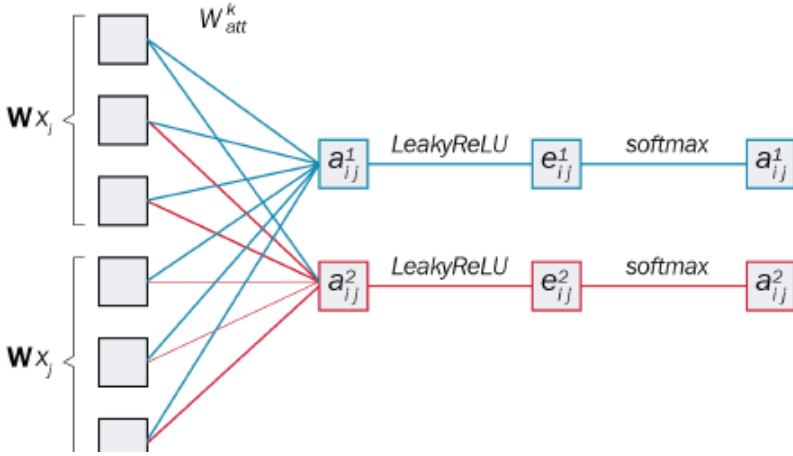


Рис. 1. Диаграмма вычисления коэффициентов внимания

В результате формула для вычисления коэффициентов α_{ij} внимания имеет следующий вид [Labonne, 2023]:

$$\alpha_{ij} = \frac{\exp(W_{att}^t \text{LeakyReLU}(W[x_i || x_j]))}{\sum_{k \in N_i} \exp(W_{att}^t \text{LeakyReLU}(W[x_i || x_j]))},$$

где имеет место:

- конкатенация признаков целевого узла и соседнего $W[x_i || x_j]$;
- применение функции активации LeakyReLU;
- умножение на матрицу весов внимания W_{att}^t ;
- softmax нормализация полученного результата.

Для множественного внимания с K «головами» формируется набор выходных признаков, которые содержат в себе данные не только об узле, но и о соседях и их влиянии:

$$h'_i = ||_{k=1}^K \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^k W^k x_i \right),$$

где σ – функция активации, а $||$ обозначает конкатенацию по всем «головам».

Стоит отметить, что от выбора количества «голов» зависит качество признаков на выходе, так как каждая из них учитывает разные зависимости узлов.

2.2. Подготовка данных

В первой итерации в качестве обучающей и тестовой выборки были созданы 6 информационных моделей жилых многоэтажных зданий, созданных в программном обеспечении (ПО) «Revit». Для формирования датасета были созданы похожие модели с небольшими различиями: количество этажей, расположение и количество конструкций на этаже, расположение лестнично-лифтовых узлов, присутствие/отсутствие стен/колонн/балок. Данные по моделям сведены в таблицу.

Таблица 1

№ модели	Кол-во этажей	Кол-во элементов	№ модели	Кол-во этажей	Кол-во элементов
1	21	1751	4	20	1152
2	21	1822	5	27	1925
3	20	1107	6	27	2005

С помощью специально разработанного инструмента были выгружены необходимые для построения графов данные в виде трех текстовых файлов.

В первом файле содержатся массивы уникальных идентификаторов элементов с уже присвоенными префиксами и данными о геометрическом расположении (рис.):

$$u_i = (id, c, x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max}),$$

где u_i – вектор признаков узла, c – категория узла, $x_{min}, y_{min}, z_{min}, x_{max}, y_{max}, z_{max}$ – координаты крайней левой нижней и крайней правой верхней точек описывающего элемент параллелепипеда.

```
[[
  [1751;-2000240;0;0;-9,84251968503937;0;0;-9,84251968503937],
  [24612;-2000240;0;0;0;0;0;0],
]]
```

Рис. 2. Пример выгруженных данных элементов

Во втором файле – массивы уникальных идентификаторов, между которыми необходимо построить ребра (рис.).

```
[[
  [509108;510268;510473;510588;510753;522444;522446;522448],
  [510268;510473;510753;515184;521567],
]]
```

Рис. 3. Пример выгруженных данных связей элементов

Третий файл датасета содержит элементы информационной модели (рис.). Такая структура обусловлена тем, что совокупность элементов позволяет определить расположение вида, а его категория – тип вида.

```
[ [
  [1752;1;509347;509461;509466;509471;509476],
  [24631;2;509347;509466;509471;509476;509599;509709],
]
```

Рис. 4. Пример выгруженных данных видов

Реализация алгоритма анализа информационных моделей выполнена в среде разработки `jupyter notebook`¹ на языке `python` с использованием библиотеки `networkx`². Для графического отображения графа была использована библиотека `matplotlib`³.

2.3. Формирование графовых представлений информационных моделей

Создание несвязного графа производится с использованием библиотеки `networkx`. В качестве входных параметров передается созданный экземпляр графа, данные о ребрах и узлах. Создание ребер выполняется по соответствующим данным, после чего для каждого полученного узла определяется вектор с данными об этом узле.

Графическое отображение полученного несвязного графа выведено при помощи функции `draw_networkx` модуля `networkx` (рис.).

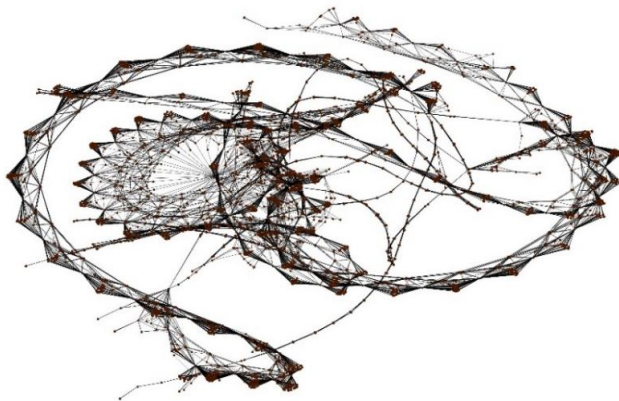


Рис. 5. Граф информационных моделей созданного датасета

¹ URL: <https://jupyter.org>.

² URL: <https://networkx.org>.

³ URL: <https://matplotlib.org>.

На представленном графе отражены пять информационных моделей жилых многосекционных зданий. Общее количество узлов составляет 9762. Также в структуре графа необходимо отметить явную сходимость нескольких узлов в одном, что говорит о наличии общего элемента, например перекрытии. С целью подготовки данных графа для обучения моделей применялся модуль `torch_geometric`¹.

3. Обучение нейросетевого алгоритма

3.1. Подготовка графа для обучения нейросети

В представленной работе для создания нейросети используются библиотеки `torch`² и `torch_geometric`. Для того, чтобы произвести обучение моделей, предоставляемых этими пакетами, необходимо подготовить данные по созданному графу, так как по умолчанию они не совместимы. Далее граф преобразуется в формат библиотеки `torch_geometric` функцией:

from_networkx(G', node_attributes, edge_attributes).

К каждому узлу добавлены три дополнительных параметра: `degree`, `clustering`, `pagerank`. Параметр `degree` хранит в себе количество связей с текущим узлом, параметр `clustering` – показатель кластеризации данного узла, параметр `pagerank` – важность данного узла среди остальных. Также в функцию добавлено масштабирование признаков с использованием класса `StandardScaler`, как часто используемый [Akhatov et. al., 2024].

После преобразования данных производится разделение на обучающую и тестовые выборки с использованием масок. В первой итерации проводилось разделение по соотношению 0.8 ед. к 0.2 ед., где 0.8 ед. – обучающая выборка, а 0.2 ед. – тестовая выборка.

3.2. Архитектура графовой нейросети

Разработанная нейронная сеть является сверточной с несколькими скрытыми слоями. Выбрана архитектура GAT, так как при помощи механизма внимания появляется возможность учитывать только важные соседние узлы, что необходимо при построении видов информационных моделей зданий. Диапазон скрытых слоев GATv2Conv выбран от 2 до 5. Также принято решение добавить функции нормализации данных и слой `dropout` с вероятностью 0.2 перед выходным слоем с целью предотвращения переобучения [Бабичева и др., 2021].

Критерием оптимизации выбран Adaptive Moment Estimation [Simon, 2024]. В качестве функции потерь была выбрана функция кросс-энтропии с весом 2, так как имеет место задача бинарной классификации.

¹ URL: <https://pytorch-geometric.readthedocs.io/en/latest/index.html>.

² URL: <https://pytorch.org>.

Стоит отметить, что в работе для каждого типа вида создается своя модель машинного обучения, так как для построения определенного вида необходимо учитывать узлы по разным критериям. Например, чтобы построить план, необходимо учесть расположение всех конструкций на этаже, а для фрагмента достаточно учитывать локальные зоны, но также важно проанализировать расстояние между конструкциями.

3.3. Обучение нейросети и анализ результатов

На первом этапе обучения количество перцептронов в скрытом слое было установлено в размере 64. Было установлено количество эпох обучения, равное 30. Также проведено обучение моделей с разным количеством скрытых слоев в диапазоне от 2 до 5 и разными массивами «голов». В результате получены графики потерь и точности для всех моделей, а также в работе представлены графики для моделей, предсказывающей планы и фрагменты планов, с 5 скрытыми слоями и для модели, предсказывающей разрезы, с 2 скрытыми слоями (рис. 6-8). Результаты тестирования точности моделей сведены в таблица.

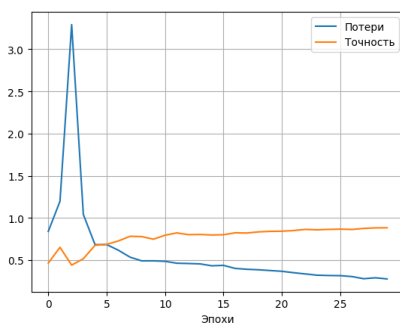


Рис. 6. Графики потерь и точности модели, предсказывающей планы

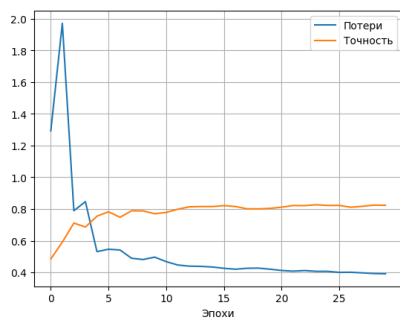


Рис. 7. Графики потерь и точности модели, предсказывающей разрезы

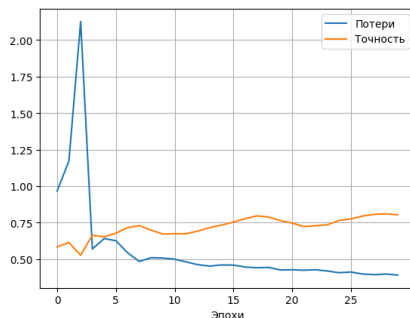


Рис. 8. Графики потерь и точности модели, предсказывающей фрагменты планов

По графикам модели планов заметно, что начиная с 20 эпохи точность растёт меньше и близка к 0.8, но ошибка продолжает существенно уменьшаться.

По графикам модели разрезов заметно, что, начиная с эпохи 10, точность и потери стабилизируются. Такое поведение связано с тем, что в моделях мало разрезов, поэтому модель запоминает только конкретные паттерны.

По графикам модели фрагментов планов заметно, что модель на протяжении всего обучения имеет высокую амплитуду точности. Это связано с тем, что модель не может найти нужный паттерн и в данном случае необходимо провести дополнительный анализ архитектуры нейросети.

По результатам, приведенным в таблица2, можно сделать следующие выводы:

1. Для модели, предсказывающей разрезы на данный момент достаточно 2 скрытых слоя нейронной сети.
2. Для моделей, предсказывающей планы и фрагменты плана, увеличение количества скрытых слоев приводит к увеличению точности.

Таблица 2

Тип вида	2 слоя	3 слоя	4 слоя	5 слоев
План	82.96%	79.81%	83.32%	87.05%
Разрез	79.59%	79.96%	80.83%	79.66%
Фрагмент плана	75.13%	81.49%	81.20%	83.32%

Заключение

В данной работе выполнен анализ существующих подходов и решений по формированию чертежей зданий. Предложен и реализован метод классификации элементов цифровых информационных моделей зданий по признаку видимости на различных типах видов, основанный на использо-

вании графовых нейронных сетей. Для проведения вычислительных экспериментов создан датасет в виде несвязного графа, содержащий множество подграфов – информационных моделей. Также в работе была реализована подготовка данных, включая извлечение топологических признаков (degree, clustering, pagerank) и нормализацию признаков.

Разработаны, обучены и протестированы модели графовых нейронных сетей с разным количеством слоев GATv2Conv от 2 до 5. Наилучшие результаты (87.05% и 83.32% соответственно) для планов и фрагментов планов показали модели с пятью скрытыми слоями. Для модели, предсказывающей разрезы на данный момент достаточно двух скрытых слоя.

Выявлены ограничения, связанные с недостатком объема данных для некоторых типов видов (фасады, фрагменты-сечения), а также нестабильностью обучения модели для фрагментов планов, что свидетельствует о необходимости расширения датасета и адаптации архитектуры под каждый отдельный тип вида. Также необходимо провести более детальное тестирование применяемых функций потерь и алгоритмов оптимизации.

Список литературы

- [Austern et al., 2024] Austern G., Bloch T., Abulafia Y. Incorporating Context into BIM-Derived Data—Leveraging Graph Neural Networks for Building Element Classification // *Buildings*. – 2024. – Vol. 14(2). – P. 527. – DOI: <https://doi.org/10.3390/buildings14020527>.
- [Çeter, 2024] Server Çeter Automated BIM Model Annotation via Graph Neural Networks: Bridging the Gap Between Design and Documentation: Master Thesis for the Master of Science Program in Civil Engineering / TUM School of Engineering and Design, 2024. – P. 89.
- [Kim et al., 2024] Kim Y., Chin S., Choo S. Rule-based automation algorithm for generating 2D deliverables from BIM // *Journal of Building Engineering*. – 2024. – Vol. 97. – P. 111033. – DOI: <https://doi.org/10.1016/j.jobe.2024.111033>.
- [Лунева и др., 2024] Лунева Е.Е., Банокин П.И. Графовые нейронные сети в задачах продленной аутентификации групповых профилей социальных сетей // *System Analysis and Mathematical Modeling*. – 2024. – Т. 6, № 3. – С. 300-311. – DOI 10.17150/2713-1734.
- [Labonne, 2023] Maxime Labonne Hands-On Graph Neural Networks Using Python: Livery Place, 35 Livery Street, Birmingham, B3 2PB, UK, Packt Publishing, 2023.
- [Akhatov et al., 2024] Akhatov A.R., Renavikar A., Rashidov A.E.O., Nazarov F.M. Optimization of the number of databases in the big data processing// *Проблемы информатики*. – 2023. – No. 1(58). – P. 33-47. – DOI 10.24412/2073-0667-2023-1-33-47.
- [Бабичева и др., 2021] Бабичева М.В., Данилов В.В., Борщевский С.В. Вероятностный подход к оптимизации нейронных сетей случайной редукцией нейронов // *Вестник Донецкого национального университета. Серия Г: Технические науки*. – 2021. – № 1. – С. 63-70.
- [Simon, 2024] Simon J.D. Prince Understanding Deep Learning. 255 Main Street, 9th Floor, Cambridge, MA 02142 MIT Press, 2024.

УДК 004.8

doi: 10.15622/rcai.2025.095

АЛГОРИТМЫ ИДЕНТИФИКАЦИИ КЛАССА В ЗАДАЧЕ РАСПОЗНАВАНИЯ КЛАВИАТУРНОГО ПОЧЕРКА С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

Д.Н. Кобзаренко (*kobzarenko_dm@mail.ru*)^{A,B}

А.Г. Мустафаев (*arslan_mustafaev@mail.ru*)^{A,B}

^A Дагестанский государственный институт народного хозяйства,
Махачкала

^B МИРЭА – Российский технологический университет, Москва

В рамках научных исследований по идентификации личности на основе клавиатурного почерка с использованием современных методов искусственного интеллекта, получены экспериментальные результаты. На основе новых разработанных алгоритмов получен максимальный результат точности распознавания классов на тестовых выборках. Результаты показывают высокую эффективность выбранных моделей, методов обработки исходных и алгоритмов идентификации класса в задаче распознавания личности по клавиатурному почерку.

Ключевые слова: клавиатурный почерк, машинное обучение, задача классификации, CatBoost, Python.

Введение

Отечественные работы относительно формулировки задачи идентификации личности по клавиатурному почерку в подавляющем большинстве носят теоретико-гипотетический характер без каких-либо описаний конкретных технических решений и проблем, возникающих в контексте рассматриваемой задачи. Авторами современных публикаций предполагается, что задача идентификации клавиатурного почерка может быть полезной в различных областях, таких как кибербезопасность, криминалистика, психология и др. Коротко остановимся на некоторых работах.

В работе [Сатыбалдиева, 2024] приведены общие сведения о том какие методы в основном используются для подобных задач в процентном соотношении и методы машинного обучения с нейронными сетями не на первом месте, что неудивительно, поскольку обзор соответствует 2000-м

годам согласно приведенному списку литературы. В современных же исследованиях подобного рода (классификация и регрессия) машинное обучение и/или нейронные сети вытеснили старые методы и алгоритмы, и являются инструментами номер 1.

Имеется готовое техническое решение подобной задачи в виде свидетельства программы для ЭВМ [Лунев, 2024], в описании которого программа позиционируется как средство, позволяющее собрать статистику и провести аутентификацию пользователя по поведенческим паттернам, свойственным каждому человеку. Оцениваются параметры, скорость и динамика набора текста, время перехода между клавишами и опечатки. Более конкретной информации (какие используются методы, алгоритмы, точность прогноза и т.д.), кроме того, что основа программы написана на Python по данному свидетельству нет, как отсутствуют и публикации, связанные с ним. Поэтому сравнивать наши результаты с приведенным программным обеспечением невозможно отсутствует.

В работе [Варламова, 2023] используется и анализируется (без машинного обучения) так называемый оценочный параметр времени удержания клавиш для определения пользователя и утверждается, что время удержания клавиш способно идентифицировать пользователей в 85% случаев. Причем в конце статьи поясняется, что суть исследования состоит в идентификации по типу «свой»-«чужой», что соответствует более простому варианту прогнозной модели – бинарной классификации.

Разработка абстрактной модели искусственной иммунной сети и ее использование для распознавания образов клавиатурного почерка описана в [Сулавко, 2020]. Здесь подробно рассматриваются математические и технические аспекты построения данной модели, ключевыми особенностями которой являются то, что система также выполняет классификацию по принципу «свой»-«чужой», и основана на однородных исходных данных (обучающие и тестовые данные вводятся в виде одной и той же небольшой фразы многократным повторением N раз, кроме того, данная фраза должна быть введена безошибочно, иначе пример не заходит в базу).

Мировой опыт решения подобной задачи состоит из подробного рассмотрения и сопоставления авторами различных вариантов моделей как классического машинного обучения наряду с нейронными сетями [Sharma et al., 2023], так и исключительно нейронных сетей различных модификаций и функций активации [Acien et al., 2021]. А с учетом того, что в задаче распознавания клавиатурного почерка в целом не удастся приблизиться к 100% результату, в последнее время решили совместить клавиатурный почерк с движением мыши [Wang et al., 2023], что несомненно дает свои преимущества с точки зрения качества обучения модели, поскольку число независимых признаков возрастает.

Исходя из опыта наших предыдущих исследований, мы установили, что перебор архитектур нейронных сетей (DNN, Conv1D, LSTM), по сравнению с другими эффективными моделями машинного обучения (XBoost, CatBoost) не дает какого-либо прироста в точности прогноза и даже порой уступает последним. Поэтому повышение эффективности в решении задачи нам представляется не в поиске оптимальной архитектуры модели и параметров (что конечно тоже важно, но не дает хорошего прироста), а в работе с исходными данными до их поступления в модель и в постобработке результатов выхода модели с целью принятия решения о прогнозе. Для этого нами разработаны рассматриваемые далее оригинальные алгоритмы принятия решения по принадлежности пула примеров, сгенерированных на основе набора на клавиатуре нескольких предложений, тому или иному объекту.

Основные характеристики задачи идентификации клавиатурного почерка в нашей концепции ее реализации заключаются в следующем:

- задача позиционируется как мультиклассовая классификация (на выходе нужно идентифицировать владельца клавиатурного почерка);
- текст, вводимый субъектом для генерации обучающего и тестового набора примеров неоднородный, не повторяющийся и не пересекающийся;
- при вводе текста субъекту не ставятся никаких ограничений (допускается свободное редактирование, исправления ошибок и т.д.), за исключением разумного лимита времени достаточного для ввода текста даже для неопытного пользователя ПК;
- признаки классификации формируются только на основании локальной динамики набора текста (именно этот параметр является существенным, что установлено предыдущими нашими экспериментами);
- задача решается на основе моделей машинного обучения, для идентификации класса используется алгоритм, который принимает решение на основе пула тестовых примеров (сгенерированных на основе набранного фрагмента текста), пропущенных через модель.

От алгоритма идентификации класса на основе пула примеров, пропущенных через модель, зависит точность распознавания. Данная работа посвящена экспериментальному изучению этого вопроса с разработкой таких алгоритмов и оценкой их точности.

1. Постановка задачи идентификации клавиатурного почерка на основе машинного обучения

Участников испытаний (число классов) 35 человек. Испытуемым предложено пройти процедуру снятия клавиатурного почерка в комфортных условиях, на любой удобной клавиатуре (мембранного типа), в том числе и на ноутбуке.

Испытуемый вводит осмысленную текстовую информацию, разделяемую на небольшие порции (по 4-6 строк), называемые *эксперимент 1, эксперимент 2, ..., эксперимент N*. Допускается перерыв любой длительности между экспериментами (чтобы не накапливалась усталость), но во время набора текста (тайпинга) в рамках одного эксперимента, испытуемый должен полностью концентрировать свое внимание на данной процедуре. В среднем на 1 эксперимент тайпинга уходит 1-2 минуты времени.

Задача испытуемого состоит в точном повторении шаблона текста, высвечиваемого на экране. При этом он может ошибаться, стирать символы, но для удачного завершения очередного эксперимента необходимо в точности повторить шаблон текста. Для защиты от случайной или специальной затяжки времени устанавливается лимит прерывающий эксперимент, который в этом случае необходимо начать заново. Работа считается законченной тогда, когда выполнены все N экспериментов.

Количество экспериментов (этапов) принято равным $N = 40$ (содержимое этапов одинаковое для всех испытуемых). Результаты экспериментов заносятся в таблицу с полями: тип операции (d – нажатие клавиши / u – отжатие клавиши), код клавиши, время операции (до миллисекунды). Количество исходных записей для одного испытуемого на основе 40 этапов набора текста оценочно составляет 15000 (плюс-минус 3000) записей, в зависимости от того насколько точно он повторяет шаблонный текст без исправлений.

Приложение для снятия клавиатурного почерка разработано нами ранее и его работа описана в [Мустафаев, 2023].

В качестве модели данных для формирования окончательной таблицы примеров (датасета) принята модель на основе трех признаков – временных разностей t_1 , t_2 , t_3 (в миллисекундах) между четырьмя ближайшими событиями из исходной базы данных. География клавиш (на основе кодов и их расположения) не учитывается, поскольку она не улучшает качества прогноза, что установлено в предыдущих экспериментах. Условия для принятия примера в датасет следующие:

1. Все четыре события должны быть исключительно от 32 клавиш клавиатуры (кроме «ё») либо от клавиши точки «.», поскольку именно последовательность буквенных клавиш составляет основу уникальности тайпинга. Можно было бы добавить клавишу «пробела», которая часто используется, но предыдущие наши эксперименты показали, что такой вариант не улучшает прогнозные свойства.

2. Временной интервал между четырьмя событиями не должен превышать 2 секунды. Суть этого условия в том, чтобы исключить цепочки действий, в которых пользователь мог отвлекаться. Другие возможные значения этого интервала пока не исследованы.

3. В [Мустафаев, 2024] нами установлен факт того, что вся уникальная информация о динамике набора текста для субъекта (для четырех событий) содержится в последовательности событий вида: d(нажатие)-u(отжатие)-d(нажатие)-u(отжатие). Остальных 15 вариантов последовательностей событий сравнительно немного, и они отбрасываются, как неинформативные.

Последний параметр результирующего датасета с примерами – `sf` отвечает за тип выборки: 0 – обучающая, 1 – тестовая № 1, 2 – тестовая № 2. Обучающая выборка примеров генерируется на основе экспериментов с 1 по 38, 39-й эксперимент – основа для тестовой выборки №1 и 40-й эксперимент – для тестовой выборки №2.

В качестве модели машинного обучения для решения задачи на данном этапе исследований принята модель градиентного бустинга, реализуемая с помощью библиотеки CatBoost (<https://catboost.ai/>). Для табличных данных (всего с тремя признаками) такая модель вполне достаточна и эффективна.

Входными данными являются массив X векторов с 3-мя признаками (нормализованными в интервале $[0, 1]$) и массив Y указаний учителя – номера классов. При прогнозировании класса через обученную модель пропускается пул тестовых примеров X в количестве n , полученных на основе набора текста одного из субъектов. На выходе снимается матрица R вероятностей принадлежности примеров одному из классов размерностью $(n, 35)$. Далее на основе матрицы R требуется определить номер класса субъекта. Для этого и необходим алгоритм идентификации.

Два алгоритма идентификации приведены нами в [Мустафаев, 2024], но, как показала практика, они недостаточно эффективны. В данной работе разработаны и исследованы три дополнительных алгоритма. Для исследования точности алгоритмов в среде Google Colaboratory разработан ноутбук с настройкой опций тестирования и визуализирующий результаты.

2. Алгоритмы идентификации субъекта клавиатурного почерка

Приведем пять алгоритмов идентификации, обозначим их латинскими буквами «A», «B», «C», «D», «E».

Примем следующие обозначения:

- На вход подается матрица вероятностей от выходов модели $R[n, m]$, где n – количество примеров от испытуемого, m – число классов.
- Требуется идентифицировать испытуемого – найти номер класса S .

2.1 Алгоритм «A»

Искомый класс идентифицируется по максимальному количеству примеров, которые могут принадлежать классу исходя из максимальной вероятности. В таком алгоритме классов-победителей может быть несколько, поскольку оценка целочисленная.

Для каждого i -го из n примеров, находим столбец j (номер класса) соответствующий максимальному значению вероятности и записываем номера классов в вектор $P_{\max}[n]$. На основе полученного вектора P_{\max} рассчитываем вектор $L[m]$, значение i -го элемента которого соответствует количеству найденных элементов со значением i в векторе P_{\max} . Результирующий номер класса S вычисляется как позиция i максимального элемента вектора L .

2.2. Алгоритм «В»

Находятся суммы вероятностей для всех примеров по классам, побеждает класс с максимальной суммой вероятности. Здесь критерий вещественный, поэтому шансы на то, что будет больше одного победителя с одной и той же суммарной вероятностью практически равны нулю.

Для входной матрицы $R[n, m]$ находятся суммы столбцов и результат заносится в вектор $P_{\text{sum}}[m]$. Результирующий номер класса S вычисляется как позиция i максимального элемента вектора P_{sum} .

2.3. Алгоритм «С»

Начиная с этого алгоритма и далее вычисления проводятся в два этапа, поэтому можно их условно назвать двухэтапными. Здесь и далее вводится целочисленная переменная, которая может принимать значение от 2 и теоретически до m (количество классов), обозначим ее через w .

На первом этапе находится вектор P_{sum} как в алгоритме «В». Далее на основе вектора P_{sum} определяется список номеров классов $F[w]$ с наибольшими вероятностями в количестве w . Назовем эти классы пулом победителей.

На втором этапе во входной матрице $R[n, m]$ обнуляются столбцы классов, номера, которых отсутствуют в $F[w]$. Оставшиеся столбцы определяют победителя (или победителей при одинаковых количествах побед в примерах) по алгоритму «А», соревнуясь только в рамках пула победителя.

2.4. Алгоритм «D»

Первый этап аналогичен первому этапу в алгоритме «С» и определяет пул победителей $F[w]$.

На втором этапе на основе пула классов победителей генерируются соревновательные пары классов каждый с каждым. Например, если список F содержит классы $[2, 5, 10]$, то имеем три соревновательные пары (2 против 5, 2 против 10, 5 против 10). Для каждой соревновательной пары из входной матрицы $R[n, m]$ выделяются два соответствующих столбца вероятностей и по алгоритму «А» выясняется победитель в этой конкретной паре. В копилку класса-победителя пары заносится 2 балла, проигравшему 0 баллов, если ничья (количество побед в примерах поровну), то оба класса получают по 1 баллу.

В конце алгоритма подсчитываются баллы, набранные классами на втором этапе в парных соревнованиях и определяется победитель по наибольшему баллам (или победители).

2.5. Алгоритм «Е»

Этот алгоритм аналогичен алгоритму «D» за исключением подсчета баллов на втором этапе в парных соревнованиях. Подсчет баллов следующий: в копилку класса добавляются баллы по количеству примеров, в которых класс победил и отнимаются баллы по количеству примеров, в которых класс проиграл.


3. Оценка эффективности алгоритмов идентификации


На основе разработанных алгоритмов реализованы программные коды, а также функции отчета результатов идентификации на тестовых примерах.

Имеется два набора тестовых примеров «тест-1» и «тест-2», в каждом наборе представлено по одному пулу примеров для каждого класса. Таким образом, в двух тестах содержатся 70 пулов примеров, по 2 на каждый класс. Точность прогноза определяется как частное от деления количества правильно идентифицированных классов на 70. Если при идентификации победителей больше одного, то считаем, что класс не распознан.

Кроме основного обучающего массива данных (True_Train) с количеством примеров 82073, с помощью библиотеки tabgan (<https://pypi.org/project/tabgan/>), сгенерированы массивы синтетических обучающих данных на основе генератора OriginalGenerator (OriginalGenerator_Train) с количеством примеров 245075 и генератора GANGenerator (GANGenerator_Train) с количеством примеров 117787, которые также участвуют в эксперименте. На основе трех массивов обучены три модели типа CatBoostClassifier. Обращаем внимание, что обучение необходимо выполнить в режиме CPU (без видеоускорителя), поскольку установлено, что в режиме GPU (с видеоускорителем) результаты ухудшаются. Данный факт не связан с особенностями решаемой задачи, причина его пока не ясна, у других разработчиков происходит тоже самое, что подтверждается форумом (<https://github.com/catboost/catboost/issues/1408>).

Для выбора опций тестирования разработан набор виджетов, в котором выбирается модель, алгоритм идентификации, тестовый набор и количество победителей первого этапа (параметр w для алгоритмов «C», «D», «E») (рис. 1).


Показать код



Модель:

True_Train

▼

Алгоритм:

E

▼

Тест:

тест-2

▼

Кол-во поб.:

3

▼

Выполнить тест

Рис. 1. Набор опций виджетов для тестирования в блокноте Google Colaboratory

В табл. 1 сведены результаты тестирования всех моделей, алгоритмов и тестовых наборов при $w = 3$. Такое значение w выбрано исходя из того, что отчет о тестировании по сумме вероятностей на алгоритме «В» для одного из нераспознанных классов показал, что по значению суммы вероятностей этот класс вошел в тройку, а остальные нераспознанные классы вошли в двойку лучших. Таким образом, чтобы не потерять искомый класс в алгоритмах «С», «D», «Е» минимальное значение w для нашего случая требуется взять равным 3.

Таблица 1

Результаты тестирования алгоритмов (при $w = 3$)

Алгоритм / тест	Модель True_Train		Модель OriginalGenerator		Модель GANGenerator	
	Нераспоз. классы	Точн.	Нераспоз. классы	Точн.	Нераспоз. классы	Точн.
«А» тест-1	2, 11, 14, 28, 29, 34	0,814	2, 14, 28, 29, 34	0,871	2, 9, 11, 14, 28, 29, 34	0,757
«А» тест-2	1, 4, 9, 25, 29, 33, 34		4, 9, 33, 34		1, 2, 4, 9, 10, 11, 18, 29, 33, 34	
«В» тест-1	2	0,938	-	0,943	14	0,914
«В» тест-2	4, 9, 33, 34		4, 9, 33, 34		2, 9, 11, 33, 34	
«С» тест-1	-	0,943	-	0,971	2, 29	0,900
«С» тест-2	9, 31, 33, 34		9, 33		1, 2, 9, 10, 33	
«D» тест-1	-	0,971	-	0,985	11, 29	0,943
«D» тест-2	9, 34		9		10, 22	
«Е» тест-1	-	0,985	-	0,985	29	0,943
«Е» тест-2	9		9		10, 22, 29	

Результаты тестирования при $w=3$ показывают, что алгоритм «Е» при всех вариантах обучения показывает лучшие результаты, которые состоят в том, что из 70 пулов примеров нераспознанным остается только один. Также видно, что модель, обученная на датасете GANGenerator существенно уступает другим двум моделям, из чего следует, что данный генератор не совсем подходит к решению текущей задачи.

В табл. 2 представлены результаты экспериментов для w меняющегося от 3 до 7 для алгоритма «Е» (который признан лучшим из табл. 1). Они показывают, что при $4 \leq w \leq 7$ модель, обученная на датасете OriginalGenerator дает 100% результат в распознавании классов.

Таблица 2

Результаты тестирования алгоритма «Е» (при $w = [3 \dots 8]$)

w / тест	Модель True_Train		Модель OriginalGenerator		Модель GANGenerator	
	Нераспоз. классы	Точн.	Нераспоз. классы	Точн.	Нераспоз. классы	Точн.
3 / тест-1	-	0,985	-	0,985	29	0,943
3 / тест-2	9		9		10, 22, 29	
4 / тест-1	-	0,985	-	1,000	29	0,957
4 / тест-2	31		-		22, 29	
5 / тест-1	-	0,985	-	1,000	29	0,957
5 / тест-2	31		-		22, 29	
6 / тест-1	-	0,971	-	1,000	29	0,957
6 / тест-2	22, 31		-		22, 29	
7 / тест-1	-	0,971	-	1,000	29	0,943
7 / тест-2	22, 31		-		3, 22, 29	
8 / тест-1	-	0,971	-	0,985	29	0,914
8 / тест-2	22, 31		22		1, 3, 22, 29, 31	

Заключение

В итоге, можно сделать вывод о том, что максимальная эффективность в решении задачи идентификации клавиатурного почерка машинным обучением достигается применением алгоритма «Е» и генерации обучающих примеров модели с помощью генератора OriginalGenerator библиотеки tabgan. При этом требует дополнительного изучения вопрос как подбирать оптимальный параметр w и как он зависит от количества классов в системе, что планируется сделать в дальнейшем.

На данный момент экспериментальная база насчитывает 35 субъектов исследования (база пополняется и далее будет расширена). Это не много, чтобы быть полностью уверенными в надежности разработанного подхода с точки зрения практического использования, но и не мало, поскольку в других подобных исследованиях их в среднем в интервале 50-150. Кроме

того, следует отметить, что даже при 35 классах, без рассмотренных оригинальных алгоритмов точность распознавания на тестовой выборке (без синтетического обогащения) составляет 93,8%, а с алгоритмом «Е» 98,5%, эта разница очень существенна и при переборе моделей ИИ с гиперпараметрами такую разницу не получить.

При практическом применении всех текущих наработок для построения реальной системы идентификации личности по клавиатурному почерку нужно учитывать факт того, что данный вид биометрии основан на приобретенной координации движений пальцев, которая со временем может меняться, поэтому в реальной системе необходимо выполнять переобучение модели на основе новых актуальных данных тайпинга. Возможно также внедрить систему механизм согласно которому система каждый раз при идентификации личности будет давать оценку на сколько легко распознается тайпинг субъекта и давать рекомендации по переобучению, если его почерк становится все больше не похож на оригинал.

Влияние на клавиатурный почерк психофизического состояния оценить сложно, возникает вопросы как оценить сиюминутного состояния человека при тайпинге или как искусственно ввести его в как-либо состояние для сбора исходного материала исследования? В данном случае мы исходим из того, что клавиатурный почерк это приобретенный навык, связанный с координацией, который выполняется на бессознательном уровне подобно управлению автомобилем, а также из того, что в процессе идентификации пользователь просто отложит процедуру в случаях нестандартных ситуаций, влияющих на его психофизическое состояние.

Список литературы

- [Варламова, 2023] Варламова С.А., Вавилина Е.А. Идентификация пользователя на основе клавиатурного почерка // Инновационное приборостроение. – 2023. – Т. 2, № 3. – С. 67-71. – doi: 10.31799/2949-0693-2023-3-67-71.
- [Лунев, 2024] Свидетельство о государственной регистрации программы для ЭВМ № 2024684582 Российская Федерация. Программа анализа клавиатурного почерка KBoardTrack: № 2024683137: заявл. 08.10.2024 : опубл. 18.10.2024 / А.А. Уксусова, Д.Л. Перчаткин, А.К. Хамитов; заявитель ООО "Системы нейробезопасности".
- [Мустафаев, 2023] Мустафаев А.Г., Кобзаренко Д.Н., Паштаев Б.Д. Нейросетевая модель идентификации клавиатурного почерка: датасет, архитектура, метрика // Промышленные АСУ и контроллеры. – 2023. – № 10. – С. 34-41. – doi: 10.25791/asu.10.2023.1466.
- [Мустафаев, 2024] Мустафаев А.Г., Кобзаренко Д.Н., Шихсаидов Б.И. Способы повышения точности в задаче идентификации клавиатурного почерка нейросетевой моделью // Промышленные АСУ и контроллеры. – 2024. – № 5. – С. 3-13.
- [Сатыбалдиева, 2024] Сатыбалдиева М.М. Исследование систем для идентификации пользователя на основе анализа клавиатурного почерка // Научный аспект. – 2024. – Т. 14, № 5. – С. 1897-1903.

- [Сулавко, 2020] Сулавко А.Е. Абстрактная модель искусственной иммунной сети на основе комитета классификаторов и ее использование для распознавания образов клавиатурного почерка // Компьютерная оптика. – 2020. – Т. 44, № 5. – С. 830-842. – doi: 10.18287/2412-6179-CO-717.
- [Acien et.al., 2021] Acien A., Morales A., Monaco J.V., Vera-Rodriguez R. TypeNet: Deep Learning Keystroke Biometrics. JOURNAL OF LATEX CLASS FILES. – 2021. – Vol. 14, No. 8. – <https://doi.org/10.48550/arXiv.2101.05570>.
- [Sharma et al., 2023] Sharma A., Jureček M., Stamp M. Keystroke Dynamics for User Identification // Computer Science (Machine Learning). – 2023. – <https://doi.org/10.48550/arXiv.2307.05529>.
- [Wang et al., 2023] Wang X., Shi Y., Zheng K., Zhang Y., Hong W., Cao S. User Authentication Method Based on Keystroke Dynamics and Mouse Dynamics with Scene-Irrelated Features in Hybrid Scenes // Sensors. – 2022, – Vol. 22(16), 6627. – <https://doi.org/10.3390/s22176627>.

УДК 004.8

doi: 10.15622/rcai.2025.096

К ВОПРОСУ О ПСИХОФИЗИОЛОГИЧЕСКИХ ХАРАКТЕРИСТИКАХ В ЧЕЛОВЕКО-МАШИННЫХ СИСТЕМАХ

М.Н. Королева (*maria.svyatkina@gmail.com*)

Московский государственный технический университет
им. Н.Э. Баумана, Москва

Человеко-машинные системы полуавтоматического или автоматического типов могут и должны учитывать особенности человека, и адаптироваться к нему. В работе предлагается рассматривать оператора в контуре системы также и в виде объекта управления (для возможностей адаптации системы). Исходя из этого, входной информационный поток должен быть подвергнут преобразованиям соответствующим целям и критериям управления. Это преобразование должно основываться на таком компоненте как модель объекта управления (оператора). Важной составляющей модели являются его психофизиологические характеристики. Такой подход характерен для широкого спектра человеко-машинных систем от систем управления технологическими процессами до выстраивания индивидуальных образовательных траекторий. Важнейшим элементом преобразования в этой работе рассматривается такой компонент как транслятор ключевых характеристик модели в форму подачи информационных единиц или их совокупности. Приведен пример изменения формы представления дидактических единиц на основе учета темперамента обучающихся.

Ключевые слова: человеко-машинные системы, психофизиологические характеристики, модель мира, объект управления, системы управления, оператор.

Введение

Говоря о современном развитии интеллектуальных систем нельзя не отметить системы фактически ставшие помощниками и ассистентами людей. Например, в части планирования дня и подготовки контента. Продвинутые пользователи уже обучили модели GPT под себя и успешно используют их для помощи в работе (конечно, не теряя здравый смысл).

Что же касается производственных систем, где нет возможности так рутинно обучать системы, мы должны говорить о разработке адаптируемых интеллектуальных систем к пользователю. То есть необходимо заложить в систему возможность “понимания”, кто такой оператор (задать его профиль в картину миру системы), чтобы система могла обучиться самостоятельно, накопив и проанализировав действия пользователя в зависимости от контекста и ситуаций. Это означает, что в дальнейшем система будет понимать кто сейчас является оператором (условно Иванов И.И. или Смирнов С.С.), как конкретно этот оператор действует обычно, какие решения ему свойственны и даже какие стратегии он применяет.

1. Анализ существующих подходов и работ

О необходимости учета личных данных и личностных характеристик говорят во многих областях: в области информационной безопасности при социо-инженерных атаках [Azarov et. al., 2020], в финансовом секторе при покупке и продаже финансовых инструментов [Стинбарджер, 2013], в эвакуационных моделях движения людей [Киселев, 2020], в вопросах безопасности труда [Муштонина, 2021], но чаще всего в сфере образования.

1.1. Адаптируемые системы в образовательной сфере

В образовательном процессе строят модель обучаемого – информационную модель, содержащую некоторую информацию о пользователе учебного курса и использующую ее для адаптации процесса работы в системе к этому пользователю. Модель предназначена для адаптации процесса обучения и привязана к модели предметной области (учебного курса).

В разных системах состав модели обучаемого различен. Так, по [Атанов, 2001], [Дорофеев и др., 2007] она представляет собой модель учебных целей, на которую наложены динамически изменяющиеся результаты, хранящиеся в треке обучения. Модель обучаемого предназначена для оценки знаний на определенный момент времени: она доступна как преподавателю, так и учащемуся. Более развёрнутые описания [Давыдова и др., 2015] могут включать:

- коммуникативную модель – данные, необходимые для однозначной идентификации студента, обеспечивающие общение с ним;
- психолого-когнитивную модель – данные для построения процесса обучения индивидуума в соответствии с его потребностями и возможностями: когнитивная, регулятивно-деятельностная и эмоционально-волевая компоненты;
- модель достижений – информация о профессиональных и общекультурных компетенциях, исходных и приобретённых в процессе обучения.

Модель курса должна взаимодействовать с моделью обучаемого для планирования учёбы и её специализации. В модель курса могут включаться [Давыдова и др., 2015] модель требуемых знаний (информация о том, чему необходимо научить обучаемых), план курса и модель учебных материалов, включающая понятия «учебные элементы», «контрольные вопросы» и т.д. Хотя адаптацию часто называют её обязательным элементом [Шаплыко, 2020], она часто понимается весьма узко – как формальный учёт пройденного материала в информационной системе. Однако известны и разработки с развитыми средствами адаптации [Рыбина, 2017], в том числе с индивидуальным планированием методики изучения курса на базе онтологий курсов и выявления проблемных зон в знаниях и умениях обучаемых. Также этот подход допускает оптимизацию решений с учётом психологического портрета обучаемого.

Хотя основные элементы интеллектуальной обучающей системы уже описаны [Добрица и др., 2019], практическое их использование до сих пор весьма ограничено. Такая обучающая система должна разрабатываться с применением методов когнитивной науки, компьютерной лингвистики, искусственного интеллекта и математики, что затрудняет как разработку, так и проверку эффективности.

Среди типичных задач обучающей системы указывают построение последовательности изучения курса, интеллектуальный анализ решений задач, помощь в решении при необходимости и мониторинг процесса обучения. Известные зарубежные интеллектуальные адаптивные платформы: Loud Cloud, Blackboard, Knewton, RealizeIT, Geekie, Smart Sparrow, TopClass – обеспечивают главным образом лишь мониторинг учебного процесса, а прочие интеллектуальные элементы разработчикам курса требуется добавлять самостоятельно.

Примером российских технологических и методологических решений для адаптивного обучения является совместный проект НИТУ «МИСиС» и Томского университета систем управления и радиоэлектроники (ТУСУР) [Кречетов и др., 2018]. В основе модели адаптации лежат классические модели кривой забывания и итеративного подхода к обучению. Результатом работы алгоритмов является множество траекторий достижения запланированных результатов обучения, только одна из которых предъявляется для освоения. Это показывает, что внедрение подобных систем в текущих условиях возможно и актуально.

1.2. Потребность в адаптируемых системах в области АСУТП

В настоящее время наблюдается кадровый кризис для областей связанных с управлением технологическими процессами – нет возможности выбирать подходящих по темпераменту и другим психическим характеристикам операторов, так как нет избыточности желающих. Однако с по-

мощью интеграции различных подходов и моделей, как то семантические сети, онтологии, модели генеративного ИИ, большие языковые модели, а также экспертные знания и оценки в области информационных интеллектуальных технологий и психологии мы можем адаптировать не только операторов к системам, но и системы к ним. Фактически мы наконец пришли к необходимости использовать уже апробированные в образовательной области технологии, дополнять их и развивать новый класс информационных систем адаптируемых к пользователю на производстве. Стоит отметить, что о создании подобных систем, говорят и в мире. Так, например, в работе [Cardin et al., 2022] говорится о проектировании и использовании цифровых двойников человека-оператора в промышленных киберфизических системах, какие характеристики необходимо «снимать» с оператора, и подробно рассмотрен этический вопрос.

Мы предлагаем пойти несколько дальше и поговорить о возможности решения встающей перед нами задачи построения модели субъекта, чтобы система могла «понимать», кто с ней работает и какие корректировки информационного потока со своей стороны она должна предоставлять, имея в виду основную целевую функцию. Модель субъекта должна включать в себя такие психофизиологические характеристики, как: темперамент, возможность к риску, смене парадигмы поведения, то есть также иметь когнитивную, регулятивно-деятельностную и эмоционально-волевою составляющие. Адаптация как раз завязана на учет психофизиологических характеристик.

2. Модель субъекта в общей структуре системы

Рассмотрим общую схему взаимодействия / информационных потоков между сущностями адаптируемой системы, представленной на рис. 1. Субъектом деятельности выступает человек (оператор или обучающийся). На вход поступает не только общий поток информации, но и оценка ситуации или состояния системы, что хорошо или плохо (оценка результата деятельности, для обучения – срез знаний или успеваемость, для технологических процессов – оценка ситуации или режима работы), а также меняется входное воздействие на человека. То есть предоставляемая информация в человеко-машинных системах должна быть адаптируемой под индивидуальные особенности – текстовое описание, визуальные образы и аудио сигналы для разных людей должны быть разной интенсивности и эмоциональной окрашенности. Для случая обучающих систем это может быть изменение формы подачи материала, последовательности (изменение программы или образовательной траектории). Для случая управления технологическими процессами – изменение интенсивности подачи информации в зависимости от психофизиологических особенностей оператора, его восприимчивости к информации. Также выбор корректирующего

воздействия или подстройка системой должны осуществляться с точки зрения итоговой цели работы системы (например, сохранение течения процесса в оптимальном режиме) и критериев деятельности (показатели по качеству и цели управления).

Рис. 1. Общая схема

Для адаптируемости система в своем контуре должна иметь модель субъекта. Эта модель может быть многогранной и должна включать в себя следующие аспекты:

- Фактические данные: фео, возраст, пол, профессия, имеющиеся компетенции, семейное положение и наличие детей, имеющиеся хронические заболевания;
- Поведенческая модель: параметры личностной тревожности и склонность к риску, стратегии принятия решений (проверка гипотез, выжидательная тактика или быстрое принятие решений) [Дозорцев и др., 2022], возможность смены парадигмы действий, здесь же настройки подачи информации на экране и учет возможных изменений экранных форм оператором;
- Эмоционально-волевая компонента: учет психических характеристик, темперамент, учет эмоционального фактора за счет фиксирования физиологических характеристик и движения мыши ПК, а также анализ взаимосвязей между оценкой ситуации и действиями оператора в истории работы.

3. Учет темперамента как одной из основных характеристик

Психофизиологические характеристики охватывают очень широкий спектр, это и темперамент, и как оператор принимает решения, свойственны ли ему те или иные стратегии поведения в экстренных ситуациях, зависят ли они от его эмоционального состояния и состояния здоровья в моменте. Отметим, что любая попытка новой классификации не отменяет предыдущие, а становится лишь дополнением – ещё одним измерением многогранной личности человека. При этом часто исследования подобного рода начинаются с попытки опереться на темперамент человека, как на наиболее исследованный набор характеристик. В работе [Грановская и др., 2018] приведено подробное описание, как действуют в различных стрессовых ситуациях люди с разным темпераментом, как влияет риск на них. Например, холерики и сангвиники работают лучше в ситуации форс-мажора, а флегматики и меланхолики действуют продуктивнее в условиях, когда направление поиска решения уже заранее определено или имеет место чёткая, формализованная и стабильная структура производственного процесса [Грановская и др., 2018]. Также показано, что у выраженного холерика время, используемое для анализа событий, короткое, что подкрепляет его тактические возможности и снижает стратегические. Поэтому в случае, когда только анализ длительных временных интервалов ведёт к оптимальному решению, холерики малопродуктивны. А у флегматиков оптимальное время анализа существенно больше, что способствует стратегическим решениям. Вместе с тем их медлительность, невозмутимое спокойствие и осмотрительность не позволяют им оперативно оценивать как обстановку, так и собственные действия. В случае необходимости принятия решения у них сразу же формируется внутренний протест, хотя в некоторых случаях уже пора принимать соответствующее решение. Соответственно, можно предположить ситуации, в которых необходимо корректирующее воздействие в представлении информации в зависимости от темперамента.

4. Пример изменения окрашенности текста на основе темперамента

В качестве простого (но важного для дальнейшей работы) примера приведем изменение формы подачи текста темы “Когнитивный подход к разработке онтологий” из учебника [Гаврилова и др., 2016, с. 213-214] с помощью системы Perplexity AI.

Исходный текст:

Если в результате анализа области и предназначения онтологии выявлено, что онтология будет использоваться для коммуникаций людей, то следует учесть некоторые когнитивные аспекты онтологий. В частности, на этапе формирования иерархии классов представляется целесообразным применить некоторые результаты гештальт-психологии [Гаврилова, 2005; Гаврилова, Воинов, 2007]. Автор гештальт-психологии Макс Вертгеймер так сформулировал основной принцип хорошего гештальта (хорошей формы) или закон прегнантности [Вертгеймер, 1987]: «Организация любой структуры в природе или в сознании должна быть настолько хороша (регулярна, полна, сбалансирована или симметрична), насколько позволяют существующие условия». Также полезными могут быть и другие когнитивно-перцептивные законы:

- закон близости – визуальные стимулы (объекты), находящиеся близко друг от друга воспринимаются как единое целое;
- закон сходства – вещи, обладающие одинаковыми свойствами, обычно воспринимаются как нечто единое (цельное);
- закон включения В. Келера – есть тенденция воспринимать только большую фигуру, а не ту меньшую, которую она включает;
- закон парсимонии – самый простой пример является самым лучшим, известен как принцип «бритвы Оккама»: «не нужно умножать сущности без необходимости».

Переформулируем эти законы для практического использования инженером по знаниям. Основная гипотеза: «Гармония = концептуальный баланс + ясность» [Гаврилова, 2005].

При этом концептуальный баланс подразумевает, что:

- понятия одного уровня иерархии на одной диаграмме связываются с родительским концептом одним и тем же типом отношения (например, «класс – подкласс» или «часть – целое»);
- глубина ветвей онтологического дерева должна быть примерно одинаковая (± 2);
- общая картинка должна быть довольно симметричной;
- перекрестные ссылки должны быть по возможности исключены.

Ясность включает:

- минимизацию. Так максимальное число концептов одного уровня или глубина ветви не должна превышать знаменитое число Ингве-Миллера (7 ± 2);
- прозрачность для чтения. Тип отношений на диаграмме должен быть по возможности очевиден, так чтобы не перегружать схему онтологии лишней информацией и опускать названия отношений.

Трудно не согласиться с более «правильной» формой онтологии А по сравнению с онтологией В. Резюмируя, можно сказать, что в случае использования онтологии для коммуникации с людьми следует стремиться к концептуальному балансу и ясности создаваемых концептуальных моделей. В случае, если требования к онтологии и предметная область не позволяют этого обеспечить (например, поставлена задача «по-крупному» проработать всю предметную область и подробнее разобрать один раздел или разные разделы предметной области находятся на разных уровнях зрелости/освоения), то следует при демонстрации онтологии человеку разделить ее на сбалансированные и ясные фрагменты (например, общий вид без деталей + отдельная проработанная ветка).

Преобразованный текст для холерика:

Если вы хотите, чтобы онтология работала для общения между людьми, действуйте решительно: учитывайте, как люди воспринимают информацию! На этапе создания иерархии классов используйте принципы гештальт-психологии – это ускорит понимание и повысит эффективность. Главное правило: структура должна быть максимально четкой, сбалансированной и симметричной, насколько это возможно. Используйте:

- *Закон близости: группируйте связанные элементы рядом.*
 - *Закон сходства: похожие объекты объединяйте.*
 - *Закон включения: выделяйте главное, не распыляйтесь на детали.*
 - *Закон парсимонии: не усложняйте – чем проще, тем лучше.*
- Ваша цель – гармония: баланс концепций и ясность!*
- *Все элементы одного уровня связывайте одинаково.*
 - *Ветви должны быть примерно одинаковой глубины.*
 - *Симметрия – признак порядка.*
 - *Минимизируйте перекрестные связи.*

Ясность – это минимум лишних деталей и прозрачность связей. Не перегружайте схему, делайте всё очевидным. Если невозможно соблюсти баланс – разбейте онтологию на понятные части: общий вид плюс отдельные проработанные ветки.

Для меланхолика:

Если анализ показал, что онтология нужна для коммуникации между людьми, важно учесть особенности восприятия информации. На этапе построения иерархии классов полезно опереться на гештальт-психологию, которая учит: структура должна быть максимально гармоничной, полной и симметричной, насколько это возможно.

Обратите внимание на следующие законы:

- *Близость*: объекты, находящиеся рядом, воспринимаются как единое целое.

- *Сходство*: похожие элементы объединяются в сознании.

- *Включение*: мы склонны видеть большую фигуру, не замечая меньшую внутри неё.

- *Парсимония*: выбирайте самые простые решения.

Практически это означает, что:

- Все элементы одного уровня связываются с родительским одним типом связи.

- Глубина ветвей должна быть примерно одинаковой.

- Старайтесь избегать перекрёстных связей.

- Число элементов на уровне не должно превышать 7 ± 2 .

Если не получается сделать схему полностью гармоничной, разделите её на более простые и понятные части для демонстрации.

Для флегматика:

Если выяснилось, что онтология будет использоваться для коммуникации, важно спокойно и последовательно учесть когнитивные аспекты. На этапе формирования иерархии классов применяйте принципы гештальт-психологии: структура должна быть такой хорошей, насколько позволяют условия – регулярной, полной, сбалансированной, симметричной.

Следуйте основным законам:

- *Закон близости*: группируйте близкие объекты.

- *Закон сходства*: объединяйте схожие элементы.

- *Закон включения*: выделяйте главное.

- *Закон парсимонии*: упрощайте, где это возможно.

В работе придерживайтесь баланса и ясности:

- Используйте одинаковые типы связей для элементов одного уровня.

- Держите глубину ветвей примерно равной.

- Стремитесь к симметрии и минимизируйте перекрёстные связи.

- Не перегружайте схему – оптимально 7 ± 2 элемента на уровне.

Если невозможно соблюсти все условия, разбейте онтологию на понятные фрагменты: общий вид и отдельные подробные ветки.

Для сангвиника:

Если ваша онтология нужна для общения между людьми – отлично! Важно сделать её понятной и привлекательной. На этапе построения иерархии классов используйте гештальт-принципы: пусть всё будет

гармонично, симметрично и легко для восприятия.

Применяйте простые правила:

- *Близость: размещайте связанные объекты рядом.*
- *Сходство: похожие вещи объединяйте.*
- *Включение: выделяйте главное, не теряйтесь в деталях.*
- *Парсимония: не усложняйте – чем проще, тем лучше!*

Главное – баланс и ясность:

- *Связывайте элементы одного уровня одинаково.*
- *Старайтесь, чтобы ветви были одинаковой глубины.*
- *Делайте схему симметричной и без лишних связей.*
- *Не перегружайте: оптимально 7 ± 2 элемента на уровне.*

Если не удаётся сделать всё идеально – покажите сначала общую картину, а потом отдельные детали. Так будет проще и интереснее!

Выделим ключевые аспекты из предложенных измененных текстов по темпераментам:

- выделена суть, сам текст сокращен;
- учтены характеристики темперамента: эмоциональность (наличие восклицаний у холерика и сангвиника), степень детализации, акцент на мотивацию или на безопасность и порядок;
- разный способ подачи информации (призыв к действию, объяснение, пошаговая инструкция, вовлекающий стиль).

Такой подход позволяет сделать информацию максимально понятной и комфортной для восприятия представителям каждого типа темперамента.

Заключение

Нам всем хочется жить в мире, где люди образованы, профессиональны и эмоционально устойчивы. Однако в реальности нет очередей на работу операторами, а анализировать и управлять технологическими процессами необходимо. При этом операторам важно оставаться в эмоционально стабильном состоянии, минимизировать стресс, и быть уверенным в принимаемых решениях. Построение модели мира субъекта, обучение систем на работу с операторами, адаптацию на основе их психофизиологических характеристик должно помочь в реализации целей управления процессами в меняющихся условиях.

Список литературы

- [Azarov et al., 2020] Azarov A., Tulupyeva T., Suvorova A., Koroleva M., Vasileva O. Assessment of the Information System's Protection Degree from Social Engineering Attack Action of Malefactor While Changing the Characteristics of User's Profiles: Numerical Experiments // In: Proceedings of the Fourth International Scientific Conference "Intelligent Information Technologies for Industry" (ITI'19), Ostrava, Czech Republic, 2019. Advances in Intelligent Systems and Computing. – 2020. – Vol. 1156. – P. 523-530. – doi: 10.1007/978-3-030-50097-9_53.

- [**Стинбарджер, 2013**] Стинбарджер Б. Психология трейдинга. Инструменты и методы принятия решений. – М.: Альпина Паблишер, 2013. – 368 с.
- [**Киселев, 2020**] Киселев С.С., Сазонов С.Ю., Сазонова Е.С. Психологическая модель поведения агента // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. – 2020. – № 9. – С. 69-74. – doi: 10.37882/2223-2966.2020.09.20.
- [**Муштонина, 2021**] Муштонина Е.А., Александрова А.В., Левчук А.А. Учет личностных характеристик и компетенций в вопросах безопасности труда при профессиональном отборе работников (на примере персонала локомотивных бригад) // III Международная научно-практическая конференция, посвященная 75-летию Победы Советского народа в Великой Отечественной Войне 1941-1945 гг. "Научный потенциал вуза – производству и образованию"(Армавир, 04-05 декабря 2020 года): Сборник статей. – Армавир: ООО «Редакция газеты «Армавирский собеседник» (Армавирская типография), 2021. – С. 227-232.
- [**Гаврилова и др., 2016**] Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. Инженерия знаний. Модели и методы: учебник. СПб.: Изд-во «Лань», 2016. – 324 с.
- [**Грановская и др., 2018**] Грановская Р.М. [и др.]. Взаимосвязь темперамента, психологических защит и совладания со стрессом // Вестник психотерапии. – 2018. – № 68(73). – С. 83-99.
- [**Атанов, 2001**] Атанов Г.А. Моделирование учебной предметной области, или предметная модель обучаемого // Образовательные технологии и общество. – 2001. – Вып. 1. – Т. 4. – С. 111-124.
- [**Голубев, 2018**] Голубев П. Интеллектуальные системы поддержки принятия решений – краткий обзор. Блог компании Open Data Scienceб 2018 // Хабр: сайт. – URL: <https://habr.com/ru/companies/ods/articles/359188/>.
- [**Давыдова и др., 2015**] Давыдова Е.Н., Сергушичева А.П. Модели обучаемого и преподавателя для мультиагентной обучающей системы // Открытое образование. – 2015. – № 5. – С. 25-31.
- [**Добрица и др., 2019**] Добрица В.П., Горюшкин Е.И. Применение интеллектуальной адаптивной платформы в образовании // Auditorium. Электронный научный журнал Курского государственного университета. – 2019. – № 1(21).
- [**Дорофеев и др., 2007**] Дорофеев А.С., Сосинская С.С. Модели обучающего курса в разработке систем дистанционного обучения // Прикладная информатика. – 2007. – № 3(9). – С. 25-37.
- [**Кречетов и др., 2018**] Кречетов И.А., Дорофеева М.Ю., Дегтярев А.В. Раскрываем потенциал адаптивного обучения: от разработки до внедрения // Материалы Международной конференции «Learning Stakeholders and Researchers Summit 2018». – М: Изд. дом ВШЭ, 2018. – С. 76-85.
- [**Кузнецов и др., 2009**] Кузнецов М. А., Пономарев С. С. Современная классификация систем поддержки принятия решений // Прикаспийский журнал управления и высокие технологии. – 2009. – № 3(7). – С. 52-58.
- [**Рыбина, 2017**] Рыбина Г.В. Интеллектуальная технология построения обучающих интегрированных экспертных систем: новые возможности // Открытое образование. – 2017. – № 4. – С. 43-57.

- [Шаплыко, 2020] Шаплыко И.М. Адаптивная модель обучаемого для системы управления обучением в ВУЗе: автореф. дисс. магистра технических наук. – Минск БГУИР, 2020. – 53 с.
- [Cardin et al., 2022] Cardin O., Trentesaux D. Design and Use of Human Operator Digital Twins in Industrial Cyber-Physical Systems: Ethical Implications // In: 14th IFAC Workshop on Intelligent Manufacturing Systems IMS 2022 (IFAC-PapersOnLine), Tel-Aviv, Israel, 28-30 March 2022. – 2022. – Vol. 55, Issue 2. – P. 360-365. – doi: 10.1016/j.ifacol.2022.04.220.
- [Дозорцев и др., 2022] Дозорцев В.М., Венгер А.Л. О проблеме доверия человека-оператора искусственному интеллекту // Автоматизация в промышленности. – 2022. – № 2. – С. 10-17.

УДК 004.056

doi: 10.15622/rcai.2025.097

ОБНАРУЖЕНИЕ АНОМАЛИЙ В КОНТЕЙНЕРНЫХ СИСТЕМАХ НА ОСНОВЕ ПРОФИЛИРОВАНИЯ И НЕЙРОННОЙ СЕТИ AE-LSTM¹

И.В. Котенко (*ivkote@comsec.spb.ru*)

М.В. Мельник (*mkmxvh@gmail.com*)

Санкт-Петербургский Федеральный исследовательский центр РАН,
Санкт-Петербург

Особенности архитектуры контейнерных систем создают дополнительные риски, позволяя злоумышленникам негативно воздействовать не только на целевой контейнер, но и на другие контейнеры, размещенные на том же хосте, а также на саму операционную систему. Для повышения точности обнаружения аномалий в контейнерных системах предложена методика создания программных систем обнаружения аномального поведения за счёт оценки состояния элементов контейнерных систем. Данная методика основана на создании эталонных профилей легитимных значений параметров элементов контейнерных систем и использовании нейронной сети AE-LSTM. В качестве таких параметров могут выступать точки монтирования, сетевые интерфейсы, ограничения на использование вычислительных ресурсов и другие. Обученная на эталонном профиле модель позволяет выполнить классификацию текущего состояния контейнера как легитимного или аномального на основе вычисления ошибки реконструкции данных. Если ошибка реконструкции превышает заданный предел – регистрируется аномалия. Точность представленного составляет от 96,9% до 98,7% в зависимости от модели и элемента контейнерной системы.

Ключевые слова: обнаружение аномалий, контейнерные системы, элементы контейнерной системы, нейронные сети.

¹ Исследование выполнено за счет гранта Санкт-Петербургского научного фонда № 23-РБ-01-09.

Введение

Контейнеры, как и виртуальные операционные системы (ОС), представляют собой динамические среды выполнения программного кода. В последнее время они становятся основополагающим элементом при развертывании различных систем, особенно в условиях, когда оптимизация вычислительных ресурсов играет решающее значение. Эффективность контейнерных систем достигается за счет интеграции только необходимых программных компонентов, что является ключевым фактором при создании и развертывании сложных высоконагруженных систем.

Помимо классических проблем безопасности (уязвимости, неправильные конфигурации и другие), контейнеры сталкиваются с рядом специфических проблем. Такие проблемы возникают из-за их архитектурных особенностей, обуславливаемых реализацией основного принципа контейнеризации – WORA (Written Once And Run Anywhere, написано один раз и работает везде) [Dakić, 2024], который позволяет контейнеру использовать ядро хостовой ОС.

Атаки могут быть обнаружены благодаря подходу на основе профилирования, поскольку они часто сопровождаются необычным поведением, которое не свойственно при штатном функционировании наблюдаемой системы. Однако, используемые профили могут содержать большие объемы информации, особенно если в качестве данных используются системные вызовы (System Calls, Syscalls). Кроме того, для реализации подхода на основе профилирования с помощью сложных методов глубокого машинного обучения, например, таких, как управляемые рекуррентные блоки (Gated Recurrent Unit, GRU) или долгая краткосрочная память (Long Short-Term Memory, LSTM), могут потребоваться значительные объемы вычислительных ресурсов и времени. В то же время может оказаться недостаточно использования простых методов, таких, как автокодировщики (Autoencoder, AE) или перцептроны (Perceptron, P).

Для повышения точности обнаружения аномального поведения и снижения затрат вычислительных ресурсов, в статье представлена методика реализации программных систем обнаружения аномального поведения контейнерных систем. Предлагаемая методика предполагает создание эталонных профилей легитимных значений параметров элементов контейнерных систем и основана на использовании модели нейронной сети AE-LSTM. Архитектура разработанной системы и реализованного прототипа содержит два компонента. Первый компонент – модуль обучения на основе профилей легитимных значений параметров элементов контейнеров, второй – модуль обнаружения.

На этапе обучения создаются эталонные профили, описывающие легитимное поведение элементов контейнерных систем для каждой группы контейнеров. Эти профили состоят из данных о значениях определенных параметров элементов контейнерной системы, собранных за определен-

ные периоды времени. Такие параметры включают сетевые интерфейсы, ограничения на использование вычислительных ресурсов, точки монтирования и другие. После обучения на эталонных данных нейронная сеть способна классифицировать текущее состояние контейнера, определяя его как легитимное или аномальное. В процессе обнаружения аномалий модель анализирует значения параметров на определенном промежутке времени и вычисляет ошибку реконструкции данных. Если ошибка реконструкции этих данных превышает заданное пороговое значение, система сигнализирует о наличии аномалии.

Оценки, полученные в ходе экспериментов, демонстрируют уровень точности показателя Assurance в диапазоне от 96,9% до 98,7% в зависимости от модели нейронной сети AE-LSTM и класса параметров элементов контейнерных систем. Кроме того, представленное решение не требует высоких затрат вычислительных ресурсов (как при обучении модели, так и при обнаружении).

1. Обзор релевантных работ

Многие исследования демонстрируют эффективное применение методов машинного обучения в качестве ключевого метода реализации подхода на основе профилирования в задачах обнаружения аномального поведения [Branitskiy et al., 2017], [Laskov et al., 2004]. В текущих исследованиях используются, в основном, модели глубокого обучения.

В [Kotenko et al., 2024] используется трассировка syscalls для отслеживания поведения в контейнерных системах. После чего выполняется построение гистограмм процессов и передача их в модель нейронной сети AE для обучения и последующего обнаружения аномальных процессов. Авторы отмечают неспособность предложенного подхода к обнаружению определенных типов атак.

В [Tien et al., 2019] представлена система KubAnomaly для обнаружения аномалий на платформе оркестрации Kubernetes. Представленный подход, напротив, основан на разбиении непрерывного потока syscalls на отдельные последовательности. Ввиду большого количества данных авторы приняли решение отслеживать только четыре категории syscalls (файловый ввод-вывод, сетевой ввод-вывод, планировщик и память). Результаты экспериментов демонстрируют точность предложенной модели около 96%.

Подходы, использованные в [Snehi et al., 2021] и [Gantikow et al., 2020], похожи и основаны на использовании модели нейронной сети LSTM и трассировки syscall. В [Snehi et al., 2021] анализируется поведение контейнерной системы на основе syscall. Процесс обнаружения основан на анализе разности предсказанной последовательности syscalls по отношению к действительной. Если значение разности превышает допустимое

пороговое значение, то поведение контейнера считается аномальным. В [Gantikow et al., 2020] представлены два подхода к обнаружению. Первый, как и в [Snehi et al., 2021], основан на трассировке и прогнозировании syscalls, второй – на прогнозировании путей к файлам/каталогам. В целом, можно отметить, что оба представленных подхода хорошо дополняют друг друга.

В [Gupta et al., 2017] используется гибридная нейронная сеть для прогнозирования потребления вычислительных ресурсов во время работы контейнера. Представленная модель состоит из классической модели LSTM и двунаправленной сети долгой краткосрочной памяти (Bidirectional LSTM, BiLSTM). Ряд экспериментов показал, что эта модель демонстрирует хорошую точность, а показатель Recall достигает 94%.

Работы [Castanhel et al., 2021] и [Cui et al., 2021] – схожи и используют общий подход на основе трассировки и техники скользящего окна для сбора и формирования последовательностей syscalls. В [Castanhel et al., 2021] сравнивается эффективность различных методов машинного обучения. В качестве входных данных используются последовательности syscalls, полученные в результате трассировки. Последовательности формируются на основе техники скользящего окна и фильтрации. Авторами отмечается, что точность метода Multilayer Perceptron (MLP) растет вместе с ростом размера скользящего окна и составляет более 89% в [Cui et al., 2021] напротив, используется гибридная нейронная сеть на основе AE-LSTM. В целом, авторы отмечают неплохую точность обнаружения аномалий неконтролируемой моделью нейронной сети.

В [Wang et al., 2023] представлена система DockerWatch на основе двух нейронных сетей CNN и CNN-LSTM, которые используют извлеченный из контейнера исполняемый файл для анализа на предмет вредоносного содержимого.

В результате проведенного анализа релевантных работ можно сделать вывод, что большинство исследований по теме обнаружения аномального поведения направлены на использование syscalls в качестве входных данных для моделей нейронных сетей. Такой подход имеет существенный недостаток, связанный с большими объемами данных, генерируемыми контейнерными системами. Обучение моделей со сложной архитектурой может быть крайне затруднено с точки зрения использования вычислительных ресурсов и времени, а обучение простых моделей может оказаться недостаточным для эффективного и своевременного обнаружения.

2. Подход к обнаружению

Работа контейнерной системы обеспечивается за счет взаимодействия ее ключевых элементов, приложения внутри такой системы и пользователей. В качестве элементов контейнерной системы могут выступать: образ

контейнера (Docker Image), контейнерный движок (Container Engine), контейнерная сеть (Container Networking), хранилище (Volumes), API (Application Programming Interface) и CLI (Command-Line Interface) и механизмы безопасности, определяемые пространством имен (Namespaces), контрольными группами (Control Groups, Cgroups), Seccomp (Secure Computing Mode), AppArmor (Application Armor), SELinux (Security-Enhanced Linux) и возможностями Linux (Capabilities).

Аномалии могут проявляться в изменениях различных значений параметров таких элементов. Например, появление новой точки монтирования может указывать на попытку атаки типа Docker Escape. Запуск контейнера с новым, ранее не использовавшимся образом, может свидетельствовать о внедрении вредоносного программного обеспечения. Изменение IP-адреса (Internet Protocol) контейнера может быть признаком подготовки к атаке “Человек посередине” (Man-in-the-Middle, MitM), IP Spoofing или попытке обойти сетевые политики и правила безопасности. Изменение параметров CapAdd и CapDrop, которые управляют добавлением и удалением Capabilities стандартного набора контейнера, может предполагать попытку повышения привилегий (Privilege Escalation) или обход механизмов безопасности.

Учитывая особенности контейнеров и природу аномалий, предлагаемый подход основывается на создании эталонных профилей, отражающих нормальное поведение элементов этих систем. Каждый профиль формируется для определенной группы контейнеров и некоторого класса параметров на последовательных временных интервалах со смещением в одну минуту. Продолжительность каждого интервала составляет 10 минут.

Таким образом, для каждой категории параметров создается отдельный профиль: для ограничений использования вычислительных ресурсов (RUL), для сетевых интерфейсов (Net), для точек монтирования (Mnt), для образов (Img). Сформированные профили используются для обучения моделей нейронной сети AE-LSTM, которая в дальнейшем применяется для обнаружения аномалий в работе контейнера путем анализа параметров его элементов.

Представленная система включает несколько ключевых компонентов: (1) сбор данных; (2) нормализация данных; (3) обучение и (4) обнаружение.

(1) Сбор данных. Для сбора значений параметров каждого элемента контейнерной системы разработан программный компонент, который в единицу времени (минуту) собирает информацию о сетевых интерфейсах, установленных ограничениях на использование вычислительных ресурсов, точках монтирования и др., как представлено в листинге 1.

Данные о параметрах элементов контейнерной системы

```

1: Net: 127.0.0.1, 172.28.1.10,...
2: RUL: {"CpuShares":0, "Memory":432013312, "KernelMemory":null,...
3: Mnt: /dev/sda2 on /etc/resolv.conf type ext4 (rw,relatime)...
```

В данном листинге в разделе в разделе “Net” представлены сетевые интерфейсы, в разделе “RUL” – установленные лимиты на использование вычислительных ресурсов, “Mnt” – представлены точки монтирования.

(2) Нормализация данных. На первом этапе формируются временные последовательности собранных значений параметров элементов контейнеров. Каждая последовательность охватывает 10-минутный интервал со смещением в 1 минуту. На втором этапе выполняется обработка данных для каждого класса параметров, которая осуществляется с учетом специфики их представления. Например, для профиля Net каждый октет сетевого интерфейса записывается как отдельный элемент в формируемую строку; в результате строка содержит описание всех сетевых интерфейсов. Для профиля RUL на основе предопределённых параметров, таких, как CpuShares, Memory, KernelMemory и др., формируется строка, в которой первое значение соответствует CpuShares, второе – Memory, третье – KernelMemory и так далее; если значение для какого-либо параметра отсутствует, используется значение по умолчанию, равное 0.

(3) Обучение. В основе предлагаемой системы лежит неконтролируемая гибридная модель нейронной сети AE-LSTM, обучение которой осуществляется на основе данных, характеризующих проведение элементов контейнерных систем. Такой подход позволяет избежать необходимости явной разметки данных, которая в реальных условиях затруднена. В результате успешного обучения каждому профилю будет соответствовать модель AE-LSTM, которая представлена на рис. 1.

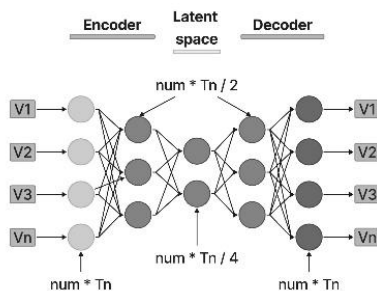


Рис. 1. Модель нейронной сети AE-LSTM

После загрузки данных в модель нейронной сети, кодировщик (Encoder) сжимает входные данные в скрытое представление (Latent space). После чего декодер (Decoder) их восстанавливает из Latent space в исходное состояние. Таким образом, модель обучается минимизировать ошибку реконструкции во время обучения.

Ключевой особенностью данной архитектуры является постепенное уменьшение размерности данных на первом ($\text{num} * T_n / 2$) и втором ($\text{num} * T_n / 4$) уровне скрытого слоя. Размер входного и выходного слоя определяется как количество значений в строке передаваемых данных (num) умноженное на длину временного периода (T_n), которая равна 10). Такая архитектура позволяет предотвращать переобучение модели, сосредотачиваясь на наиболее значимых признаках. В качестве функций активации используются ReLU (Rectified Linear Unit), которые обеспечивают быстрое вычисление за счёт использования простой функции: если входное значение положительное, то функция возвращает это значение, а если отрицательное или равно 0, то функция возвращает 0.

Кроме того, в ходе ряда экспериментов было установлено, что использование модели AE-LSTM, в отличие от классической модели AE, обеспечивает значительно более высокие показатели точности.

При обучении модели нейронной сети использовалась комбинация алгоритма оптимизации ADAM (Adaptive Moment Estimation) и функции потерь MSE (Mean Squared Error). ADAM быстро и эффективно настраивает параметры модели, такие как веса (Weights) и смещения (Biases). Это позволяет автоматически подстраивать скорость обучения для каждого из этих параметров. Функция MSE, измеряя средний квадрат разности между предсказанными и фактическими значениями, обеспечивает плавные градиенты, что упрощает и стабилизирует процесс обучения, приводя к более точным прогнозам.

(4) Обнаружение. На этапе обнаружения данные за 10-минутный интервал передаются в соответствующую модель для классификации состояния элемента контейнера как легитимного или аномального. Для поступивших данных вычисляется MSE между исходными и реконструированными данными. Если это значение превышает установленное пороговое значение, данные классифицируются как аномальные.

Пороговое значение устанавливается как среднее значение MSE плюс коэффициент, умноженный на стандартное отклонение. Среднее и стандартное отклонение MSE рассчитывается на основе вычисления MSE реконструкции каждого примера из набора данных, представляющих нормальное поведение. Это обеспечивает гибкую настройку чувствительности, балансируя между точностью и ложными срабатываниями.

3. Эксперименты

Для подтверждения эффективности разработанной методики и программного прототипа был создан испытательный стенд, подготовлены наборы данных, обучены модели нейронной сети и выполнены оценки точности обученных моделей.

Испытательный стенд. Испытательный стенд включает две виртуальные машины под управлением Ubuntu 22.04.5 LTS. Первая виртуальная машина содержит программный компонент для сбора и передачи значений параметров элементов контейнеров в модель нейронной сети, а также два Docker-контейнера: RDP-сервер (Remote Desktop Protocol) и SSH-сервер (Secure Shell). Вторая виртуальная машина также работает под управлением Ubuntu 22.04.5 LTS и используется для взаимодействия с Docker-контейнерами посредством RDP и SSH. Это взаимодействие включает в себя запуск утилит, выполнение команд ОС, обновление параметров элементов контейнерных систем и другие.

Наборы данных. Всего сформировано четыре набора данных, которые получены в результате работы двух контейнеров.

Два набора данных (rdp-n.txt и ssh-n.txt) включают данные легитимных параметров элементов контейнерных систем, на которых были обучены модели нейронной сети AE-LSTM.

Набор данных rdp-n.txt содержит: редактирование сетевых интерфейсов с использованием легитимных значений, монтирование диска внутри контейнера (легитимные пути целевых устройств – /dev/sda2/ и /dev/sda3/ и другие, легитимные пути назначения – /mnt/share_1, /mnt/share_2, /mnt/share_3 и другие).

Набор данных ssh-n.txt содержит: изменение ограничений на использование вычислительных ресурсов (cpuset-cpus, memory, memory-swap, cpu-shares, cpu-quota, cpu-period, memory-reservation) на легитимные значения; запуск контейнеров на основе проверенных образов.

Два остальных набора данных (rdp-a.txt и ssh-a.txt) включают как данные легитимных значений параметров, так и не легитимных.

Набор данных rdp-a.txt содержит: редактирование сетевых интерфейсов с использованием легитимных и не легитимных значений, монтирование диска внутри контейнера (легитимные пути целевых устройств – /dev/sda2/ и /dev/sda3/ и другие, легитимные пути назначения – /mnt/share_1, /mnt/share_2, /mnt/share_3 и другие, нелегитимные пути целевых устройств – /dev/sdb2/ и /dev/sdb3/ и другие, нелегитимные пути назначения – /mnt/escape_1, /mnt/escape_2, /mnt/escape_3 и другие).

Набор данных ssh-a.txt содержит: изменение ограничений на использование вычислительных ресурсов (cpuset-cpus, memory, memory-swap, cpu-shares, cpu-quota, cpu-period, memory-reservation на легитимные и не легитимные значения), запуск контейнеров на основе проверенных и не проверенных образов.

Обучение. Обучение моделей проводилось на виртуальной машине, которая включает следующие характеристики: ОС: Ubuntu 22.04.5 LTS; RAM (Random Access Memory): 8 GB (Gigabyte); CPU (Central Processing Unit): AMD Ryzen 5 5600X 2 Virtual Core Processor 3.70 GHz. Время для обучения каждой модели составило не более 1 минуты.

В ходе ряда экспериментов было установлено, что при использовании представленной архитектуры наилучшие показатели обнаружения модели достигаются при уровне потерь в диапазоне от 0,5% до 1%. Такой уровень потерь достигается при установленных параметрах batch size (количество примеров, используемых в одной итерации обучения), равном 1, и epoch (количество итераций обучения), равном 5.

В результате были обучены четыре модели нейронной сети AE-LSTM. Данные для каждой модели были агрегированы из двух наборов данных (rdp-n.txt и ssh-n.txt), а затем с помощью фильтрации предобработаны для обучения каждой модели в соответствии с типом параметров элементов контейнеров. Таким образом, модель Net содержит данные о легитимных сетевых интерфейсах, RUL – данные о легитимных значениях ограничения вычислительных ресурсов, MNT – данные о легитимных точках мониторинга, Img – данные о легитимных образах контейнеров. Кроме того, наборы данных были оптимизированы путем удаления дублирующихся последовательностей, в результате чего из нескольких идентичных последовательностей оставалась только одна.

Результаты обнаружения. Оценка эффективности разработанного прототипа проводилась с помощью меры Accuracy на наборах данных: rdp-n.txt, ssh-n.txt, rdp-a.txt и ssh-a.txt). Результаты точности разработанных моделей для модели net составляют 98,9 %, для rul – 96,9 %, mnt – 98,7 %, img – 97,2 %.

4. Анализ результатов

В данной статье представлена методика разработки программных систем обнаружения аномального поведения в контейнерах, которая основана на оценке состояния их элементов. Ключевой особенностью и научной новизной предлагаемого решения является использование значений параметров элементов контейнерной системы на временных промежутках в качестве входных данных для моделей нейронной сети AE-LSTM.

В отличие от предыдущих исследований [Kotenko et al., 2024], [Tien et al., 2019], [Snehi et al., 2021], [Gantikow et al., 2020], которые используют syscalls в качестве входных данных для моделей нейронных сетей, предлагаемое решение использует данные о параметрах элементов контейнерных систем. Вследствие чего нагрузка на вычислительные ресурсы значительно ниже, чем при использовании syscalls.

Следует отметить, что работы [Castanhel et al., 2021], [Cui et al., 2021], в которых используется техника скользящего окна для формирования последовательностей syscalls, ввиду большого потока данных, генерируемых динамичными контейнерами, могут столкнуться со сложностями при обработке данных. В отличие от этого подхода, метод анализа параметров элементов контейнерных систем не сталкивается с данной проблемой, поскольку объем таких данных невелик, а их анализ осуществляется каждую минуту. При необходимости интервал между итерациями анализа можно сократить, а длину временной последовательности увеличить.

В [Gupta et al., 2017] представлен подход на основе прогнозирования показателей нагрузки на вычислительные ресурсы во время работы контейнера. Применённый подход не сможет обеспечить обнаружение аномалий, природа которых не связана с изменением нагрузки на вычислительные ресурсы, например, запуск контейнера с установленными ограничениями на использование вычислительных ресурсов, изменение сетевых интерфейсов, монтирование диска хостовой операционной системы из контейнера. В отличие от этого подхода, предложенный в данном исследовании подход, напротив, сможет обеспечить обнаружение таких аномалий.

Однако, у предложенного решения есть ряд недостатков. Ввиду того, что учитываются только данные о параметрах элементов контейнера, некоторые атаки не могут быть обнаружены. К таким атакам можно отнести Crypto Mining, Privilege Escalation.

Заключение

В статье представлена методика реализации системы обнаружения аномального поведения в контейнерных системах за счёт оценки состояния элементов контейнерных систем. Предлагаемая методика основана на создании эталонных профилей, отражающих состояние элементов контейнера на временных интервалах.

Новизной и отличительной особенностью предлагаемого подхода к обнаружению аномального поведения в контейнерных системах является анализ временных последовательностей значений параметров элементов контейнерных систем с использованием моделей AE-LSTM.

Результаты проведенного эксперимента показывают, что точность реализованного прототипа составляет от 96,9% до 98,7% в зависимости от модели нейронной сети AE-LSTM и класса параметров элементов контейнерной системы. Следует отметить, что предлагаемый подход нацелен на обнаружение только тех аномалий, природа которых напрямую связана с изменением значений параметров элементов контейнера. Другие аномалии не будут выявлены. Таким образом, данное решение может быть использовано в качестве дополнительного инструмента для обнаружения

аномалий, учитывая, что оно позволяет быстро обучать модели нейронной сети и не требует значительных затрат вычислительных ресурсов для своей работы.

В рамках дальнейших исследований планируется расширить набор анализируемых параметров элементов контейнера с целью повышения точности обнаружения аномалий. В частности, будут исследованы: NetworkMode (режим работы сети), CgroupnsMode (режим пространств имен для групп контроля), IpcMode (режим межпроцессного взаимодействия), CapAdd (возможности, предоставленные контейнеру) и другие параметры, потенциально влияющие на безопасность и стабильность работы контейнеров.

Список литературы

- [Branitskiy et al., 2017] Branitskiy A., Kotenko I. Hybridization of computational intelligence methods for attack detection in computer networks // *Journal of Computational Science*. – Elsevier, 2017. – No. 23. – P. 145-156. – doi: 10.1016/j.jocs.2016.07.010.
- [Castanhel et al., 2021] Castanhel G.R., Heinrich T., Ceschin F., Maziero C. Taking a peek: An evaluation of anomaly detection using system calls for containers // 2021 IEEE Symposium on Computers and Communications (ISCC), Athens, Greece, 2021. – P. 1-6. – doi: 10.1109/ISCC53001.2021.9631251.
- [Cui et al., 2020] Cui P., Umphress D. Towards unsupervised introspection of containerized application // *Proceedings of the 2020 10th International Conference on Communication and Network Security*, New York, NY, United States, 2020. – P. 42-51. – doi: 10.1145/3442520.3442530.
- [Dakić, 2024] Dakić P. Software compliance in various industries using ci/cd, dynamic microservices, and containers // *Open Computer Science*. – 2024. – Vol. 14(1). – P. 20240013. – doi: 10.1515/comp-2024-0013.
- [Gantikow et al., 2020] Gantikow H., Zöhner T., Reich C. Container anomaly detection using neural networks analyzing system calls // 2020 28th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), Västerås, Sweden, 2020. – P. 408-412. – doi: 10.1109/PDP50117.2020.00069.
- [Gupta et al., 2017] Gupta S., Muthiyar N., Kumar S., Nigam A., Dinesh D.A. A supervised deep learning framework for proactive anomaly detection in cloud workloads // 2017 14th IEEE India Council International Conference (INDICON), Roorkee, India, 2017. – P. 1-6. – doi: 10.1109/INDICON.2017.8488109.
- [Kotenko et al., 2024] Kotenko I., Melnik M., Abramenko G. Anomaly detection in container systems: using normal process histograms and an autoencoder // *IEEE 25th International Conference of Young Professionals in Electron Devices and Materials (EDM 2024)*, Altai, Russian Federation, 2024. – P. 1930-1934. – doi: 10.1109/EDM61683.2024.10615118.
- [Laskov et al., 2004] Laskov P., Schafer C., Kotenko I. Intrusion detection in unlabeled data with one-class Support Vector Machines // *Lecture Notes in Informatics*, No. 46, Dortmund, Germany, July 2004. – P. 71-82.

- [**Snehi et al., 2021**] Snehi J., Bhandari A., Baggan V., Snehi M., Kaur H. AIDAAS: Incident handling and remediation anomaly-based IDaaS for cloud service providers // 2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART), MORADABAD, India, 2021. – P. 356-360. – doi: 10.1109/SMART52563.2021.9676296.
- [**Tien et al., 2019**] Tien C.W., Huang T.Y., Tien C.W., Huang T.C., Kuo S.Y. KubAnomaly: Anomaly detection for the Docker orchestration platform with neural network approaches // Engineering reports. – 2019. – Vol. 1(5). – P. e12080. – doi: 10.1002/eng2.12080.
- [**Wang et al., 2023**] Wang Y., Wang Q., Qin X., Chen X., Xin B., Yang R. DockerWatch: a two-phase hybrid detection of malware using various static features in container cloud // Soft Computing. – 2023. – Vol. 27(2). – P. 1015-1031. – doi: 10.1007/s00500-022-07546-2.

УДК 004.891

doi: 10.15622/rcai.2025.098

ПРОГРАММНАЯ ПЛАТФОРМА ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА В АУДИОЗАПИСЯХ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ¹

Е.С. Мытарин (*e.mytarin@ulstu.ru*)

В.С. Мошкин (*v.moshkin@ulstu.ru*)

А.Ю. Журавлев (*zhuravlev.a@ulstu.ru*)

Ульяновский государственный технический университет,
Ульяновск

В статье представлено описание гибридной методики программная платформа для автоматизированного обнаружения мошенничества в аудиозаписях телефонных разговоров, основанная на нейросетевых моделях и онтологическом анализе. Реализующая предложенный подход программная платформа интегрирует методы глубокого обучения, такие как распознавание речи (Vosk, Whisper), лемматизацию текста, анализ интонации и сопоставление с онтологической базой данных для оценки вероятности мошеннических действий. Решение ориентировано на применение в сфере финансовой безопасности и противодействия социальной инженерии. Результаты проведенных экспериментов в сравнении с современными подходами работы демонстрируют эффективность платформы в обработке аудиоданных, визуализации результатов и интеграции с базами данных.

Ключевые слова: противодействие мошенничеству, нейронные сети, онтологии, распознавание речи, социальная инженерия.

Введение

Современные мошеннические схемы, включая телефонное мошенничество, характеризуются высокой адаптивностью и использованием психологических приемов, что затрудняет их своевременное выявление. Традиционные методы анализа аудиоданных зачастую не справляются с обработкой

¹ Работа выполнена в рамках государственного задания № 075-03-2023-143 по проекту «Исследование интеллектуального предиктивного мультимодального анализа больших данных, конструирование признаков гетерогенных динамических данных для машинного обучения».

нечетких речевых паттернов и требуют значительных вычислительных ресурсов. В этой связи актуальной задачей становится разработка интеллектуальных систем, способных автоматизировать процесс детекции мошенничества на основе машинного обучения и формализованных знаний [1].

Одним из перспективных направлений является применение нейросетевых моделей, позволяющих анализировать речевые данные с учетом контекста и семантики. В сочетании с онтологическим подходом, обеспечивающим структурированное представление экспертных знаний, такой метод повышает точность классификации подозрительных диалогов [2].

В данной работе рассматривается программная платформа, интегрирующая методы глубокого обучения и онтологический анализ для предиктивной оценки вероятности мошенничества в аудиозаписях. Решение ориентировано на практическое применение в сфере финансовой безопасности [3].

1. Сравнительный анализ методов и программного обеспечения детектирования мошенничества в аудиозаписях

а. Программные аналоги

В настоящее время для решения задачи выявления мошеннических звонков применяются различные специализированные программные решения. Среди существующих разработок можно выделить несколько ключевых продуктов.

IBM Watson Speech to Text представляет собой платформу для преобразования речевых данных в текстовый формат с возможностью последующего семантического анализа. Система обеспечивает высокую точность распознавания речи, однако не содержит специализированных функций для детекции мошеннических схем [4].

Google Cloud Speech-to-Text – облачное решение для обработки аудиопотоков, поддерживающее интеграцию с NLP-алгоритмами. Несмотря на широкие возможности обработки естественного языка, платформа не предлагает готовых механизмов для выявления мошеннических сценариев [5].

Pindrop – специализированная система для обнаружения телефонного мошенничества, использующая анализ голосовых биометрических параметров и поведенческих паттернов. Решение демонстрирует высокую эффективность, но требует значительных вычислительных ресурсов и ориентировано преимущественно на корпоративный сегмент [6].

Nice Perform – программный комплекс для мониторинга звонков в колл-центрах, включающий функции анализа подозрительных диалогов. Система обладает развитыми возможностями машинного обучения, однако не поддерживает работу с пользовательскими онтологиями [7].

OpenVoice OS – открытая платформа для обработки голосовых данных с возможностью подключения сторонних аналитических модулей. Гибкость архитектуры компенсируется отсутствием предустановленных алгоритмов детекции мошенничества [8].

Каждый из рассмотренных инструментов обладает уникальным набором функций и может быть эффективен в определенных сценариях работы с аудиоданными. Однако ключевым ограничением существующих решений является отсутствие комплексного подхода, сочетающего прогностический анализ на основе нейросетевых моделей с гибкой системой настраиваемых онтологий для выявления мошеннических схем. Большинство представленных систем функционируют исключительно в режиме реактивного анализа, не обеспечивая предиктивной оценки потенциальных угроз.

b. Методы анализа

Для решения задачи выявления мошеннических звонков применяются различные программные решения, каждое из которых обладает уникальными характеристиками. Ниже представлен обзор ключевых методов и их параметров, основанный на анализе современных исследований. Основные критерии оценки эффективности применяемых методов: точность, скорость, интерпретируемость, устойчивость к шуму, потребление памяти [9].

Методы на основе NLP

- Точность – очень высокая (98.53%).
- Скорость – высокая (1-5 секунд на обработку записи).
- Интерпретируемость – средняя (правила анализа могут быть понятны, но сложность моделей затрудняет полное объяснение).
- Устойчивость к шуму – высокая (способность фильтровать шумы, но зависимость от качества распознавания речи).
- Потребление памяти – низкое (локальная обработка данных, 50-150 МБ) [10].

RAG-модели на основе LLM

- Точность – очень высокая (97.98%).
- Скорость – средняя (3-15 секунд, включая транскрипцию, поиск в базе знаний и генерацию ответа).
- Интерпретируемость – высокая (предоставляет обоснования решений).
- Устойчивость к шуму - высокая (использование cosine similarity для учета ошибок транскрипции).
- Потребление памяти – высокое (из-за использования больших языковых моделей, 4-8 ГБ) [11].

Гибридные системы (Pindrop)

- Точность – высокая (до 90%).
- Скорость – средне-высокая (зависит от сложности анализа, 1.5-15 секунд).

- Интерпретируемость – средняя (комбинация методов может усложнить объяснение).
- Устойчивость к шуму – высокая (использование акустических свойств голоса).
- Потребление памяти – средняя (из-за комплексного анализа, 2.5-4 ГБ) [12].

Методы на основе рекуррентных нейронных сетей (RNN)

- Точность – высокая (93.79%).
- Скорость – низкая (60–600 секунд).
- Интерпретируемость – низкая (затруднено понимание временных зависимостей).
- Устойчивость к шуму – высокая (эффективность для временных рядов).
- Потребление памяти – высокое (1–8 ГБ) [13].

Генеративно-сопоставительные сети (GAN)

- Точность – высокая (88-92%).
- Скорость – низкая (требуется время для обучения и анализа, 30-180 секунд).
- Интерпретируемость – низкая (сложность архитектуры GAN).
- Устойчивость к шуму – средняя (чувствительность к качеству данных).
- Потребление памяти – очень высокое (3–15 ГБ) [14]

В связи с тем, что все рассмотренные подходы не показывают максимальную точность при решении подобной задачи, актуальной задачей является разработка эффективного подхода для предиктивного выявления мошенничества в аудиозаписях, интегрирующей современные методы нейросетевого анализа и обработки естественного языка с учетом требований к скорости обработки и точности детекции, а также его реализация в виде программной платформы

2. Гибридная методика определения мошенничества посредством анализа аудиофайлов телефонных разговоров

Для решения поставленной задачи был разработан подход к детектированию мошенничества в аудиозаписях телефонных разговоров.

Блок-схема предложенной методики анализа содержания голосовых аудиофайлов включает представлена на рис. 1.

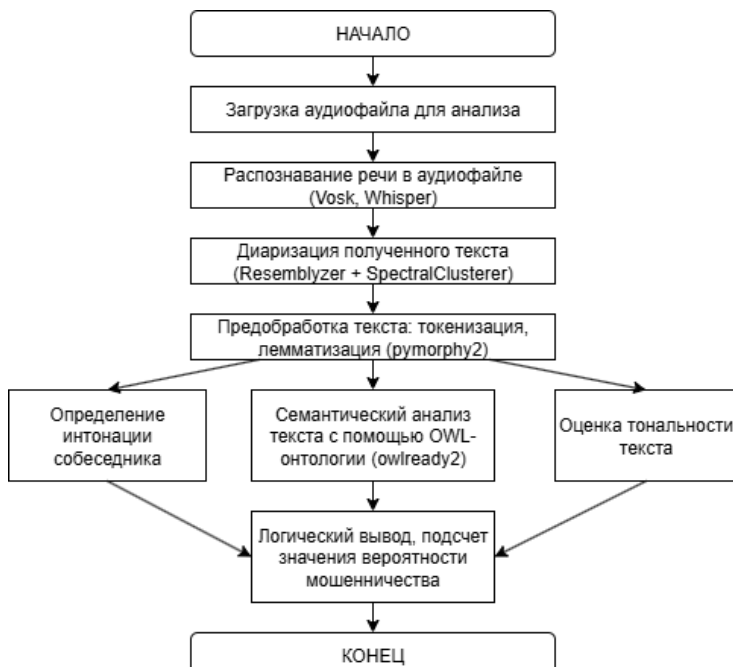


Рис. 1. Блок-схема предложенной методики анализа содержания голосовых аудиофайлов телефонных разговоров

Рассмотрим наиболее важные элементы предложенного подхода:

1. Распознавание речи реализовано с использованием моделей Vosk (на базе Kaldi) и Whisper (от OpenAI):

- *Vosk* – это инструмент для офлайн-распознавания речи, построенный поверх фреймворка Kaldi, одного из самых известных проектов в области ASR (Automatic Speech Recognition). Он использует гибридную архитектуру HMM + DNN, сочетающую статистическое моделирование с нейросетевыми признаками.
- *Whisper* – это открытая нейросетевая модель, разработанная OpenAI, основанная на архитектуре трансформеров. Она относится к классу end-to-end моделей, обученных на огромном объеме мульти-язычных данных, включая шумные и реальные аудиозаписи. Модель, по которой будет произведено распознавание речи, может быть выбрана пользователем в интерфейсе платформы.

2. **Диаризация аудио** – это сегментация аудио по говорящим с целью анализа текста только одного говорящего (предположительного мошенника).
3. **Предобработка текста** включает токенизацию и лемматизацию.
4. **Семантический анализ** предобработанного текста речи собеседника с использованием разработанной OWL-онтологии.

OWL-онтология – это одна из форм графового-семантического представления знаний, реализующая дескрипционную логику 2-ого порядка. Используемая в рамках данного исследования лингвистическая OWL-онтология имеющую следующую структуру:

$$O_F = (T_F, P_F),$$

где T_F – множество объектов классов онтологии, представляющих собой ключевые термины, которые используются при реализации схем телефонного мошенничества с целью кражи денежных средств.

Например, к подобным терминам относятся словосочетания «безопасный счет», «служба безопасности», «подозрительная активность» и пр.

P_F – множество свойств объектов:

$$P_F = \{P_F^L, P_F^S\}$$

где P_F^L , – множество бинарных лингвистических свойств терминов (ObjectProperties). Например, синонимия, антонимия, паронимия и пр.

P_F^S – свойство типа данных (DatatypeProperty), отражающее вероятность принадлежности термина к множеству терминов, употребляемых в телефонных разговорах мошенников, $p_F^S \in P_F^S, p_F^S \in [0,1]$.

В рамках данного исследования набор терминов и отношения между ними формировались специалистом самостоятельно на основе опыта анализа мошеннических телефонных звонков.

В дальнейшем планируется автоматизация процесса формирования этой лингвистической OWL-онтологии путем обработки текстов записей разговоров мошенников и извлечения наиболее аутентичных терминов.

5. **Определение интонации** собеседника реализовано посредством анализа аудиосигнала (среднее значение амплитуды).
6. **Оценка тональности текста** – это классификация текста с применением модели BERT (используемые классы тональности: положительный, отрицательный и нейтральный).
7. **Логический вывод** результата анализа включает:
 - 1) процент вероятности мошенничества;
 - 2) число найденных подозрительных слов в тексте;
 - 3) результат вычисленной интонации.

Таким образом, научной новизной предложенного решения является гибридизация современных нейросетевых моделей, моделей онтологического анализа и NLP в задачах обнаружения мошенничества в аудиозаписях телефонных разговоров.

3. Архитектура программной платформы

Разработанная программная платформа, реализующая предложенную методику, состоит из следующих программных модулей:

- Модуль обработки аудио преобразует записи в формат WAV и выполняет диаризацию речи.
- Модуль распознавания речи преобразует аудио в текст путем использования моделей Vosk и Whisper.
- Модуль обработки текста выполняет лемматизацию и анализ интонационных особенностей речи с применением библиотеки rumorphy2.
- Модуль работы с онтологией обеспечивает хранение и обработку данных с использованием библиотеки owlready2.
- Модуль хранения данных организует работу с базой данных на основе SQLite.
- Модуль визуализации отображает аналитические данные с помощью библиотеки matplotlib.

На рис. 2 представлена диаграмма компонентов разработанной программной платформы.

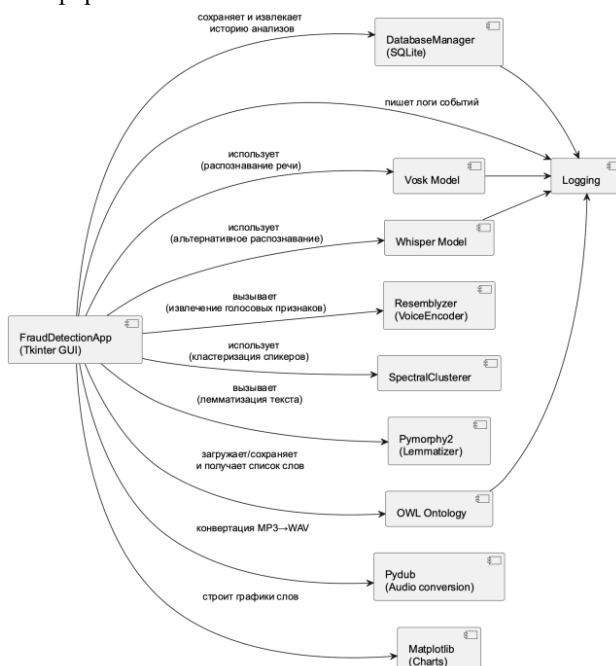


Рис. 2. Диаграмма компонентов

4. Эксперименты по определению мошенничества посредством анализа аудиофайлов телефонных разговоров

Для оценки эффективности разработанной программной платформы был проведён эксперимент на наборе аудиозаписей телефонных разговоров. Аудиозаписи были предоставлены Отделом по борьбе с киберпреступностью УМВД по Ульяновской области. Всего в выборку вошли 10 подтверждённых мошеннических звонков и 5 контрольных «чистых» записей, не содержащих признаков противоправного содержания. Такой набор позволил проверить работоспособность методики не только на единичном примере, но и при сравнительном анализе.

В ходе эксперимента система работала в двух режимах распознавания речи: Vosk и Whisper-small. Дополнительно учитывались результаты онтологического анализа, лемматизации текста и сегментации аудио по говорящим (Resemblyzer + SpectralClusterer). Для каждого разговора фиксировались следующие показатели:

- количество обнаруженных подозрительных терминов из онтологии (всего в базе 120 слов и словосочетаний);
- рассчитанная вероятность мошенничества (на основе весов терминов);
- оценка интонации собеседников (повышенный/спокойный/обычный тон);
- распределение подозрительных слов во времени.

4.1. Результаты на наборе записей

В среднем для мошеннических звонков система выявила от 8 до 15 ключевых терминов (например: «банк», «безопасный счёт», «служба безопасности», «подтверждение», «перевод»). Для контрольных звонков количество совпадений не превышало 2–3 слов, при этом итоговая вероятность мошенничества не превышала 5–7%, что указывает на низкий уровень ложных срабатываний. Для реальных мошеннических разговоров вероятность колебалась в диапазоне 15–42%, в зависимости от насыщенности диалога терминологией из онтологии.

Whisper показал более высокую точность распознавания речи на шумных записях (≈92%) по сравнению с Vosk (≈84%). При этом время обработки Whisper было выше (7–12 секунд против 4–8 секунд для Vosk). Использование онтологического анализа позволило повысить качество классификации: например, при распознавании только по ключевым словам точность составила 75%, а при учёте лемматизации и весов терминов – 83%.

4.2. Сравнительная оценка с другими подходами

Для сопоставления эффективности предложенного решения были рассмотрены результаты, опубликованные для других методов (см. раздел 1b). Итоговые показатели сведены в табл. 1.

Таблица 1

Сравнение точности и скорости различных методов обнаружения
мошенничества

Подход/Система	Точность, %	Время обработки	Особенности
Whisper + Онтология	83,2	7-12с	Учитывает семантику, интонацию, редактируема онтология
Vosk + Онтология	74,3	4-8с	Быстрое онлайн распознавание, хуже при шуме
IBM Watson STT + NLP	86-90	2-4с	Нет поддержки онтологии
Google Speech-to-Next + NLP	85-91	2-4с	Облачное решение
Pindrop (гибридная система)	85-90	1,5-15с	Голосовая биометрия, корпоративный сегмент
RNN-модели	93	60-600с	Высокая точность, но низкая скорость
GAN	88-92	30-180с	Высокие вычислительные затраты

Из табл. 1 видно, что предложенная методика занимает промежуточное положение: по точности она сопоставима с облачными решениями (Google, Watson), но выигрывает за счёт встроенной онтологической базы и анализа интонации. По времени обработки система уступает RNN и GAN, но работает существенно быстрее, что делает её применимой для прикладного использования.

4.3. Выводы по серии экспериментов

Проведённая серия экспериментов показала, что разработанная программная платформа обеспечивает комплексный анализ аудиозаписей, включающий:

- высокоточное распознавание речи на основе современных нейросетевых моделей;
- онтологический анализ, позволяющий учитывать семантические связи терминов;
- оценку интонации и сегментацию по говорящим, что повышает интерпретируемость результата.

Таким образом, система способна не только фиксировать факт употребления подозрительных слов, но и выделять наиболее значимые фрагменты диалога, что делает её полезным инструментом для служб безопасности и противодействия телефонному мошенничеству.

Заключение

В рамках выполнения данного проекта была спроектирована и реализована программная платформа, предназначенная для автоматизированного анализа телефонных разговоров с целью выявления признаков мошенничества. Основу интеллектуальной составляющей системы составляют нейросетевые модели распознавания речи, алгоритмы обработки естественного языка, а также разработанная онтологическая модель, содержащая набор ключевых слов и их весов, отражающих степень вероятности мошеннического характера обращения.

Особое внимание в разработке было уделено построению структурированной и редактируемой онтологии, которая обеспечивает гибкость и расширяемость системы. Благодаря встроенному редактору словаря пользователь может самостоятельно изменять содержание онтологии – добавлять новые термины или удалять устаревшие, что позволяет оперативно адаптировать платформу к меняющимся сценариям мошенничества.

Дальнейшее развитие проекта может быть связано с внедрением онлайн-анализа телефонных звонков в реальном времени, расширением онтологической базы за счёт подключения к внешним источникам знаний, применением более глубоких моделей анализа эмоций, а также интеграцией с системами безопасности организаций для оперативного реагирования на инциденты.

Список литературы

1. Кулешов В.В., Карасев П.И., Х.А.Х. Шамсулдин, Али А.А.Х.Н.А. Обзор методов использования нейросетевого фишинга // Информационная безопасность и защита персональных данных. Проблемы и пути их решения: Сборник материалов и докладов XV межрегиональной научно-практической конференции, Брянск, 28 апреля 2023 года / под общей редакцией О.М. Голембиовской. – Брянск: Брянский государственный технический университет, 2023. – С. 151-153.
2. Трушин И С., Алексеев А.А. Применение современных нейронных сетей в речевых технологиях для задачи разведки // Вестник науки. – 2024. – Т. 4, № 6(75). – С. 1209-1217.
3. Ермаков С.Р., Зубов Н.А. Развитие системы предотвращения подделки голосовой биометрии // Научно-технический вестник Поволжья. – 2024. – № 3. – С. 175-178.
4. IBM Watson Speech to Text. – <https://www.ibm.com/products/speech-to-text> (дата обращения: 27.05.2025).
5. Google Cloud Speech-to-Text. – <https://cloud.google.com/speech-to-text> (дата обращения: 27.05.2025).
6. Pindrop. – <https://www.pindrop.com> (дата обращения: 27.05.2025).
7. Nice Perform. – <https://bslgroupp.com/support-for-discontinued-products/nice-perform> (дата обращения: 27.05.2025).
8. OpenVoice OS. – <https://www.openvoiceos.org> (дата обращения: 27.05.2025).

9. **Lependin A.A., Filin Y.A., Malinin P.v.** Speech Replay Spoofing Attack Detection System Based on Fusion of Classification Algorithms // Известия Алтайского государственного университета. – 2018. – No. 1(99). – P. 107-112.
10. **Zhao Q., Chen K., Li T., Yang Y., Wang X.** Detecting telecommunication fraud by understanding the contents of a call // Cybersecur – 2018. – Vol. 1, No. 8.
11. **Singh G., Singh P., Singh M.** Advanced real-time fraud detection using RAG-based LLMs.
12. **Пономарев К.Г.** Способы генерации голосовых дипфейков и методы их выявления // Молодежь. Наука. Инновации. – 2024. – Т. 1. – С. 172-176.
13. **Pustovoytov P.S., Seilova N.A.** Анализ методов подделки голоса: риски, случаи и стратегии защиты // International Journal of Information and Communication Technologies. – 2024. – Vol. 5, No. 1(17). – P. 120-132.
14. **Пономарев К.Г., Верещагина Е.А.** Математический аппарат и технологическая инфраструктура системы прогнозирования синтетических голосовых дипфейков // Инженерный вестник Дона. – 2024. – № 6(114). – С. 338-353.

УДК 004.85:614.841.2:528.854

doi: 10.15622/rcai.2025.099

ОЦЕНКА ПЛОЩАДЕЙ ВОЗГОРАНИЯ НА ОСНОВЕ СПЕКТРАЛЬНЫХ ИНДЕКСОВ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ¹

Ю.В. Пестова (*yupest@gmail.com*)

О.А. Николайчук (*nikoly@icc.ru*)

Институт динамики систем и теории управления
им. В.М. Матросова СО РАН, Иркутск

В работе представлены результаты применения метода машинного обучения Random Forest для оценки площадей возгорания (гарей) лесных массивов на примере Киренского лесничества Иркутской области с использованием данных дистанционного зондирования Земли и вычислительной платформы Google Earth Engine. Для обучения модели обоснован исходный набор спектральных индексов, подготовлены данные для рассматриваемой территории, предложен алгоритм подготовки и обработки данных, получена точность модели на валидационной выборке в виде оценки Ассигасы 0.975. Результаты будут применены для решения задачи прогнозирования риска лесных пожаров территории.

Ключевые слова: лесные пожары, гарь, площади возгорания, спектральные индексы, методы машинного обучения, Random Forest.

Введение

Лесные пожары являются проблемой во всем мире, оказывая влияние на безопасность людей и инфраструктуры, биологическое разнообразие и экологию. Важным направлением исследований для повышения эффективности управления снижением риска и смягчением последствий лесных пожаров является оценка площадей возгорания (гарей) на основе данных дистанционного зондирования земли и методов машинного обучения.

¹ Работа выполнена в рамках гранта № 075-15-2024-533 Министерства науки и высшего образования РФ на выполнение крупного научного проекта по приоритетным направлениям научно-технологического развития (проект «Фундаментальные исследования Байкальской природной территории на основе системы взаимосвязанных базовых методов, моделей, нейронных сетей и цифровой платформы экологического мониторинга окружающей среды»).

Современные системы дистанционного зондирования Земли предоставляют два принципиально разных типа данных для мониторинга последствий лесных пожаров – оптические и радарные. Оптические датчики (Sentinel-2, Landsat 8/9, MODIS), работающие в видимом, ближнем, среднем и тепловом инфракрасных диапазонах, позволяют анализировать характеристики поврежденной растительности через комбинации мультиспектральных каналов и вычисления специализированных индексов на их основе [Рязанов и др., 2024].

К основным мультиспектральным каналам оптических датчиков относятся каналы: видимые цвета спектра «Blue», «Green», «Red», невидимый ближний инфракрасный канал «NIR», инфракрасные каналы «SWIR-1/2» и др. При дешифровании космических снимков применяются определенные комбинации спектральных каналов в зависимости от типа исследуемых объектов, которые выделяются при картографировании за счет цветового контраста по отношению к фону [Шихов и др., 2020]. Таким образом, применение комбинации спектральных каналов к космическим снимкам формирует композитное изображение, где исследуемые объекты имеют особую окраску, например, тепловые аномалии выглядят красноватыми или желтыми, затопленные территории имеют темно-синий и почти черный цвет, растительность отображается в оттенках темно и светло зеленого. Отличительной особенностью спектрального образа лесных гарей могут быть более низкие, чем у вырубок, коэффициенты отражения в видимой и в ближней инфракрасных зонах спектра. Это обусловлено наличием продуктов горения (зола) на поверхности почвы.

В дешифровании данных дистанционного зондирования Земли широкое распространение получило применение вегетационных индексов, которые представляют собой математические комбинации отражательной способности в различных спектральных диапазонах, используемые для оценки состояния растительности. На рубеже 2000-х годов в ряде исследований [Vilar et al., 2015], [Belhadj-Khedher et al., 2018], [Bentekhici et al., 2020], [Kurbanov et al., 2022] были использованы данные спутников AVHRR, MODIS, IRS, SPOT/HRVIR и Landsat для определения термических точек по разностному спектральному индексу растительности dNDVI. В качестве альтернативы NDVI использовался нелинейный индекс глобального мониторинга окружающей среды (GEMI) [Pinty et al., 1992], показавший более высокие результаты при определении гарей [Guindos-Rojas et al., 2018]. Помимо GEMI для оценки площади гарей используется индекс BAI, сочетающий минимальное отражение выжженной растительности в ближнем инфракрасном диапазоне и максимальное отражение в красном диапазоне. BAI был создан для определения гарей на изображениях NOAA-AVHRR [Chuvieco et al., 2002], в дальнейшем его

модификация была разработана под изображения MODIS в ближнем и среднем инфракрасных диапазонах (BAIM) для картографирования свежих гарей в средиземноморских экосистемах [Martín et al., 2006]. Высокое спектральное различие между сгоревшими и нетронутыми территориями позволяет определить двухспектральный индекс среднего инфракрасного диапазона (MIRBI) [Trigg et al., 2001], [Barboza Castillo et al., 2020], [Hawbaker et al., 2020]. MIRBI эффективно устраняет ошибки, возникающие при анализе смешанных пикселей на границе воды и земли, а также на участках, ранее подвергшихся пожарам [Santana et al., 2018]. Наряду с индексом NDVI, который остается одним из основных индексов, используемых для оценки лесных пожаров на основе данных дистанционного зондирования, в последнее время активно применяется нормализованный коэффициент выгорания (NBR) [De Santis et al., 2009], [Veraverbeke et al., 2011], [He et al., 2019]. Итальянские ученые на примере северо-восточной части острова Сицилия отмечают недостатки применения NBR: наличие водоемов или облаков приводит к увеличению значений пикселей, что неизбежно приводит к ошибкам оценки гарей. На примере снимков, полученных с Sentinel-2, вводится индекс NBR+, учитывающий отражательную способность воды сине-зеленых (Blue, Green) волн [Alcaras et al., 2022]. Также используется индекс dNBR, представляющий собой разность значений NBR до и после пожара [Tian et al., 2018], [Salguero et al., 2020]. Единого мнения среди исследователей об особенностях применения различных индексов до сих пор не выработано, однако установлено, что индекс RdNBR обеспечивает более высокую точность оценки степени ожога, чем dNBR [Cansler et al., 2012], также не удалось определить закономерности между данными индексами и степенью выгорания, которая меняется в зависимости от площади пожара и типа экосистемы (саванна, бореальные и тропические леса).

Целью данного исследования является решение задачи повышения точности оценки площадей возгорания на основе классификации спектральных индексов методом машинного обучения Random Forest.

1. Обоснование выбора спектральных индексов

Для создания модели оценки площадей возгорания выполнено обоснование вегетационных индексов для формирования набора влияющих факторов на основе визуализации динамики изменения растительного покрова и корреляционного анализа.

Для вегетационных индексов: чувствительных к растительности, к пигментам (SIPI) и хлорофиллу (CIRE, GCI), влажности (NDWI), старению (PSRI), выгоранию и устойчивых к почве (SAVI, MSAVI) были со-

ставлены композиты снимков за 30 дней до пожара, сразу после пожара и через год после пожара в Катангском районе Иркутской области за период 07.08.2020 г. – 05.09.2020 г. Анализ композитов позволил провести оценку пригодности различных спектральных индексов для включения в модель оценки площадей возгорания и постпожарных последствий.

Индексы, традиционно используемые для оценки состояния растительности (ARVI, DVI, EVI, GARI, IPVI, NDVI, RVI, VARI), демонстрируют высокую чувствительность к восстановлению растительного покрова и степени повреждений, особенно на интервале через год после пожара. Среди них NDVI, ARVI и RVI наглядно отражают восстановительные процессы.

Для выявления участков, подвергшихся интенсивному выгоранию, наибольшую информативность показывают индексы, специально разработанные для оценки последствий пожаров – BAI, BAIS2, NBR, NBR2 и NBRPLUS. Эти индексы позволяют зафиксировать изменения в спектральных характеристиках поверхности непосредственно после воздействия огня. Наибольшее покрытие возгорания демонстрируются на картах вегетационных индексов BAI и BAIS2 сразу после пожара, а также NBRPLUS через год. Карта NBR в большей степени отражает степень повреждения растительного покрова.

Индексы, чувствительные к биохимическим характеристикам растительности (SIPI, CIRE, GCI, PSRI, SAVI, MSAVI), играют важную роль в более детальной интерпретации степени деградации покрова (). PSRI позволяет отследить процессы старения и деградации листвы, заметные сразу после пожара, а спустя год – восстановительные процессы. Индексы SIPI и CIRE дают представление о фотосинтетической активности, а SAVI и MSAVI показывают устойчивость к фоновому влиянию почвы, что особенно ценно в редколесье и на открытых участках.

Для уточнения результатов визуального анализа для каждого индекса рассчитаны дифференциальные показатели как разница между средними медианами значений за период в течении месяца после пожара и через год. Таким образом, был проведен сравнительный анализ чувствительности вегетационных индексов к подверженности пожарам. Результаты представлены с помощью визуализации тепловой карты, на которой интенсивность цвета подсвечивает наибольшие и наименьшие изменения в вегетационных индексах за периоды, на основе чего определен период изменения (рис. 1).

Группа индексов	Спектральный индекс	Период изменения	Изменение через месяц	Изменение через год
Биохимические	SIPI	месяц	0,050	0,010
	PSRI	месяц	0,039	0,005
	NDWI	месяц	-0,006	0,028
	SAVI	год	-0,021	-0,023
	MSAVI	год	-0,021	-0,022
	GCI	год	-0,311	-0,744
Выгорание	CIRE	год	-0,039	-0,130
	BAIS2	месяц	0,050	0,043
	BAI	месяц	20,906	10,423
	NBRPLUS	год	-0,010	-0,048
	NBR2	год	-0,016	-0,018
	NBR	год	-0,041	-0,045
Растительность	VARI	месяц	-0,052	-0,042
	EVI	месяц	-0,029	-0,016
	DVI	месяц	-0,013	-0,011
	RVI	год	-0,205	-0,668
	NDVI	год	-0,017	-0,042
	IPVI	год	-0,009	-0,021
	GARI	год	-0,019	-0,031
	ARVI	год	0,051	-0,014

Рис. 1. Сравнение чувствительности вегетационных индексов к изменениям растительного покрова после пожара через месяц и год

2. Обучение модели Random Forest

В качестве основных входных данных для обучения модели классификации гарей использовались скорректированные на атмосферу сцены спутника Sentinel-2 (коллекция COPERNICUS / S2_SR_HARMONIZED). В качестве геометрии области исследования выбрана территория Киренского лесничества Иркутской области. Сбор, обработка данных, обучение модели и ее интерпретация проводились на облачной геопространственной платформе Google Earth Engine, которая предоставляет доступ к архиву спутниковых данных и инструментам для их анализа, в частности, в научных и некоммерческих целях.

Разметка обучающей выборки произведена в геоинформационной системе QGIS, были подготовлены 375 точечных образцов с целевой бинарной переменной «class», где 1 – «гарь», а 0 – «не гарь». Затем выполнена интеграция факторов, характеризующих возгорания на основе спектральных характеристик спутниковых снимков. Для этого были рассчитаны вегетационные индексы, которые формируют признаковое пространство обучающего набора данных. Его формирование основано на следующих этапах сбора и подготовки данных:

1. Сбор данных. Коллекции снимков Sentinel-2 получены за период с марта по октябрь с различными интервалами: за 2019 г.; за 2020 г.; помесечно за 2019 г.

2. Маскирование облаков и атмосферных артефактов. Для каждой сцены Sentinel-2 использовались QA-каналы (QA60) с масками кучевых и перистых облаков (битовые маски 10 и 11). Пиксели с облаками исключены, а значения отражательной способности приведены к диапазону [0, 1] делением на 10 000.

3. Расчет спектральных индексов. Для каждого изображения рассчитаны вегетационные индексы: NDVI, EVI, SAVI, MSAVI, ARVI, GARI, VARI, IPVI, RVI, DVI, SIPI, PSRI, CIRE, BAIS2, BAI, GCI, NDWI, NBR, NBR2, NBRPLUS.

4. Агрегирование вегетационных индексов на основе композита снимков. Получены медианные изображения по каждому временному интервалу.

5. Генерация временных разностей на основе месячных изменений. Для каждого текущего и последующего помесечного интервала итеративно рассчитаны дифференциальные значения вегетационных индексов. Полученные разностные изображения сформировали коллекцию снимков, которая была приведена к двум композитам снимков: максимальных и минимальных изменений.

6. Генерация временных разностей на основе годовых изменений. На основе медианных композитов снимков за 2019 г. и 2020 г. было получено изображение разности вегетационных индексов.

Итоговое изображение включает годовые (названия индексов имеют префикс «d» и постфикс «_year»), месячные дельты (с префиксом «d» в названии) и экстремумы месячных медианных композитов вегетационных индексов. Зона исследования дополнительно ограничена порогом $NDWI < -0.1$ и маской растительности для исключения пикселей водного покрова Земли.

В результате этапа подготовки обучающего набора данных была проведена интеграция рассчитанных значений вегетационных индексов подготовленного изображения с набором размеченных точечных образцов по разрешению 30 м.

Для определения набора независимых факторов из признакового пространства вегетационных индексов осуществлен анализ корреляции. Матрица корреляции, рассчитанная по метрике Спирмана отражает только те значения индексов, которые оказывают значительное влияние на целевую переменную «class» в порядке убывания (в первом столбце таблицы на рисунке 2). Исходя из значений корреляции, основными факторами, которые в наибольшей степени обладают спектральными характеристиками возгораний, являются: годовые разностные dNBR_year, dBAI_year, dBAIS2_year, dNBR2_year, экстремумы BAI, BAIS2, NBR и экстремумы месячных изменений dBAI, dNBR. При этом BAI и его модификация BAIS2 имеют практически одинаковые спектральные характеристики, что

отражено высоким показателем корреляции. Для спектральных индексов группы, чувствительных к пигменту и хлорофиллу корреляционный анализ показывает умеренную связь целевой переменной «class» с индексом годовой разности dSIPI_year, указывающий на стресс хлорофилла. Среди вегетационных индексов, чувствительных к почве, воде и старению, наибольшую, но несущественную зависимость с гарями имеет индекс PSRI. Он отражает старение растений по соотношению каротиноидов и хлорофилла за счет спектральных характеристик и также чувствителен к гарям, поскольку косвенно указывает на снижение водного статуса и повышение сухости растительности, что характерно для зон, подвергшихся воздействию огня и стрессовым условиям.

	class	dNBR_year	BAI	BAIS2	dBAI	dNBR2_year	dNBRPLUS_year	dBAIS2	dNBR	NBR	dBAIS2_year	dBAI_year
dNBR_year	0.789	1.000	0.662	0.684	0.497	0.714	0.590	0.360	-0.472	-0.630	-0.824	-0.794
BAI	0.762	0.662	1.000	0.910	0.629	0.506	0.325	0.407	-0.525	-0.720	-0.578	-0.656
BAIS2	0.740	0.684	0.910	1.000	0.660	0.482	0.356	0.508	-0.628	-0.871	-0.589	-0.617
dBAI	0.659	0.497	0.629	0.660	1.000	0.312	0.164	0.755	-0.843	-0.649	-0.397	-0.517
dNBR2_year	0.622	0.714	0.506	0.482	0.312	1.000	0.751	0.181	-0.259	-0.439	-0.727	-0.583
dNBRPLUS_year	0.489	0.590	0.325	0.356	0.164	0.751	1.000	0.069	-0.151	-0.368	-0.661	-0.414
dBAIS2	0.460	0.360	0.407	0.508	0.755	0.181	0.069	1.000	-0.759	-0.498	-0.307	-0.346
dNBR	-0.603	-0.472	-0.525	-0.628	-0.843	-0.259	-0.151	-0.759	1.000	0.733	0.358	0.422
NBR	-0.676	-0.630	-0.720	-0.871	-0.649	-0.439	-0.368	-0.498	0.733	1.000	0.532	0.507
dBAIS2_year	-0.727	-0.824	-0.578	-0.589	-0.397	-0.727	-0.661	-0.307	0.358	0.532	1.000	0.755
dBAI_year	-0.768	-0.794	-0.656	-0.617	-0.517	-0.583	-0.414	-0.346	0.422	0.507	0.755	1.000

Рис. 2. Матрица корреляции вегетационных индексов, чувствительных к выгоранию и целевой переменной «class»

Размеченная выборка подверглась случайному разбиению в соотношении 70 к 30 на тренировочную и валидационную выборки, соответственно. Первая используется для обучения классификатора, а вторая – для оценки его точности.

Для бинарной классификации («гарь» и «не гарь») использована реализация алгоритма метода машинного обучения Random Forest (метод из Google Earth Engine «smileRandomForest») с гиперпараметром числа деревьев – 32, набором признаков из наиболее информативных индексов: dNBR_year, BAI, dBAI, dNBR2_year, dNBRPLUS_year, dBAIS2, dVARI_year, dARVI_year, PSRI и прогнозируемым классом «class».

Оценка качества модели измерена метрикой точности Ассигасу, что составила на валидационной выборке 0.975. Точечные образцы на территории гарей имеют ошибку предсказания в 6%, что может быть по причине сохранившейся растительности внутри границ возгорания (рис. 3).

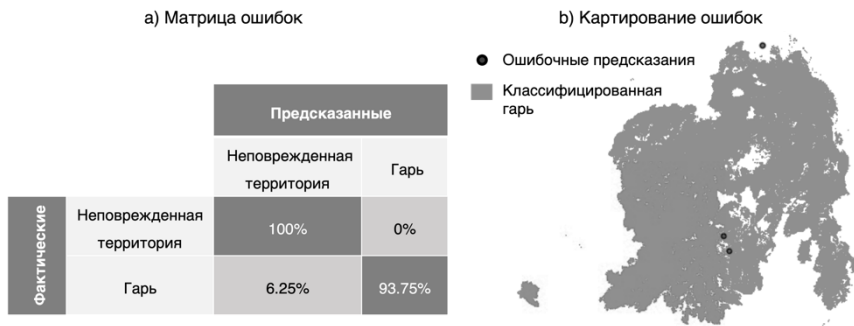


Рис. 3. Матрица ошибок результатов классификации модели (а) и ошибочные предсказания модели на примере площади возгорания (б)

Анализ важности факторов показал наиболее значимыми вегетационные индексы, чувствительные к выгоранию (dNBR_year, BAI, dBAI, dNBR2_year), наименее – к растительности (dVARI_year, dARVI_year), при этом индекс PSRI оказывает большее влияние, чем индексы возгорания dBAIS2 и dNBRPLUS_year (рис. 4). За счет этого модель четче распознает границы возгораний и менее чувствительна к вырубкам.

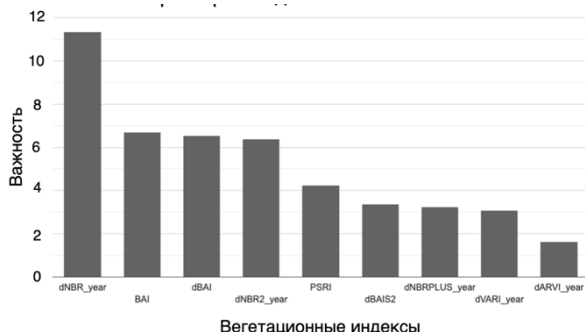


Рис. 4. Важность факторов модели Random Forest

3. Оценка эффективности модели

Сопоставление полученных границ гарей и эталонных пожаров, которые предоставлены Институтом космических исследований РАН, проводилось с целью оценки площадей возгорания и их уточнения для построения моделей прогнозирования риска пожарной опасности.

Индекс пересечения с объединением (IoU, Intersection over Union) является одной из ключевых метрик для оценки качества пространственного совпадения между предсказанными и эталонными контурами пожа-

ров. В задачах классификации и детекции объектов на основе геоданных IoU используется для количественной оценки степени совпадения между двумя полигонами: предсказанным участком возгорания и эталонным контуром пожара. Метрика рассчитывается как отношение площади пересечения двух геометрий к площади их объединения. Оценка IoU составила 0.58, что свидетельствует о среднем уровне совпадения между результатами классификации и фактическими границами пожаров. В контексте геопространственного анализа принято считать, что значения IoU выше 0.5 указывают на приемлемое качество пространственного соответствия, в то время как значения ниже 0.5 свидетельствуют о существенных расхождениях. Таким образом, результат в 0.58 подтверждает, что модель успешно идентифицирует основные очаги возгораний (рис. 5).

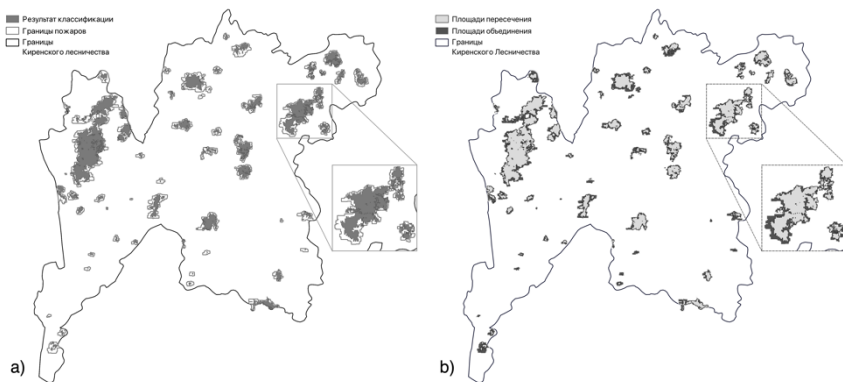


Рис. 5. Площади предсказанных гарей и эталонных пожаров (a) и визуализация IoU (b)

Следует отметить, что при визуальном анализе пространственного расположения контуров классифицированных гарей (белые линии) и эталонных границ (черные линии) видно, что автоматическая классификация зачастую точнее отражает фактическую конфигурацию гарей (рис. 6). Также отметим, что данные о границах гарей в эталонном наборе данных получены путем автоматического распознавания термоточек с точностью один километр на один пиксель, а затем уточнены вручную путем разметки. В ряде случаев наблюдается, что эталонные данные оказываются менее детализированными или содержат смещения, что может объяснять невысокое значение метрики IoU, однако, визуальная оценка изображений подтверждает более высокую точность и достоверность предложенной модели для выделения гарей (рис. 6).

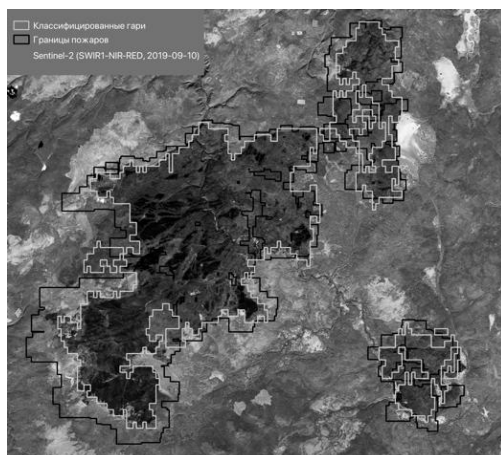


Рис. 6. Сравнение контуров границ возгораний в результате классификации и пожаров на спутниковом снимке Sentinel-2 в синтезе каналов SWIR-1, NIR и RED

Заключение

Полученные результаты будут применены для решения задачи прогнозирования риска лесных пожаров территории в качестве отдельного этапа верификации исходных данных о лесных пожарах, где границы гарей важны при генерации событий «отсутствие пожара» для повышения точности прогнозирования.

Список литературы

- [Рязанов и др., 2024] Рязанов С.С., Кулагина В.И. Обзор российских и иностранных источников мультиспектральных снимков для создания систем агроэкологического мониторинга // Российский журнал прикладной экологии. – 2024. – № 2. – С. 4-18. – doi: 10.24852/2411-7374.2024.2.04.18.
- [Шихов и др., 2020] Шихов А.Н., Герасимов А.П., Пономарчук А.И., Перминова Е.С. Тематическое дешифрирование и интерпретация космических снимков среднего и высокого пространственного разрешения / Перм. гос. нац. исслед. у-т. – Пермь, 2020. – С. 33.
- [Vilar et al., 2015] Vilar L., Camia A., San-Miguel-Ayanz J. A comparison of remote sensing products and forest fire statistics for improving fire information in Mediterranean Europe // Eur. J. Remote Sens. – 2015. – Vol. 48. – P. 345–364. – doi: 10.5721/EuJRS20154820.
- [Belhadj-Khedher et al., 2018] Belhadj-Khedher C., Koutsias N., Karamitsou A., El-Melki T., Ouelhazi B., Hamdi A., Nouri H., Mouillot F. A revised historical fire regime analysis in Tunisia (1985–2010) from a critical analysis of the National fire database and remote sensing // Forests. – 2018. – Vol. 9, 59. – doi: 10.3390/f9020059.

- [Bentekhici et al., 2020] Bentekhici N., Bellal S.A., Zegrar A. Contribution of remote sensing and GIS to mapping the fire risk of Mediterranean forest case of the forest massif of Tlemcen (North-West Algeria) // *Nat. Hazards*. – 2020. – Vol. 104. – P. 811-831. – doi: 10.1007/s11069-020-04191-6.
- [Kurbanov et al., 2022] Kurbanov E., Vorobev O., Lezhnin S., Sha J., Wang J., Li X., Cole J., Dergunov D., Wang Y. Remote Sensing of Forest Burnt Area, Burn Severity, and Post-Fire Recovery: A Review // *Remote Sens.* – 2022. – Vol. 14, 4714. – doi: 10.3390/rs14194714.
- [Pinty et al., 1992] Pinty B., Verstraete M.M. GEMI: A non-linear index to monitor global vegetation from satellites // *Vegetatio*. – 1992. – Vol. 101. – P. 15-20. – doi: 10.1007/BF00031911.
- [Guindos-Rojas et al., 2018] Guindos-Rojas F., Arbelo M., García-Lázaro J.R., Moreno-Ruiz J.A., Hernández-Leal P.A. Evaluation of a Bayesian algorithm to detect burned areas in the Canary Islands' dry woodlands and forests ecoregion using MODIS data // *Remote Sens.* – 2018. – Vol. 10, 789. – doi: 10.3390/rs10050789.
- [Chuvieco et al., 2002] Chuvieco E., Martín M.P., Palacios A. Assessment of different spectral indexes in the red-near-infrared spectral domain for burned land discrimination // *Int. J. Remote Sens.* – 2002. – Vol. 23. – P. 5103-5110. – doi: 10.1080/01431160210153129.
- [Belhadj-Khedher et al., 2006] Martín M.P., Gómez I., Chuvieco E. Burnt Area Index (BAIM) for burned area discrimination at regional scale using MODIS data // *For. Ecol. Manag.* – 2006. – Vol. 234, S221. – doi: 10.1016/J.FORECO.2006.08.248.
- [Martín et al., 2001] Trigg S., Flasse S. An evaluation of different bi-spectral spaces for discriminating burned shrub-savannah // *Int. J. Remote Sens.* – 2001. – Vol. 22. – P. 2641-2647. – doi: 10.1080/01431160110053185.
- [Barboza Castillo et al., 2020] Barboza Castillo E., Turpo Cayo E.Y., de Almeida C.M., Salas López R., Rojas Briceño N.B., Silva López J.O., Barrena Gurbillón M.Á., Oliva M., Espinoza-Villar R. Monitoring Wildfires in the Northeastern Peruvian Amazon Using Landsat-8 and Sentinel-2 Imagery in the GEE Platform // *ISPRS Int. J. Geo-Inf.* – 2020. – Vol. 9, 564. – doi: 10.3390/ijgi9100564.
- [Hawbaker et al., 2020] Hawbaker T.J., Vanderhoof M.K., Schmidt G.L., Beal Y.-J., Picotte J.J., Takacs J.D., Falgout J.T., Dwyer J.L. The Landsat burned area algorithm and products for the conterminous United States // *Remote Sens. Environ.* – 2020. – Vol. 244, 111801. – doi: 10.1016/j.rse.2020.111801.
- [Santana et al., 2018] Santana N.C., De Carvalho Júnior O.A., Gomes R.A.T., Guimarães R.F. Burned-area detection in Amazonian environments using standardized time series per pixel in MODIS data // *Remote Sens.* – 2018. – Vol. 10, 1904. – doi: 10.3390/rs10121904.
- [De Santis et al., 2009] De Santis A., Chuvieco E. GeoCBI: A modified version of the Composite Burn Index for the initial assessment of the short-term burn severity from remotely sensed data // *Remote Sens. Environ.* – 2009. – Vol. 113. – P. 554-562. – doi: 10.1016/j.rse.2008.10.011.
- [Veraverbeke et al., 2011] Veraverbeke S., Lhermitte S., Verstraeten W., Goossens R. A time-integrated MODIS burn severity assessment using the multi-temporal differenced normalized burn ratio (dNBRMT) // *Int. J. Appl. Earth Obs. Geoinf.* – 2011. – Vol. 13. – P. 52-58. – doi: 10.1016/j.jag.2010.06.006.

- [He et al., 2019] He Y., Chen G., De Santis A., Roberts D.A., Zhou Y., Meentemeyer R.K. A disturbance weighting analysis model (DWAM) for mapping wildfire burn severity in the presence of forest disease // Remote Sens. Environ. – 2019. – Vol. 221. – P. 108-121. – doi: 10.1016/j.rse.2018.11.015.
- [Alcaras et al., 2022] Alcaras E., Costantino D., Guastaferro F., Parente C., Pepe M. Normalized Burn Ratio Plus (NBR+): A New Index for Sentinel-2 Imagery // Remote Sens. – 2022. – Vol. 14, 1727. – doi: 10.3390/rs14071727.
- [Tran et al., 2018] Tran B.N., Tanase M.A., Bennett L.T., Aponte C. Evaluation of spectral indices for assessing fire severity in Australian temperate forests // Remote Sens. – 2018. – Vol. 10, 1680. – doi: 10.3390/rs10111680.
- [Salguero et al., 2020] Salguero J., Li J., Farahmand A., Reager J.T. Wildfire trend analysis over the contiguous United States using remote sensing observations // Remote Sens. – 2020. – Vol. 12, 2565. – doi: 10.3390/rs12162565.
- [Cansler et al., 2012] Cansler C.A., McKenzie D. How robust are burn severity indices when applied in a new region? Evaluation of alternate field-based and remote-sensing methods // Remote Sens. – 2012. – Vol. 4. – P. 456-483. – doi: 10.3390/rs4020456.

УДК 004.94

doi: 10.15622/rcai.2025.100

МОДЕЛИРОВАНИЕ ПЕРЕДАЧИ ИМПЛИЦИТНОГО СМЫСЛА В УСТНОЙ РЕЧИ

В.Л. Стефанюк (*stefanuk@iitp.ru*) ^{А,В}

Л.В. Савинич (*savinitch@iitp.ru*) ^А

^А Институт проблем передачи информации им. А.А. Харкевича РАН,
Москва

^В Российский университет дружбы народов, Москва

Работа посвящена моделированию процесса передачи имплицитной информации, то есть информации, содержащей скрытый смысл высказывания. Указываются языковые стратегии, используемые для передачи, выявления и интерпретации имплицитного смысла. Исследуются просодические средства, такие как: интонация, высота тона, ударение, длительность звука, интенсивность, темп, которые широко используются в речи при передаче подтекста и которые выступают маркерами для выявления и интерпретации скрытого смысла в интерактивном общении. Анализируемые примеры сопровождаются аудиозаписью и графиками движения основного тона и интенсивности звука.

Ключевые слова: моделирование, импликация, информация, просодия.

Введение

Настоящая статья продолжает исследование понятия импликации и способов её выражения в речевом потоке. Научная новизна исследования заключается в моделировании процесса употребления просодических средств, мотивации их использования для передачи скрытого смысла, а также их роли в адекватном понимании высказывания. Данный анализ вносит определенный вклад в важное направление искусственного интеллекта – моделирование рассуждений, – одним из основателей которого является Д.А. Поспелов [Поспелов, 1989].

Модель передачи имплицитного смысла предполагает, что говорящий использует разнообразные языковые стратегии, чтобы передать смысл, выраженный неявно. Эти стратегии основаны на понимании слушателем

непрямого, скрытого смысла, подразумеваемого говорящим. Слушатель, в свою очередь, замечая семантическую неполноту, недосказанность в потоке речи, должен выявить, осмыслить и интерпретировать содержащийся скрытый смысл высказывания, основываясь на общих знаниях, опыте и контексте ситуации.

К таким языковым стратегиям относят, например, сарказм, иронию, слова и фразы в переносном смысле, недоговорённость, а также просодические средства.

Просодические средства – это единицы языка, которые даже носители языка порой воспринимают с трудом и далеко не всегда могут повторить интонационную конструкцию, хотя сами постоянно, бессознательно её воспроизводят.

Просодические средства включают в себя не только интонацию, ударение, высоту тона, но также интенсивность звучания, темп речи, длительность звука, паузы, артикуляцию, которые широко используются в интерактивном общении [Янко, 2001], [Кодзасов, 2009], [Savinitch, 2019], [Oxford Handbook of Language Prosody, 2020], [Couper-Kuhlen, 2021], [Stefanuk et al., 2023].

По сути, эти просодические средства могут выступать в качестве контекстуальных подсказок, помогающих передать имплицитный смысл – «смысл, выходящий за рамки того, что „дано“ самой языковой формой, или того, что буквально „сказано“» [Verschueren, 1999]. Другими словами, имплицитный смысл – это то, что подразумевается, но не выражается явно.

1. Материал и методы

Материалом исследования послужили примеры, взятые из русского дискурса, а именно: из речи государственного обвинителя в суде, а также из чтения литературного произведения профессиональным артистом.

Данная работа посвящена детальному анализу просодических средств из выступления известного советского сатирика.

Анализ и моделирование передачи имплицитной информации проводились в нескольких аспектах: семантическом, сопоставляющим значения слов; прагматическом, описывающим стратегии говорящего; стилистическом, в сочетании с иронией; социокультурном, учитывающим контекст ситуации, то есть социальную, культурную и историческую обстановку, в которой происходит общение.

Анализируемые примеры записаны на аудионоситель в режиме реального времени, поэтому наш доклад сопровождается аудиозаписью. Физические параметры записи представлены на графиках, которые мы называем тонограммами.

2. Использование иронии при передаче имплицитного смысла

Рассматриваемый пример передачи имплицитного смысла иллюстрирует использование просодических ресурсов в сочетании со стилистическим приемом иронии.

Ирония – это стилистический оборот, в котором преднамеренно утверждается противоположное тому, что думают о лице или предмете; тонкая, скрытая насмешка.

Существует значительное количество литературы по просодии иронии [Attardo et al., 2013], где предпринимаются попытки описать «иронический тон голоса». Аналогичным образом, в работе [Bryant, 2013] автор сосредоточен на иронической коммуникации – в частности, на вербальной иронии в сочетании с просодией речи.

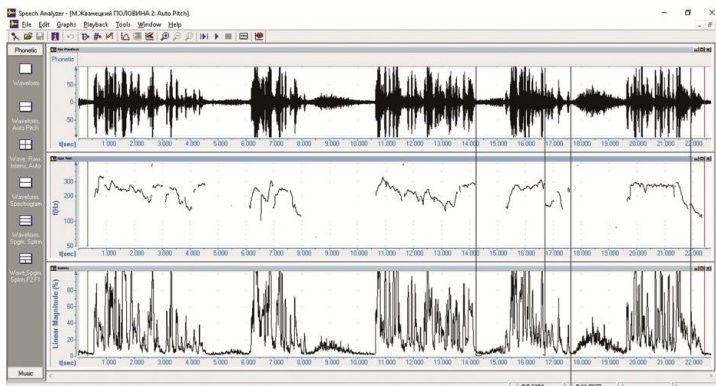
Предлагаемый нами пример был выбран из 15 миниатюр, записанных на аудионоситель во время 45 минутного выступления известного сатирика Михаила Жванецкого. В анализируемых предложениях стрелки, расположенные после соответствующих слов, указывают направление движения тона.

«Половина пассажиров следит за тем, чтобы другая половина брала билеты↗ в трамвае, и первая половина уже не берёт↗, – она на службе ↘. Таким↗ образом, ездит по билетам только половина ↘ населения.»

В этом отрывке сатирик иронически высмеивает тех пользователей общественного транспорта, которые, в ожидании прибытия в пункт назначения, следят за тем, чтобы все пассажиры брали билеты за проезд. Однако они сами билеты не берут, потому что как бы занимаются общественно полезной деятельностью – как будто «на службе». Таким образом, только половина населения ездит по билетам, нанося убытки экономике.

В первом предложении ирония выражена столкновением одновременно активируемых значений слова *служба* – лексического и контекстуального. Основное лексическое значение слова *служба* – «работа, исполнение каких-либо обязанностей» [Словарь рус. яз., 1984: 143]. В ситуации, описываемой сатириком, слово приобретает новое, контекстуальное, значение: добровольно принять на себя обязанности общественного «контролера» (как бы «на службе») и поэтому не покупать билет за проезд. Оба значения находятся в явном противоречии друг с другом (рис. 1).

На средней панели тонограммы 1, в сегменте между третьей и четвертой вертикальными линиями, график тонограммы иллюстрирует резкое падение высоты тона на ударных и заударных слогах носителя акцента *на службе ↘* в диапазоне от 226 Гц до 146,5 Гц. При этом падение тона сопровождается удлинением ударного гласного, которое длится 571 мсек. Для сравнения, предшествующий акцентоноситель *берёт↗*, состоящий из двух слогов, длится 268 мсек.

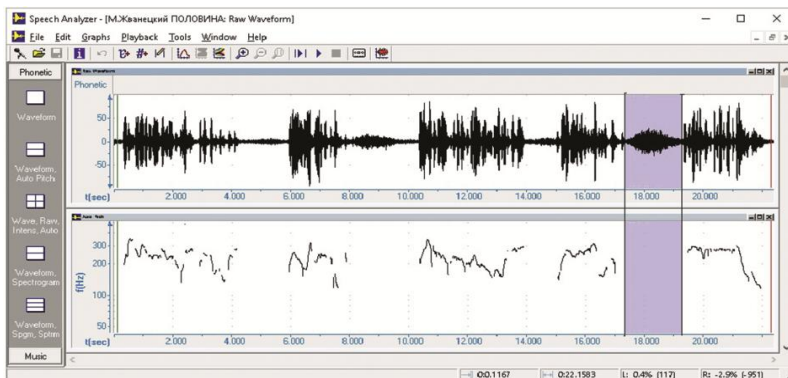


билеты↗ не берёт↗
 'tickets↗' 'does not take↗'

Рис. 1. Тонограмма 1

На нижней панели тонограммы 1 представлен график интенсивности звука, отмечающий увеличение громкости на носителе акцента *на службе*↗. Таким образом, проиллюстрированные просодические признаки (резкое падение высоты тона на ударных и заударных слогах, удлинение ударного гласного и усиление громкости) указывают на усиленное выделение словосочетания *на службе*↗, тем самым подчёркивая важную смысловую часть высказывания.

Далее, после выделения слова, сатирик выдерживает продолжительную паузу, которая длится 2 сек. На тонограмме 2 (рис. 2) пауза обрамлена двумя вертикальными линиями.



на службе
 'in the service'

Рис. 2. Тонограмма 2

На нижней панели выделенного сегмента график тонограммы отсутствует, что указывает на незаполненную паузу (т.е. тишину). В то же время на осциллограмме верхней панели в сегменте паузы зафиксированы возрастающие и убывающие колебания, вызванные фоновым шумом смеха аудитории. То есть запланированная пауза выдерживается сатириком после слова, несущего основное ироническое содержание, которое зачастую трудно осознать сразу. Поэтому сатирик оставляет длительную паузу, чтобы дать аудитории время осмыслить и отреагировать на имплицитный смысл миниатюры. Этой стратегии сатирик последовательно придерживается на протяжении всего выступления.

3. Грамматические аномалии

Завершая анализ, рассмотрим ряд сознательных нарушений правил русской грамматики, в которых говорящий стремится выделить семантически значимое высказывание. В заключительном предложении сатирик произносит кульминационную фразу миниатюры с изменением ударения в слове:

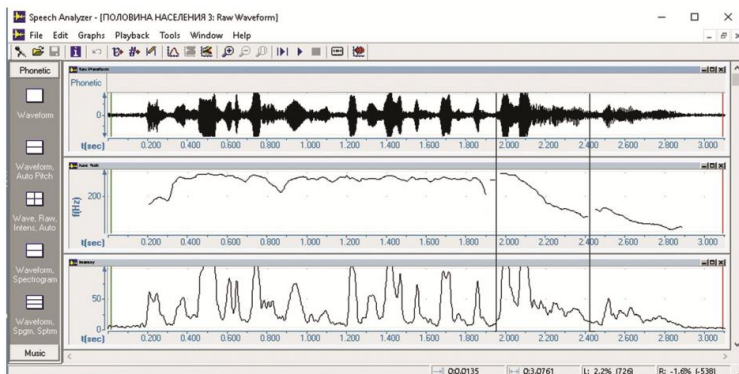
Таким→ образом, ездит по билетам только половина↘ населения.

Что касается переноса ударения, необходимо напомнить, что с акцентом в повествовательных предложениях произносятся определенные словоформы в соответствии с некоторыми базовыми принципами русского языка [Ковтунова, 1976], [Светозарова, 1993]. Носителем акцента в данном словосочетании должна быть словоформа *население* (*половина населения*). Однако сатирик стратегически переносит ударение на другую словоформу – *Половина↘* с резким падением высоты тона на ударных и заударных слогах.

На средней панели тонограммы 3 (рис. 3) две вертикальные линии ограничивают значительное понижение высоты тона на слове *Половина↘* в диапазоне от 297 Гц до 137 Гц.

Нижняя панель тонограммы 3 иллюстрирует интенсивность звука – акустическое свойство громкости. В выделенном сегменте уровень интенсивности на слове колеблется от –22,1 дБ до 0,3 дБ, а пик интенсивности приходится на первый слог – *Половина*. Кроме того, интересно, что пик интенсивности на первом слоге не совпадает с нормативным фиксированным ударением, которое должно приходиться на третий слог лексемы: *полоВина*.

Этот перенос фиксированного ударения в начальную позицию, а также усиление громкости звука объясняются прагматическим намерением говорящего. Подобная прагматическая стратегия подчеркивает семантическое содержание высказывания и, в конечном итоге, способствуют иронической передаче подтекста.



Таким образом, ездит по билетам только половина населения.
 'Thus, only half of the population travels by ticket.'

Рис. 3. Тонограмма 3

Заключение

Таким образом, в настоящей статье рассмотрена проблема передачи имплицитной информации в речи, наряду со стратегиями её использования. Проиллюстрированы и детально описаны просодические средства языка, которые способствуют передаче скрытого смысла высказывания. Эти просодические средства выступают маркерами для выявления и интерпретации имплицитного смысла.

С этой целью выявлены и проанализированы такие просодические элементы, как интонация, высота тона, интенсивность звучания, длительность, темп речи, пауза.

В результате исследования, нами представлена компьютерная модель, которая средствами искусственного интеллекта позволяет строить модель процесса при передаче имплицитной информации, что внесёт определённый вклад в важное направление искусственного интеллекта – моделирование рассуждений.

Список литературы

- [Ковтунова, 1976] Ковтунова И.И. Современный русский язык: Порядок слов и актуальное членение предложений. – М.: Просвещение, 1976. – С. 76-78.
 [Кодзасов, 2009] Кодзасов С.В. Исследования в области русской просодии. – М.: Языки славянских культур, 2009. – 491 с.
 [Поспелов, 1989] Поспелов Д.А. Моделирование рассуждений. Опыт анализа мыслительных актов. – М.: Радио и связь, 1989. – 114 с.
 [Светозарова, 1993] Светозарова Н.Д. Акцентно-ритмические инновации в русской спонтанной речи // Проблемы фонетики. – 1993. – Т. 1. – С. 189-199.

- [Словарь рус. яз.] Словарь русского языка. Институт русского языка РАН. – М.: Русский язык, 1984. – С. 143.
- [Янко, 2001] Янко Т.Е. Коммуникативные стратегии русской речи. – М.: Языки славянских культур, 2001. – 384 с.
- [Attardo et al., 2013] Attardo S., Pickering L., Baker A. Prosodic and multimodal markers of humor in conversation. // *Prosody and Humor*. John Benjamins publishing company. Amsterdam, 2013. – Vol. 55. – P. 37-60.
- [Bryant, 2013] Bryant, G. A. Verbal irony in the wild // *Prosody and Humor*. John Benjamins publishing company. Amsterdam, 2013. – Vol. 55. – P. 103-120.
- [Couper-Kuhlen, 2021] Couper-Kuhlen E. Prosody. – 2021. – P. 1-24. – <https://www.benjamins.com/online/hop/articles/pro1/print>.
- [Oxford Handbook of Language Prosody, 2020] Oxford Handbook of Language Prosody. – Oxford University Press. New York, 2020. – P. 125-165.
- [Savinitch, 2019] Savinitch L.V. Pragmatics of Pauses // In: Proc. of the 1st International Seminar on the Foundations of Speech – Pausing, Breathing and Voice. University of Southern Denmark, Sønderborg, Denmark December 1–3, 2019. – P. 112-113.
- [Stefanuk et al., 2023] Stefanuk V.L., Savinitch L.V. Automated Discovery of Direct Speech to Increase Quality and Battery Saving in Telephony // In: Proc. of the Seventh International Scientific Conference “Intelligent Information Technologies for Industry” (ITI’23). – Springer Cham, 2023. – Vol. 2. – P. 169-178.
- [Verschuieren, 1999] Verschueren J. Understanding Pragmatics. – Oxford University Press. New York, 1999. – 25 p.

УДК 004.021

doi: 10.15622/rcai.2025.101

ДЕЦЕНТРАЛИЗОВАННАЯ МНОГОКРИТЕРИАЛЬНАЯ МАРШРУТИЗАЦИЯ

Н.П. Фомин (*heartmarshall@yandex.ru*)^A

К.С. Яковлев (*yakovlev@isa.ru*)^{A,B}

^A Санкт-Петербургский государственный университет,
Санкт-Петербург

^B Федеральный исследовательский центр
«Информатика и управление» РАН, Москва

Рассматривается задача децентрализованной многокритериальной маршрутизации в компьютерных сетях, где каждый узел самостоятельно формирует маршрутные решения на основе локально доступной информации. Современные протоколы маршрутизации зачастую ограничены алгоритмами поиска пути, минимизирующими значение скалярной целевой функции, что может приводить к неоптимальному выбору маршрута в сложных сетевых сценариях, когда имеется несколько равнозначных критериев качества. В работе предложены два новых децентрализованных метода маршрутизации, основанных на современных алгоритмах многокритериального поиска. Первый, метод моделирования поведения узлов, позволяет каждому узлу строить локальную таблицу маршрутизации с учетом узла-отправителя, гарантируя соблюдение заданных ограничений стоимости пути и детерминированность. Второй, жадный метод, принимает решение о следующем переходе, основываясь только на адресате пакета, предлагая компромисс между вычислительной сложностью и строгостью ограничений. Экспериментально подтверждена эффективность обоих подходов и представлены их сравнительные характеристики по качеству решений и ресурсоемкости.

Ключевые слова: многокритериальная маршрутизация, Парето-оптимальность, алгоритмы поиска пути, компьютерные сети.

Введение

Компьютерные сети являются важнейшей инфраструктурой современного мира, и от их производительности и стабильности зависит качество многих сфер жизнедеятельности человека. Ключевую роль в функциони-

ровании сетей играет алгоритм маршрутизации пакетов данных, который определяет пути их следования от отправителя к получателю. В большинстве сетей решения о маршрутизации принимаются децентрализованно: каждый узел (маршрутизатор) самостоятельно определяет куда отправить пакет на основе локально доступной ему информации и заложенного в него алгоритма. Выбор оптимального маршрута критически важен, так как каждый транзитный узел вносит задержку и потенциально увеличивает риск ошибок передачи данных.

Традиционные протоколы маршрутизации основаны на адаптации к децентрализованному сценарию работы таких известных (централизованных) алгоритмов поиска пути как, например, Дейкстры [Dijkstra, 1959] или Беллмана-Форда [Bellman, 1958]. Важной особенностью этих (и подобных им) алгоритмов является тот факт, что они являются однокритериальными. Такой подход, с одной стороны, упрощает реализацию, с другой – может приводить к выбору неоптимальных маршрутов на практике, особенно при наличии разнородных требований к качеству соединения, т.е. когда речь идет о необходимости учета нескольких критериев качества маршрута.

Подобный учет возможен в современных алгоритмах многокритериального поиска кратчайшего пути [Stewart et al., 1991], [Mandow et al., 2010], [Hernández Ulloa et al., 2020], [Hernández et al., 2023], способных находить множество Парето-оптимальных решений, т.е. осуществлять многокритериальную оптимизацию. Однако, все эти алгоритмы являются централизованными и, следовательно, напрямую неприменимыми в реальных задачах маршрутизации пакетов в компьютерных сетях, когда каждый узел принимает локальное решение о пересылке пакета на один из смежных узлов, обладая лишь локальной информацией. На устранение этого недостатка и направлена данная работа, целью которой является разработка и исследование новых методов децентрализованной многокритериальной маршрутизации, которые используют возможности современных централизованных алгоритмов поиска, в частности, алгоритма BOD [Hernández et al., 2023].

1. Постановка задачи

Сеть представляется в виде ориентированного графа $G = (S, E)$, где S – множество узлов, E – множество ребер. Каждому ребру $e \in E$ сопоставлен d -мерный вектор положительных стоимостей $\tilde{c}(e) = (c_1(e), \dots, c_d(e))$. Путь $\pi = (s_0, \dots, s_k)$ – последовательность вершин графа, между которыми существуют рёбра. Каждый путь имеет стоимость $\tilde{C}(\pi)$ равную сумме стоимостей рёбер между вершинами пути.

Вектор \tilde{p} доминирует вектор \tilde{q} (обозначается $\tilde{p} < \tilde{q}$), если $\forall j: p_j \leq q_j$ и $\exists i: p_i < q_i$. Путь π является Парето-оптимальным, если его стоимость $\tilde{C}(\pi)$ не доминируется стоимостью любого другого пути между теми же

начальным и конечным узлами. Задача многокритериального поиска кратчайшего пути заключается в нахождении множества всех Парето-оптимальных путей от s_{start} к s_{target} . В постановке “от одного ко многим” результатом для каждого узла s является Парето-оптимальное множество $sols(s)$, содержащее все недоминируемые векторы стоимостей путей из s_{start} в s .

В децентрализованной системе каждый узел $s \in S$ определяет следующий переход $nextHop_s(p) \in Succ(s)$ для пакета $p = (s_{\text{source}}, s_{\text{dest}})$, где $Succ(s)$ – множество смежных с s узлов. Совокупность функций $rules = \{nextHop_s \mid s \in S\}$ образует правила маршрутизации. Трассировка $trace(p, rules)$ определяет путь пакета в сети.

Критерии для оценки множества правил $rules$:

1. **Ограниченность:** если \exists путь π от s_{source} к s_{dest} с $\tilde{C}(\pi) \leq \tilde{C}_{\text{limit}} = (C_1, \dots, C_d)$, то $\tilde{C}(trace(p, rules)) \leq \tilde{C}_{\text{limit}}$.
2. **Детерминированность:** $trace(p, rules)$ всегда приводит к единственному пути.

Экземпляр задачи – пара $(G, \tilde{C}_{\text{limit}})$. Рассматриваются две постановки для $nextHop_s(p)$: зависимость этой функции от пары $(s_{\text{source}}, s_{\text{dest}})$ или только от s_{dest} .

2. Анализ существующих подходов

Задача многокритериальной маршрутизации активно исследуется в контексте обеспечения качества соединения в компьютерных сетях. Существующие подходы можно условно разделить на несколько категорий.

Наиболее распространенным методом является агрегация нескольких метрик в единую скалярную функцию стоимости, как правило, путем их взвешенной суммы. На основе этой объединенной метрики затем выполняется алгоритм поиска кратчайшего пути, например, Дейкстры. Хотя такие протоколы, как OSPF, являются децентрализованными, сам подход со взвешенной суммой имеет фундаментальные недостатки, рассмотренные в обзорных работах по QoS-маршрутизации [Wang & Crowcroft, 1996]. Выбор весовых коэффициентов нетривиален и, по сути, является глобальным решением, плохо адаптирующимся к динамическим условиям сети. Кроме того, любой фиксированный набор весов приводит к потере информации о Парето-оптимальных маршрутах.

Другой класс подходов формулирует задачу как поиск кратчайшего пути с ограничениями. В этом случае одна метрика минимизируется при условии, что остальные не превышают заданных порогов. Эта задача в общем случае является NP-трудной [Jaffe, 1984]. Из-за этого исследования в данной области часто фокусируются на разработке эвристических алгоритмов, которые находят приближенные решения за приемлемое время [Korkmaz & Krungz, 2001]. Однако эвристики не гарантируют нахождения оптимального пути, а их децентрализованная реализация остается сложной задачей.

Именно эту проблему решает предложенный в работе Метод Моделирования. Он является полностью децентрализованным, но, в отличие от стандартных эвристик, путем моделирования принятия решений другими узлами обеспечивает строгие гарантии соблюдения сквозных ограничений. Жадный метод, в свою очередь, представляет собой практический децентрализованный компромисс: он не предоставляет полных гарантий, но, в отличие от простой агрегации метрик, учитывает Парето-оптимальность на локальном уровне. Таким образом, данное исследование предлагает новые механизмы именно для децентрализованной многокритериальной маршрутизации, анализируя компромисс между полнотой гарантий соблюдения ограничений и ресурсоемкостью.

2. Предлагаемые методы

Для решения поставленной задачи децентрализованной многокритериальной маршрутизации предлагаются два метода, использующие алгоритм BOD для нахождения Парето-оптимальных множеств путей $sols(s)$. Методы различаются объемом используемой информации для формирования правил $nextHop_s(p)$ и, как следствие, обеспечиваемыми гарантиями ограниченности и детерминированности в обмен на вычислительную сложность.

При этом стоит отметить невозможность нахождения детерминированного ограниченного метода при зависимости $nextHop_s(p)$ только от s_{dest} (т.е., без учета s_{source}). Доказательством служит следующий сценарий, иллюстрируемый на рис. 1.

Рассмотрим узел s_{cur} , который получил два пакета, оба адресованные узлу s_{goal} , но пришедшие по разным предшествующим путям π и π' (из разных s_{source} , как показано на рис. 1,б). Эти пути имеют различные накопленные стоимости $c(\pi)$ и $c(\pi')$ (рис. 1а). Если s_{cur} принимает решение о следующем переходе, основываясь только на адресате s_{goal} , он должен выбрать один и тот же следующий узел (например, с ребром стоимости (g_1, g_2) или с ребром (g'_1, g'_2) на рис. 1,а) для обоих пакетов. Однако, как видно из рис. 1,а, выбор одного ребра может привести к тому, что суммарная стоимость пути для одного пакета $(c(\pi) + (g_1, g_2))$ уложится в заданные ограничения C_{limit} , в то время как для другого пакета $(c(\pi') + (g_1, g_2))$ эти ограничения будут превышены. Аналогичная ситуация может возникнуть при выборе другого ребра. Таким образом, не существует единого детерминированного выбора следующего перехода для s_{cur} , зависящего только от s_{dest} , который гарантировал бы соблюдение ограничений C_{limit} для обоих пакетов несмотря на то, что для каждого из них по отдельности может существовать допустимый маршрут при учете s_{source} .

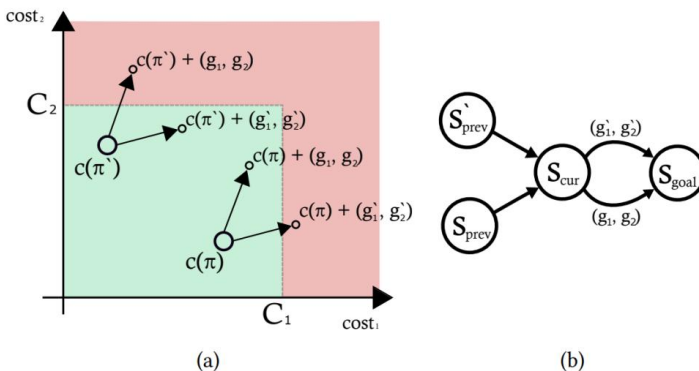


Рис. 1. (a) Пример возможных исходов при продолжении некоторых путей π и π' из вершины s_{cur} и (b) рисунок подграфа, в котором содержится эта вершина

Метод моделирования поведения узлов (Modeling Method)

Данный метод нацелен на построение полного множества правил маршрутизации *rules*, удовлетворяющего критериям ограниченности и детерминированности. Каждый узел cur_node формирует свою локальную таблицу маршрутизации $nextHop_{cur_node(p)}$, учитывая как узел-отправитель s_{source} , так и узел-адресат s_{dest} пакета p . Это достигается путем моделирования принятия решений другими узлами для предсказания полных путей и их соответствия ограничениям \tilde{C}_{limit} .

Процесс построения таблицы маршрутизации для cur_node включает:

1. Определение множества локально достижимых из cur_node узлов (далее R_{loc}), пути к которым удовлетворяют \tilde{C}_{limit} и запись начальных правил для них.
2. Определение множества потенциальных узлов-отправителей S_{pot} , для которых cur_node достижим в пределах \tilde{C}_{limit} . Для этого используется BOD на графе с инвертированными ребрами.
3. Для каждой пары $(s_i \in S_{pot}, s_j \in R_{loc})$ моделируется поиск пути из s_i в s_j . Если оптимальный путь проходит через cur_node , то определяется следующий узел после cur_node и соответствующее правило $(s_i, s_j, next_hop)$ добавляется в таблицу cur_node .
4. Опциональная оптимизация таблицы путем замены набора правил $(s_k, t, next_hop_common)$ на одно правило $(*, t, next_hop_common)$, если $next_hop_common$ одинаков для всех s_k .

Псевдокод метода представлен в листинге 1.

Листинг 1

Алгоритм моделирования поведения узлов

```

Input: A single routing problem ( $S, E, \vec{c}, C_1, C_2, s_{cur\_node}$ )
Output: route_table
1   $sols \leftarrow BOD^1(S, E, \vec{c}, s_{cur\_node})$ 
2  for each  $t \in S$  do
3      if  $sols(t) \neq \emptyset$  then
4           $route\_table.write(s_{cur\_node}, t, first\_step(choose(sols(t))))$ 
5           $reachable \leftarrow t$ 
6   $sols \leftarrow BOD^2(S, reverse(E), \vec{c}, s_{cur\_node})$ 
7  for each  $t \in S$  do
8      if  $sols(t) \neq \emptyset$  then
9           $possible\_senders \leftarrow p$ 
10 for each  $t \in possible\_senders$  do
11      $sols \leftarrow BOD^3(S, E, \vec{c}, t)$ 
12     for each  $p \in reachable$  do
13         if  $next\_node(sols(p)) \neq \emptyset$  then
14              $route\_table.write(t, p, next\_node(choose(sols(p))))$ 
15 for each  $target \in reachable$  do
16      $senders_{target} \leftarrow \emptyset$ 
17     for each  $sender \in possible\_senders$  do
18         Add  $route\_table.get\_next\_hop(sender, target)$  to  $senders_{target}$ 
19     if  $|senders_{target}| \leq 1$  then
20         // Заменяем все записи, содержащие  $target$  в поле адресата пакета на
           строку с универсальным переходом ( $target, *, next\_hop$ )
21          $route\_table.prune(target)$ 
22 return  $route\_table$ 

```

Метод жадной маршрутизации (Greedy Method)

Жадный метод предлагает упрощенный подход, где узел cur_node формирует правила $nextHop_{cur_node}(p)$, основываясь только на адресате $sdest$ и Парето-оптимальных путях из cur_node . Метод не моделирует поведение других узлов.

Алгоритм состоит из одного запуска BOD из cur_node для определения $sols(t)$ для всех достижимых t . Правило $nextHop_{cur_node}(*, t)$ формируется применением функции $choose(sols(t))$ для выбора первого шага пути. Псевдокод представлен в листинге 2.

Алгоритм жадной маршрутизации

Input: A single routing problem $(S, E, \vec{c}, C_1, C_2, s_{start})$
Output: *route_table*

```

1  $sols \leftarrow BOD^1(S, E, \vec{c}, s_{start})$ 
2 for each  $t \in S$  do
3   if  $sols(t) \neq \emptyset$  then
4      $route\_table.write(s_{start}, t, first\_step(choose(sols(t))))$ 
5 return route_table

```

Исследованы три варианта функции $choose(sols(s))$ для двухкритериального случая ($d=2$):

1. **Ближайший к «идеальной» точке:** среди решений, удовлетворяющих \tilde{C}_{limit} (или всех, если таких нет), выбирается минимизирующее сначала евклидово расстояние до $(0,0)$, затем до прямой $g_1=g_2$. См. рисунок №2.
2. **Мин. стоимость по критерию 1:** выбор решения с минимальной первой компонентой стоимости g_1 (с приоритетом для удовлетворяющих \tilde{C}_{limit}).
3. **Мин. стоимость по критерию 2:** аналогично варианту 2, но для g_2 .

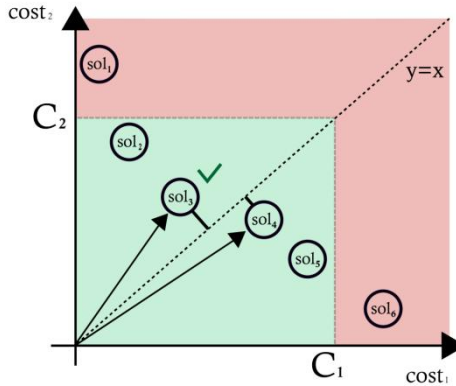


Рис. 2. Демонстрация работы варианта №1 функции $choose$ на Парето-множестве из 6 решений: выбрано решение sol_3

Жадный метод уступает методу моделирования в гарантиях ограниченности и детерминированности (последнее зависит от реализации функции $choose$), но значительно выигрывает в скорости выполнения.

3. Экспериментальные исследования

Для оценки эффективности и сравнения предложенных методов моделирования и жадной маршрутизации были проведены экспериментальные исследования.

3.1. Условия эксперимента

Эксперименты проводились на синтетически сгенерированном направленном графе, моделирующем компьютерную сеть и состоящем из 1024 узлов. Средняя степень исхода вершины составляла четыре. Каждое ребро графа характеризовалось двумя критериями ($d=2$): задержкой (delay) и вероятностью потери пакета (loss), значения которых генерировались случайным образом в диапазонах $[0.00497, 0.02953]$ и $[0.00869, 0.04347]$ соответственно.

В качестве ограничений на суммарную стоимость пути $\tilde{C}_{limit} = (C_1, C_2)$ использовались три пары значений: $(0.6, 0.6)$, $(0.7, 0.7)$ и $(0.8, 0.8)$.

Эксперимент состоял из двух основных этапов:

1. Построение таблиц маршрутизации: для каждого узла сети и для каждого из исследуемых методов (Modeling и трех вариантов Greedy, различающихся функцией *choose*) генерировалось полное множество правил маршрутизации. Для Modeling функция *choose* выбирала решение с минимальной первой компонентой стоимости среди удовлетворяющих \tilde{C}_{limit} (или из всех, если таковых нет).
2. Трассировка пакетов: генерировался набор из 10000 тестовых пакетов – случайных пар отправитель-получатель. Для каждого пакета и для каждой построенной таблицы маршрутизации выполнялась трассировка для определения итогового пути и его стоимости.

Также для оценки ресурсоемкости Modeling метода было проведено сравнение с его версией, где оптимизации алгоритма BOD (используемые на шагах 1-3 метода, такие как ограничение глубины поиска по \tilde{C}_{limit} или сохранение только первого/следующего шага вместо полных предков) были отключены (далее Modeling-no-opt).

Все эксперименты проводились на персональном компьютере с процессором AMD Ryzen 5 7600X и 64 Гб оперативной памяти.

3.2. Показатели качества

Для сравнения методов использовались следующие метрики:

1. Процент недоставленных пакетов: доля пакетов, для которых в результате трассировки либо не был найден путь до адресата, либо стоимость найденного пути превысила ограничения \tilde{C}_{limit} . Эта метрика напрямую отражает соблюдение критерия ограниченности.
2. Среднее время построения таблицы маршрутизации для одного узла. Характеризует вычислительную сложность каждого метода на этапе подготовки.

3. Максимальное количество состояний в очереди OPEN алгоритма BOD: используется для оценки пикового потребления памяти алгоритмом BOD в ходе работы методов.

3.3. Результаты и их анализ

Основные результаты экспериментов представлены в табл. 1–3.

Таблица 1

Процент пакетов, для которых не найден путь, стоимость которого не превышает ограничения

Algorithm	Experiments		
	C=(0.6, 0.6)	C=(0.7, 0.7)	C=(0.8, 0.8)
	bad pockets (%)	bad pockets (%)	bad pockets (%)
Greedy	18	14	9
	14	11	6
	15	10	7
Modeling	0	0	0

Как видно из табл. 1, Modeling метод успешно доставил 100% пакетов в рамках заданных ограничений для всех исследованных \tilde{C}_{limit} , подтверждая свою способность обеспечивать ограниченность. Методы Greedy показали наличие недоставленных пакетов, причем их доля уменьшалась с ростом значений \tilde{C}_{limit} . Среди вариантов Greedy метода, наилучшие результаты показали варианты, выбирающие решение по минимальной стоимости одного из критериев: Greedy-2 и Greedy-3. Greedy-1 ("ближайший к идеальной точке") оказался наименее эффективным по данной метрике.

Таблица 2

Среднее время выполнения алгоритма

Algorithm	Experiments		
	C=(0.6, 0.6)	C=(0.7, 0.7)	C=(0.8, 0.8)
	avg time (s.)	avg time (s.)	avg time (s.)
Greedy	0.3	0.42	0.51
Modeling	95	186	305

Табл. 2 демонстрирует значительное различие в вычислительной сложности. Greedy требует на порядки меньше времени для построения таблиц, так как выполняет лишь один проход BOD для каждого узла. Modeling метод, включающий многократные запуски BOD для моделирования, существенно более ресурсоемок.

Таблица 3

Сравнение потребления ресурсов у методов моделирования
поведения узлов

\vec{C}	Algorithm	average time(s)	max OPENlen	solutions count
(0.6, 0.6)	Modeling	95	3406	7280
	Modeling_no_opt	81	3406	8032
(0.7, 0.7)	Modeling	186	3630	8290
	Modeling_no_opt	200	3630	8600
(0.8, 0.8)	Modeling	305	3910	9028
	Modeling_no_opt	285	3910	9240

Результаты, представленные в табл. 3, показывают эффективность оптимизаций, примененных в алгоритме BOD в рамках Modeling метода. Версия Modeling с оптимизациями требует несколько меньшего среднего времени и, что более важно, работает с меньшим или таким же пиковым размером очереди OPEN и количеством генерируемых решений, что подтверждает целесообразность внесенных модификаций в BOD для снижения ресурсоемкости без ущерба для качества основного метода.

В целом, экспериментальные результаты подтверждают теоретические предположки: Modeling обеспечивает более строгие гарантии ценой вычислительной сложности, тогда как Greedy предлагает более быстрый, но менее надежный в плане соблюдения ограничений подход. Выбор между методами зависит от конкретных требований к системе маршрутизации.

Заключение

В работе предложены новые методы децентрализованной многокритериальной маршрутизации, основанные на адаптации методов централизованного эвристического поиска. Проведены их экспериментальных исследования.

Проведенное исследование имеет ряд ограничений, которые определяют направления для дальнейшей работы. Так, экспериментальная оценка была проведена на графах с фиксированными параметрами. Для подтверждения обобщаемости полученных результатов возможно проведение исследований с расширенными наборами топологий графов. Также стоит провести анализ влияния ключевых топологических на производительность предложенных методов.

Другим важным направлением для будущих исследований является анализ эффективности алгоритмов в динамических сетевых условиях, где метрики каналов связи могут изменяться во времени. Такое исследование позволит оценить не только устойчивость найденных решений, но и накладные расходы, связанные с обновлением таблиц маршрутизации.

Список литературы

- [Bellman, 1958] Bellman R. On a routing problem // Quart. Appl. Math. – 1958. – Vol. 16. – P. 87-90. – doi: 10.1090/qam/102435.
- [Dijkstra, 1959] Dijkstra E.W. A note on two problems in connexion with graphs // Numerische Mathematik. – 1959. – Vol. 1, No. 1. – P. 269-271.
- [Hernández, 2023] Hernández C., Yeoh W., Baier J.A., Zhang H., Suazo L., Koenig S., Salzman O. Simple and efficient bi-objective search algorithms via fast dominance checks // Artificial Intelligence. – 2023. – Vol. 314. – P. 103807. – doi: 10.1016/j.artint.2022.103807.
- [Hernández Ulloa, 2020] Hernández Ulloa C., Yeoh W., Baier J.A., Zhang H., Suazo L., Koenig S. A Simple and Fast Bi-Objective Search Algorithm // ICAPS. – 2020. – P. 143-151. – doi: 10.1609/icaps.v30i1.6655.
- [Jaffe, 1984] Jaffe J.M. Algorithms for finding a collection of circuit-disjoint paths with certain constraints // Networks. – 1984. – Vol. 14, No. 1. – P. 95-114. – doi: 10.1002/net.3230140108.
- [Korkmaz & Krunz, 2001] Korkmaz T., Krunz M. A randomized algorithm for finding a path subject to multiple QoS constraints // Computer Networks. – 2001. – Vol. 36. No. 2-3. – P. 251-268. – doi: 10.1016/S1389-1286(00)00209-7.
- [Mandow, 2010] Mandow L., De La Cruz J.L.P. Multiobjective A* search with consistent heuristics // J. ACM. – 2010. – Vol. 57, No. 5. – P. 27:1-27:25. – doi: 10.1145/1754399.1754404.
- [Stewart, 1991] Stewart B.S., White III C.C. Multiobjective A* // J. ACM. – 1991. – Vol. 38, No. 4. – P. 775-814. – doi: 10.1145/115783.115786.
- [Wang & Crowcroft, 1996] Wang Z., Crowcroft J. Quality-of-service routing for supporting multimedia applications // IEEE Journal on selected areas in communications. – 1996. – Vol. 14, No. 7. – P. 1228-1234. – doi: 10.1109/49.536364.

УДК 519.841

doi: 10.15622/rcai.2025.102

ИНТЕРПРЕТИРУЕМЫЕ МОДЕЛИ-ЗАМЕСТИТЕЛИ С TREESHAP-АНАЛИЗОМ ДЛЯ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ОБЛАСТИ ПОЖАРНОЙ БЕЗОПАСНОСТИ

Р.Ш. Хабибулин (*kh-r@yandex.ru*)

Академия ГПС МЧС России, Москва

В данной работе представлен формализованный подход к созданию и валидации моделей-заместителей машинного обучения для аппроксимации байесовских вычислений в задачах ранжирования комплексов противопожарных мероприятий. Предложенный подход основан на использовании древовидных алгоритмов (*Random Forest*, *XGBoost*, *LightGBM*) в сочетании с методом *TreeSHAP* для обеспечения интерпретируемости управленческих решений. На примере объектов хранения нефти и нефтепродуктов проведено экспериментальное исследование на синтетическом датасете из 2000 сценариев. Разработано программное обеспечение, включающее в себя комплексную систему валидации с метриками корреляции ранжирования и анализом влияния на принятие решений.

Ключевые слова: *TreeSHAP*, пожарная безопасность, системы поддержки принятия решений, интерпретируемое машинное обучение.

Введение

Принятие решений по выбору противопожарных мероприятий на различных объектах защиты представляет собой комплексную многокритериальную задачу, осложненную противоречиями между требованиями к точности, скорости, стоимости и объяснимости управленческих решений. Лица, принимающие решения (ЛПР), должны ранжировать альтернативные комплексы противопожарных мероприятий по многим критериям, включая такие важные как стоимость, время реализации и эффективность снижения пожарного риска в условиях возможной неопределенности исходных данных и субъективности экспертных предпочтений [Бурков, 2010]. Существующие широко известные методы многокритериального анализа (*TOPSIS* (ранжирование альтернатив на основе их близости к иде-

альному решению), *ELECTRE* (оценивает отношение превосходства одной альтернативы над другой), Метод анализа иерархий) в целом не предназначены для оценки интервальной неопределенности исходных данных и вариативности весов критериев, что может привести к принятию субоптимальных решений [Андрейчук, 2022]. Байесовские методы стохастического доминирования, хотя и обеспечивают математически строгий учет неопределенности через распределения Дирихле и итерации Монте-Карло, требуют тысяч вычислений для получения одного ранжирования комплекса противопожарных мероприятий, что делает их трудно применимыми в интерактивных системах поддержки принятия решений (СППР) [Агасиев и др., 2024]. Современные алгоритмы машинного обучения способны обеспечить требуемую скорость вычислений, однако их непрозрачность ("черный ящик") не удовлетворяет существующим требованиям к обоснованности принимаемых решений в области обеспечения пожарной безопасности [Кузнецова и др., 2018].

Таким образом, ключевая проблема заключается в отсутствии формализованного подхода, который одновременно обеспечивал бы математическую строгость байесовского подхода, вычислительную эффективность машинного обучения и полную интерпретируемость результатов для ЛПР. Таким образом, цель данной работы состоит в разработке и валидации моделей-заместителей (суррогатных моделей) машинного обучения с *SHAP*-интерпретацией (метод интерпретации моделей машинного обучения, основанный на теории значений Шепли из кооперативной теории игр) [Mitchell et. al., 2022], способных аппроксимировать байесовские вычисления стохастического доминирования при сохранении точности ранжирования и обеспечении объяснений каждого управленческого решения. В работе применяется метод *TreeSHAP* (оптимизированный алгоритм вычисления значений *SHAP* для ансамблей деревьев) для объяснения решений в задачах управления противопожарными мероприятиями в рамках риск-ориентированного подхода и разработке комплексной системы валидации моделей-заместителей с учетом специфики ранжирования альтернатив.

1. Краткий обзор подходов и методов

Модели-заместители как метод аппроксимации вычислительно затратных процессов получили широкое распространение во многих предметных областях: от замены *CFD*-симуляций в аэродинамике до аппроксимации конечноэлементного анализа в конструкторском проектировании [Хведчук и др., 2018]. Однако применение таких моделей в задачах управления пожарной безопасностью остается малоизученным направлением, что объясняется специфическими требованиями к надежности и интерпретируемости решений в критически важных системах обеспечения безопасности. Параллельно развитие объяснимого искусственного

интеллекта привело к созданию универсальных методов интерпретации машинного обучения, таких как локальные линейные аппроксимации *LIME* (метод локальной интерпретации, не зависящий от типа модели) [Волков и др., 2023], теоретико-игровой подход *SHAP* на основе значений Шепли [Воробьев, 2021], и его реализация *TreeSHAP* для древовидных моделей [Deb et. al., 2021]. *TreeSHAP* обеспечивает точные объяснения за полиномиальное время $O(TLD)$, где T - количество деревьев, L - листьев, D - глубина, что критично для интерактивных систем поддержки принятия решений.

Анализ классов алгоритмов машинного обучения для рассматриваемых задач моделирования в контексте поддержки управления пожарной безопасностью выявил различия в их применимости. Древовидные алгоритмы (*Random Forest*, *XGBoost*, *LightGBM*) демонстрируют оптимальное сочетание характеристик: нативную поддержку метода *TreeSHAP* без дополнительных вычислительных затрат, обработку нелинейных взаимодействий между интервальными признаками, достаточно высокую робастность к выбросам в данных экспертных оценок, быстрое действие как при обучении, так и при прогнозировании, и практическую интерпретируемость результатов. Нейронные сети, несмотря на высокую точность аппроксимации сложных нелинейных зависимостей, чувствительны к гиперпараметрам, имеют склонность к переобучению. Линейные модели, хотя и обладают высокой скоростью и естественной интерпретируемостью, не позволяют адекватно моделировать сложные взаимодействия между интервальными границами критериев и байесовскими весами, что приводит к существенной потере точности аппроксимации. Методы опорных векторов (*SVM*) демонстрируют хорошую обработку нелинейности и робастность, однако требуют применения модель-независимой версии *Kernel SHAP* для интерпретации, что увеличивает вычислительную сложность. Таким образом, древовидные алгоритмы представляют собой оптимальный выбор для моделирования в задачах ранжирования комплексов противопожарных мероприятий, обеспечивая необходимое сочетание точности, скорости и объяснимости решений.

2. Методология исследования

Исходная задача формализуется как байесовская модель стохастического доминирования для N альтернативных комплексов противопожарных мероприятий, где каждый комплекс i характеризуется интервальными оценками критериев $[a_{ik}, b_{ik}] \subset \mathbb{R}^+$ для $k = 1, 2, 3$ (стоимость выполнения противопожарных мероприятий, время выполнения противопожарных мероприятий, дефицит пожарного риска на объекте защиты соответственно) и априорными параметрами распределения Дирихле $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ для весов критериев. Полная байесовская модель выполняет M итераций ме-

тодом Монте-Карло, генерируя на каждом шаге m случайные значения критериев $X_{ik}^m \sim U(a_{ik}, b_{ik})$ и веса $W^m \sim \text{Dirichlet}(\alpha)$, далее вычисляется нормированная взвешенная полезность каждой меры:

$$U_i^{(m)} = \sum_{k=1}^3 W_k^{(m)} \cdot \tilde{X}_{ik}^{(m)}, \quad (1)$$

где $\tilde{X}_{ik}^{(m)} = \frac{X_{ik}^{(m)} - \min_j X_{jk}^{(m)}}{\max_j X_{jk}^{(m)} - \min_j X_{jk}^{(m)}}$ – нормированные значения критериев,

$W_k^{(m)}$ – k -й компонент весового вектора в итерации m , $U_i^{(m)}$ – полезность меры i в итерации m . Вероятность стохастического доминирования меры i над мерой j оценивается как:

$$P_{ij} = \Pr(U_i \leq U_j) \approx (1/M) \sum_{m=1}^M \mathbb{1}[U_i^{(m)} \leq U_j^{(m)}], \quad (2)$$

где $\mathbb{1}[\cdot]$ – индикаторная функция, M – количество итераций Монте-Карло, P_{ij} – вероятность доминирования. Интегральный балл меры для ранжирования вычисляется как $s_i = \sum_{j \neq i} P_{ij}$, однако вычислительная сложность $O(M \cdot N^2)$ делает метод затруднительным для интерактивных систем поддержки принятия решений.

Модель-заместитель аппроксимирует байесовские вычисления через регрессионное отображение $f_\theta: \mathbb{R}^d \rightarrow \mathbb{R}$, где вектор признаков $z_i \in \mathbb{R}^d$ формируется объединением интервальных границ и ожидаемых весов:

$$z_i = [a_{i1}, b_{i1}, a_{i2}, b_{i2}, a_{i3}, b_{i3}, \mathbb{E}[W^1], \mathbb{E}[W^2], \mathbb{E}[W^3]]^T, \quad (3)$$

где a_{ik}, b_{ik} – нижняя и верхняя границы k -го критерия для комплекса противопожарных мероприятий i , $\mathbb{E}[W_k] = \alpha_k / \sum_j \alpha_j$ – ожидаемый вес k -го критерия при распределении Дирихле, $d = 9$ – размерность пространства признаков. Аппроксимация имеет вид:

$$\hat{s}_i = f_\theta(z_i), \quad (4)$$

где \hat{s}_i – предсказанный интегральный балл комплекса противопожарных мероприятий i , θ – параметры машинной модели, оптимизируемые минимизацией функции потерь $L(\theta) = \sum_i (s_i - \hat{s}_i)^2$ между истинными и предсказанными баллами.

Процесс обучения включает генерацию сценариев с различными комбинациями интервалов и весов, запуск полной байесовской модели для получения истинных баллов, и обучение моделей-заместителей различных типов (*Random Forest*, *XGBoost*, *LightGBM*) с последующей валидацией. Система валидации учитывает специфику задач ранжирования через метрики корреляции рангов, где коэффициент корреляции Спирмена ρ_s оценивает сохранение порядка альтернатив:

$$\rho_s = 1 - (6 \sum_i d_i^2) / (N(N^2 - 1)), \quad (5)$$

где $d_i = rank(s_i) - rank(\hat{s}_i)$ – разность рангов комплекса противопожарных мероприятий i в истинном и предсказанном ранжированиях, N – количество комплексов противопожарных мероприятий. Дополнительно вычисляется точность совпадения топ- K рекомендаций как $Acc_K = |Top_K(s) \cap Top_K(\hat{s})|/K$, где $Top_K(\cdot)$ возвращает множество индексов K лучших комплексов противопожарных мероприятий.

Интерпретация решений модели-заместителя обеспечивается методом *TreeSHAP*, который для древовидных алгоритмов вычисляет значения Шепли $\phi_i p$ для признака p комплекса противопожарных мероприятий i согласно формуле:

$$\phi_i p = \sum_{S \subseteq Fp} (|S|! (|F| - |S| - 1)! / (|F|!)) [f_\theta(z_{i,S \cup p}) - f_\theta(z_{i,S})], \quad (6)$$

где F – множество всех признаков, S – подмножество признаков без p , $z_{i,S}$ – вектор признаков комплекса противопожарных мероприятий i с нулевыми значениями для признаков не из S , $\phi_i p$ – вклад признака p в предсказание для комплекса противопожарных мероприятий i . Аддитивность *SHAP* гарантирует равенство $\sum_p \phi_i p + \phi^0 = \hat{s}_i$, где ϕ^0 – базовое значение модели, что обеспечивает полное объяснение каждого управленческого решения через распределение ответственности между влияющими факторами.

Основные этапы алгоритма для разработанной модели:

1. Ввод параметров объекта: тип объекта защиты

2. Установка ограничений (интервалы): максимальный бюджет, максимальное время реализации, минимальное требуемое снижение риска

3. Формирование альтернатив

Автоматическая генерация множества из 6 типовых комплексов противопожарных мероприятий

Загрузка характеристик из базы данных противопожарных мероприятий

Каждый комплекс мероприятий характеризуется интервальными оценками

4. Настройка критериев

Установка весов критериев (стоимость, время, риск)

Генерация параметров распределения Дирихле для учета неопределенности предпочтений

5. Выбор вычислительного метода

Выбор из 2 режимов:

Точный режим:

Байесовская модель стохастического доминирования

Время выполнения: ~52 мс

Максимальная точность

Быстрый режим:

Модель-заместитель XGBoost

Время выполнения: ~2.8 мс

Точность $R^2 = 0,9687$

6. Вычисление и ранжирование

Расчет интегральных баллов

Сортировка комплексов противопожарных мероприятий по убыванию баллов

Определение 3 лучших рекомендуемых решений

7. Интерпретация решений

TreeSHAP анализ для модели-заместителя

Вычисление SHAP значений для каждого фактора

Определение вклада каждого фактора в итоговое решение

8. Формирование результатов

Отчет включает:

Ранжированный список комплексов противопожарных мероприятий с баллами

3 лучших рекомендуемых решений с обоснованием

Визуализация (вортерфолл диаграммы, радарные графики)

Детальные объяснения через SHAP значения

9. Сохранение и итерация

Оptionальное сохранение результатов в базу данных решений (база знаний)

Возможность проведения нового анализа с измененными параметрами

Накопление истории решений для дальнейшего анализа

3. Разработка программного обеспечения и проведение компьютерного моделирования

Экспериментальная валидация моделей-заместителей проводилась на синтетическом датасете из 2000 сценариев, где каждый сценарий включал $N = 15$ альтернативных комплексов противопожарных мероприятий с интервальными оценками критериев и случайными параметрами распределения Дирихле $\alpha \in [0,5, 5,0]$. Полная байесовская модель выполняла $M = 5000$ итераций Монте-Карло для каждого сценария.

Для проведения экспериментов было разработано специализированное программное обеспечение на языке *Python* 3.9 с использованием библиотек: *NumPy* и *SciPy* для численных вычислений и статистических операций, *Pandas* для обработки табличных данных, *Scikit-learn* для базовых алгоритмов машинного обучения и метрик валидации, *XGBoost*, *LightGBM* для градиентного бустинга, *SHAP* для интерпретации моделей, *Matplotlib* и *Seaborn* для визуализации результатов (рис. 1).

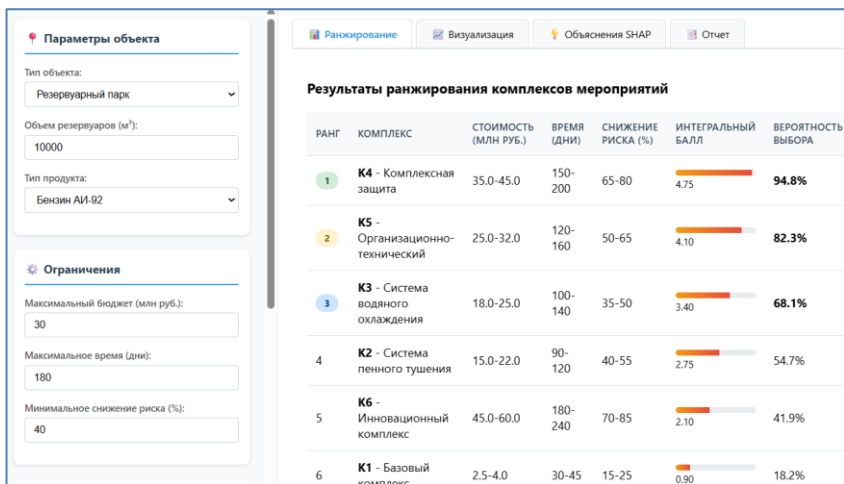


Рис. 1. Интерфейс программного обеспечения

Данные разделялись в пропорции 70%/15%/15% для обучения, валидации и тестирования соответственно. Обучались четыре типа моделей-заместителей: *XGBoost*, *Random Forest*, *LightGBM* и *Gradient Boosting* с последующей комплексной оценкой по метрикам регрессии, корреляции ранжирований и влияния на принятие решений.

XGBoost продемонстрировал превосходство по большинству метрик, достигнув коэффициента детерминации $R^2 = 0,967$ при минимальной частоте критических ошибок (0,3%), где худшая мера ошибочно выбиралась как лучшая. Корреляция Спирмена $\rho = 0,984$ указывает на высокое сохранение порядка ранжирования, что является ключевым требованием для задач принятия решений. Точность совпадения трех лучших рекомендаций составила 89%, что означает правильную идентификацию приоритетных мер в 9 случаях из 10.

TreeSHAP анализ 1000 объяснений (рис. 2) показал, что решения ЛПП в первую очередь определяются максимальным снижением риска (31,2% влияния) и минимальной стоимостью (24,7%), что соответствует принципу осторожного планирования в условиях неопределенности.

Ожидаемые веса критериев составляют 18,3% влияния, подтверждая важность учета субъективных предпочтений экспертов в рассматриваемой модели.

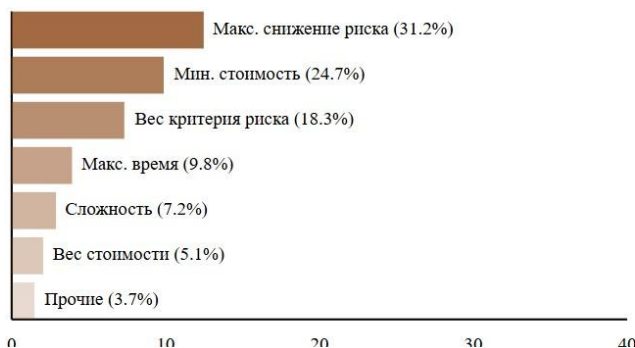


Рис. 2. *SHAP*-анализ ключевых факторов принятия решений

Локальные объяснения для конкретных комплексов противопожарных мероприятий (рис. 3) демонстрируют аддитивность *SHAP*-значений: например, комплекс противопожарных мероприятий №4 получает балл 2,0 как сумму базового значения (2,0) и вкладов отдельных факторов (-0,9 за экономичность, +1,5 за эффективность, -0,6 за длительность реализации), что обеспечивает полную прозрачность каждого решения для ЛПР.

Достигнутая точность ($R^2 = 0,9687$) превышает пороговое значение 0,95, рекомендуемое для критически важных приложений. Доверительный интервал [0,965, 0,972] указывает на статистическую устойчивость результата.

TreeSHAP обеспечивает полную декомпозицию каждого решения на вклады отдельных факторов. Доминирование фактора "максимальное снижение риска" (31,2%) соответствует принципу приоритета безопасности в управлении пожароопасными объектами.

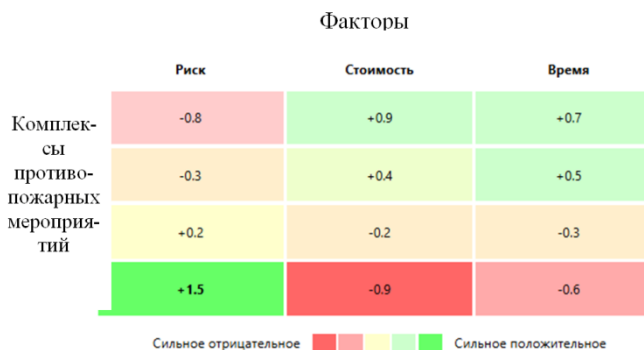


Рис. 3. Тепловая карта влияния факторов принятия решений

На основе модельных результатов разработаны практические рекомендации по выбору архитектуры моделей-заместителей: для критически важных решений в области пожарной безопасности рекомендуется модель *XGBoost* с параметрами и регуляризацией, обеспечивающими баланс точности и защиты от переобучения, тогда как для быстрого прототипирования подходит модель *Random Forest* со 100 деревьями. Установлены критерии приемлемости моделей-заместителей: коэффициент детерминации (R^2) $\geq 0,90$ для объяснения вариации, коэффициент корреляции Спирмена $\rho \geq 0,95$ для корреляции рангов, точность по тройке лучших (*Top-3 Accuracy*) $\geq 0,80$ для совпадения ключевых рекомендаций, частота критических ошибок (*Critical Error Rate*) $\leq 1\%$ для исключения катастрофических ошибок и среднюю абсолютную процентную ошибку (*MAPE*) $\leq 15\%$ для приемлемой точности.

Предложенный формализованный подход решает поставленную задачу создания быстрых, точных и интерпретируемых моделей для поддержки принятия решений в области пожарной безопасности. Достигнутое ускорение вычислений в 18–20 раз при сохранении точности выше 94% для практически значимых метрик делает подход применимым для создания интерактивных СППР, способных работать в режиме реального времени.

4. Заключение и выводы по результатам работы

Результаты проведенной работы направлены на решение противоречия между требованиями к математической строгости, вычислительной эффективности и интерпретируемости в задачах принятия решений по выбору комплекса мероприятий пожарной безопасности (на примере объектов хранения нефти и нефтепродуктов).

Экспериментальная валидация на 2000 сценариях подтверждает стабильность и робастность предложенного подхода с вариацией точности менее 1% при кросс-валидации, что обеспечивает надежность применения в реальных условиях рассматриваемых объектов защиты.

Практическая значимость исследования подтверждается созданием программного прототипа, демонстрирующего возможность интеграции практико-ориентированных технологий принятия решений в критически важные процессы управления пожарной безопасностью.

Перспективным направлением развития проводимого исследования являются методологические расширения: разработка мультифидельностных подходов с иерархией моделей различной точности (быстрый *Random Forest* для первоначальной оценки \rightarrow точный *XGBoost* для финального ранжирования \rightarrow полная байесовская модель для решений с наибольшей точностью). Дальнейшие исследования будут направлены на валидацию на реальных данных, рассмотрение комплексов противопожарных мероприятий на другие объекты защиты и интеграцию с системами автоматизированными СППР.

Список литературы

- [Агасиев и др., 2024] Агасиев Т.А., Карпенко А.П. Байесовская оптимизация с прогнозированием наилучших значений гиперпараметров суррогатной модели // Системы компьютерной математики и их приложения. – 2024. – № 25. – С. 62-67.
- [Андрейчук, 2022] Андрейчук А.А. Эффективный поиск ограниченно-субоптимальных решений задачи многоагентного планирования // Искусственный интеллект и принятие решений. – 2022. – № 1. – С. 57-70.
- [Бурков, 2010] Бурков Е.А. Определение субъективности и надежности экспертных оценок на основе анализа статистических данных // Известия СПбГЭТУ ЛЭТИ. – 2010. – № 9. – С. 33-38.
- [Волков и др., 2023] Волков Е.Н., Аверкин А.Н. Возможности применения объяснительного искусственного интеллекта для обнаружения глаукомы на примере метода LIME // Международная конференция по мягким вычислениям и измерениям. – 2023. – Т. 1. – С. 177-180.
- [Воробьев, 2021] Воробьев А.В. Метод выбора модели машинного обучения на основе устойчивости предикторов с применением значения Шепли // Экономика. Информатика. – 2021. – Т. 48, № 2. – С. 350-359.
- [Кузнецова и др., 2018] Кузнецова А.В., Сенько О.В., Кузнецова Ю.О. Преодоление проблемы "черного ящика" при использовании методов машинного обучения в медицине // Врач и информационные технологии. – 2018. – № S1. – С. 74-80.
- [Хведчук и др., 2018] Хведчук В.И., Антоник И.А. Базовые элементы FEM - анализа для электрических схем // Вестник Брестского государственного технического университета. Физика, математика, информатика. – 2018. – № 5(113). – С. 48-52.
- [Deb et. al., 2021] Deb D., Smith R.M. Application of Random Forest and SHAP Tree Explainer in Exploring Spatial (In)Justice to Aid Urban Planning // ISPRS International Journal of Geo-Information. – 2021. – Vol. 10, No. 9. – P. 629.
- [Mitchell et. al., 2022] Mitchell R., Frank E., Holmes G. GPUTreeShap: massively parallel exact calculation of SHAP scores for tree ensembles // PeerJ. Computer Science. – 2022. – Vol. 8. – P. e880. – DOI 10.7717/peerj-cs.880.

УДК 004.89

doi: 10.15622/rcai.2025.103

ПОДХОД К ПОИСКУ КОМПОНЕНТОВ ДЛЯ ПРОГРАММНЫХ СИСТЕМ НА ОСНОВЕ АНАЛИЗА МУЛЬТИМОДАЛЬНЫХ ДАННЫХ¹

С.С. Шевалдов (*sshevaldov@yandex.ru*)

А.А. Филиппов (*al.filippov@ulstu.ru*)

Ульяновский государственный технический университет, Ульяновск

В статье рассматривается решение проблемы поиска компонентов (зависимостей, библиотек) во внешних репозиториях при разработке программных систем с учетом функциональных требований и других особенностей проекта. Важность и сложность данной проблемы объясняется мультимодальной природой данных о компонентах, сложностью интерпретации и оценки параметров компонентов с учетом различных требований и ограничений. Каждый внешний компонент программной системы, имеет ряд параметров, относительно которых разработчик принимает решение о возможности его использования в проекте. Для решения проблемы поиска подходящих компонентов, ускорения процесса выбора наиболее подходящего компонента, а также повышения качества разрабатываемой программной системы предложен подход к поиску компонентов для программных систем на основе анализа мультимодальных данных. Для проверки адекватности предложенного подхода разработан прототип поисковой системы, который обеспечивает сбор, обработку и анализ мультимодальных данных о компонентах с целью получения комплексной оценки для ранжирования результатов.

Ключевые слова: мультимодальные данные, компоненты, программная система, информационный поиск, контекст.

¹ Исследование выполнено в рамках государственного задания № 075-03-2023-143 по проекту «Исследование интеллектуальной предиктивной аналитики на базе интеграции методов конструирования признаков гетерогенных динамических данных для машинного обучения и методов предиктивного мультимодального анализа данных».

Введение

В настоящее время при разработке программных систем для повышения скорости разработки активно применяются принципы компонентно-ориентированного программирования [Nierstrasz, 1992], [Лаврищева, 2018]. Временные затраты сокращаются за счет повторного использования существующих реализаций требуемых модулей и функций. Компоненты обычно оформляются в виде библиотек и помещаются в специализированные хранилища – репозитории [Trew, 2006]. Для различных платформ и/или языков программирования используются отдельные репозитории. В свою очередь компоненты могут использовать возможности других компонентов, в результате чего формируется дерево зависимостей, на основе которого определяется перечень необходимых для подключения к проекту компонентов.

Корректный выбор компонентов является важным шагом процесса проектирования, который существенно влияет на качество разрабатываемой программной системы. Проблема качественного и обоснованного выбора зависимостей обуславливается требованиями к проекту и широким набором характеристик компонентов. При выборе компонентов необходимо учитывать:

- свойства и требования разрабатываемой программной системы;
- квалификацию и опыт разработчиков;
- функциональные возможности, характеристики и качество самих компонентов.

Данные параметры представляют собой множество мультимодальных данных.

Таким образом, возникает задача разработки метода информационного поиска компонентов для программной системы с учетом различных параметров проекта и множества компонентов, представленных набором мультимодальных данных. При этом такая поисковая система должна ранжировать найденные компоненты на основе значения комплексной оценочной величины.

Процесс выбора оптимальных и качественных компонентов с учетом всех ограничений и требований является сложной и ресурсоемкой задачей, т.к. требует анализа данных из различных источников с учетом динамики развития отрасли в целом.

1. Анализ проблемы

В данной работе изучается проблема поиска компонентов для разрабатываемой программной системы, наиболее точно отвечающим динамически изменяющимся требованиям. Мультимодальный характер данных, на основе которых производится выбор компонентов, обусловлен наличием как качественных, так и количественных показателей, с различными методами измерения и представления.

В работе [Перл, 2022] мультимодальные данные определяются как объект или процесс, который описывается различными характеристиками, представленными в соответствующих модальностях. Под модальностью понимается один из нескольких атрибутов, которые описывают конкретную характеристику объекта. Совокупность характеристик одного мультимодального объекта обозначается сущностными представлениями или образами. Каждый мультимодальный объект необязательно имеет сущностное представление во всех модальностях. Признаки могут быть несовместимы друг с другом (например, имя и периодичность). Объект, имеющий несколько описывающих его модальностей, можно назвать мультимодальным.

Важной проблемой является формулировка общего понятия мультимодальных данных. Несмотря на широкую распространённость термина «мультимодальность» в различных науках, часто под ним понимаются разные понятия и явления. Так, в логике принято говорить о модальностях речи и о модальностях, как отражении вероятности высказывания [Nguyen, 2006].

В транспортной логистике говорят о мультимодальности, как о вариативности и комбинировании видов транспорта при доставке грузов [Karimi, 2018].

В биологии и медицине говорят о мультимодальности как о множественном представлении изучаемого или используемого объекта (мультимодальный МРТ, мультимодальная терапия и пр.).

Другим широко распространённым примером мультимодальных данных является наличие разнообразных форматов данных, в первую очередь это изображения, видео и аудио [Перл, 2022].

По своей природе мультимодальные объекты описываются рядом признаков, которые часто не совместимы друг с другом (например, цветом, весом, ценой), и эти характеристики могут рассматриваться как модальности конкретного объекта. В работе [Перл, 2022] определены необходимые виды связей в наборе мультимодальных данных.

Задаваемые извне связи, о которых известно перед началом анализа мультимодальных данных обозначаются как «прямые». Вычисленные связи, отражающие собой меру подобия (силу связывания w), обозначаются как «косвенные» связи. В работе [Kalyonova, 2017] подробно рассмотрены связи в мультимодальных данных.

Базовая модель представления мультимодальных данных включает набор аспектов, позволяющих сформировать конечное представление базовой структуры мультимодальных данных. Множество входных данных, разделённых по модальностям, рассматривается в виде

$$X^n = \{x_1, x_2, \dots, x_n\},$$

где n – количество уникальных мультимодальных объектов.

Для выполнения анализа полученные наборы данных передаются классификаторам в соответствии с модальностями. Методы классификации зависят от модальностей и не являются четко заданными в рамках всего алгоритма анализа. Классификация проводится с целью определения степени взаимосвязи между представлениями различных мультимедальных объектов.

Результат анализа обычно представляет собой иерархический граф значений в разрезе модальностей. Полученный граф описывает положение конкретной характеристики в общем дереве модальностей.

В рамках поставленной задачи учитываются следующие характеристики компонента:

- Безопасность – сведения об обнаруженных в компоненте уязвимостях (CVE).
- Поддержка проекта – активность разработки, количество незакрытых ошибок, наличие и качество документации.
- Тяжеловесность (зависимости самого компонента) – размер дерева зависимостей компонента с учетом транзитивных связей.
- Совместимость – функциональные и нефункциональные требования, совместимость с платформой и окружением проекта.
- Избыточность – степень использования компонента с учетом его тяжеловесности.
- Популярность – количество использований компонента и степень упоминания компонента в поисковых системах.
- Качество – количество запросов о компоненте с упоминанием различных проблем и ошибок.

Количественными показателями в данном случае являются, к примеру, количество функций компонента, количество незакрытых сообщений об ошибках в репозитории проекта, размер компонента с учетом зависимостей и т. д. Качественными параметрами будут являться: оценка качества поддержки компонента, соответствие требованиям проекта, популярность компонента, и др. Для определения популярности и качества компонента используется анализ динамики количества запросов в поисковых системах с использованием различных шаблонов: «how to», «error», «download» и т.д.

- источниками мультимодальных сведений являются:
- репозитории компонентов для платформы и/или языка программирования (mvn, npm, nuget и др.);
- репозитории компонента, расположенные на IT-хостингах (GitHub, GitLab и др.);
- репозитории сведений об уязвимостях;
- поисковые системы (Yandex, Google и др.);
- большие языковые модели;
- различные сторонние сервисы с полезной статистикой.

Мультимодальные сведения о компонентах позволят реализовать алгоритм информационного поиска компонентов с учетом особенностей и требований конкретного проекта.

Информационный поиск – процесс поиска неструктурированной информации, удовлетворяющей информационные потребности [Закалин, 2018].

Процесс информационного поиска обычно включает в себя последовательность следующих операций, направленных на сбор, обработку и предоставление необходимой информации пользователям с учетом их информационной потребности:

- Уточнение информационной потребности и формулировка запроса.
- Установление совокупности возможных информационных источников.
- Извлечение информации из выявленных информационных источников.
- Ознакомление с полученной информацией и оценка результатов поиска.

В рамках данного исследования информационный поиск компонента рассматривается как комбинация полнотекстового поиска и поиска по метаданным, а информационная потребность пользователя формируется исходя из функциональных и нефункциональных требований проекта и особенностей проекта. Особенности проекта программной системы формируют контекст, в рамках которого учитывается информационная потребность пользователя.

Таким образом, требуется разработка подхода к информационному поиску компонентов для программных систем с учетом мультимодального представления данных и ограничений контекста проекта.

2. Предложенный подход

В данном разделе будет рассмотрен подход к ранжированию компонентов программной системы с учетом степени соответствия информационной потребности на основе анализа мультимодальных данных.

Формально проект программной системы можно представить в виде следующего выражения:

$$P = \langle D_1, D_2, \dots, D_n \rangle,$$

где $D_i = \langle Dp, C, F, T, R, M \rangle$ – компонент программной системы, в котором

Dp – дерево зависимостей i -го компонента;

C – множество уязвимостей i -го компонента;

F – множество функций i -го компонента;

T – информация о динамике популярности i -го компонента;

R – информация о процессе разработки i -го компонента;

M – метаданные i -го компонента.

На данный момент контекст проекта, который описывает особенности окружения и ограничения проекта, включает только множество функциональных требований. Расширение контекста планируется выполнить в рамках дальнейших исследований.

Тогда, алгоритм поиска компонентов по функциональным требованиям формально можно представить в виде следующей функции:

$$IR: F \times \tilde{F} \rightarrow \tilde{D}.$$

Функция IR осуществляет поиск компонентов \tilde{D} , множество функций которых F соответствует функциональным требованиям проекта \tilde{F} .

Функциональные требования проекта \tilde{F} представляют собой текстовое описание, составленное на основе технического задания к разрабатываемой программной системе.

Так как репозитории компонентов содержат большое количество элементов, что затрудняет процесс получение всех данных из такого репозитория, было решено использовать большую языковую модель (LLM) в качестве реализации функции IR . В будущем планируется разработка алгоритма для генерации поискового запроса для выбранного репозитория компонентов на основе функциональных требований проекта \tilde{F} .

Для получения множества компонентов \tilde{D} формируется специальный запрос (промпт) для LLM. Для получения промпта используется заранее подготовленный шаблон, в который подставляются указанные функциональные требования \tilde{F} . Шаблон также определяет формат выходных данных для LLM. В итоге ответ LLM содержит в себе перечень имен библиотек, которые в большинстве случаев соответствуют указанным функциональным требованиям. Планируемая интеграция с поисковыми движками существующих репозиторий компонентов позволит сократить количество ошибок работы LLM за счет определения возможных галлюцинаций.

Дальнейшая работа алгоритма поиска компонентов заключается в ранжировании полученного множества компонентов \tilde{D} по заранее установленным правилам.

Формально функцию вычисления комплексной оценки для компонента на основе анализа мультимодальных данных можно представить следующим выражением:

$$Rate(\tilde{D}_i) = w_1 * g(Dp) + w_2 * h(C) + w_3 * k(F) + w_4 * l(T) + w_5 * m(R),$$

где w_1, w_2, w_3, w_4, w_5 – веса оценок отдельных параметров компонента;

$g(Dp), h(C), k(F), l(T), m(R)$ – функции для вычисления значений оценки отдельных параметров компонента, представленных мультимодальными данными, на основе нечетких экспертных правил. Вопрос расчета данных параметров выходит за рамки данной статьи и будет рассмотрен в одной из следующих публикаций.

Для проверки адекватности предложенного алгоритма поиска компонентов для программной системы на основе анализа мультимодальных данных был разработан прототип поисковой системы.

3. Архитектура и основные функции поисковой системы

В ходе разработки прототипа поисковой были разработаны следующие подсистемы:

1. Пользовательский интерфейс (UI). Интерфейс для ввода функциональных требований и отображения множества найденных компонентов с указанием оценки. Система позволяет ознакомиться с данными каждого из найденных компонентов, а также получить детальные сведения, на основе которых была получена оценка.

2. Подсистема работы с LLM. Данная подсистема формирует промпт на основе введенных пользователем функциональных требований и осуществляет взаимодействие с LLM.

3. Подсистема загрузки. На основе полученных от LLM имен компонентов данная подсистема производит загрузку данных из внешних источников с последующим их сохранением в хранилище (базе данных). В случае, если данные о компоненте содержатся в хранилище и загрузка данных производилась недавно, производится обращение к подсистеме хранения для их получения.

4. Подсистема хранения. Подсистема хранения позволяет получить ранее загруженные данные о компонентах, а также рассчитать значение комплексной оценки для каждого из них.

5. Хранилище на базе СУБД Postgres.

Для демонстрации примера работы прототипа поисковой системы будем использовать следующий поисковый запрос (функциональные требования): «работа с данными в формате dataframe». В текущей реализации в качестве ограничений контекста указывается язык программирования, в данном случае «python». Контекст учитывается при формировании промпта для LLM.

Например, оценка качества процесса разработки компонента $m(R)$ в данной реализации определяется отношением количества закрытых задач (issues) в репозитории проекта компонента к общему числу задач.

В результате работы поисковой системы пользователь получит список компонентов (в данном случае количество ограничено) для указанного языка программирования, отсортированный в порядке убывания вычисленной комплексной оценки $Rate(\tilde{D}_i)$.

На рис. 1 представлен результат работы прототипа поисковой системы.

Введите ключевые слова требуемого функционала

dataframe	pip	Найти зависимости			
Имя зависимости	Код последнего релиза	Дата последнего релиза	Ссылка на репозиторий	Количество закрытых issues/общее количество issue, %	Пакетный менеджер
Pandas	2.2.3	2024-09-20	https://github.com/pandas-dev/pandas	86.49	pip
SciPy	1.9.3	2022-10-20	https://github.com/scipy/scipy	84.06	pip
NumPy	2.2.6	2025-05-17	https://github.com/numpy/numpy	82.66	pip

Рис. 1. Пример результата работы прототипа поисковой системы

Как видно из рис. 1, в библиотеке (компоненте) Pandas закрыто 86,49% задач. На основе только лишь этого значения можно сделать вывод о том, что данная библиотека имеет более высокое качество поддержки.

Для расчета оценки популярности компонента $l(T)$ была собрана статистика использования компонентов в открытых проектах в рамках IT-хостинга GitHub. В ходе такого анализа производился анализ открытых проектов на языке python с целью выявления их зависимостей (используемых компонентов). В результате была получена статистика о частоте использования различных компонентов. Также была собрана статистика о наиболее популярных комбинациях компонентов в программных системах. Иллюстративный пример полученных статистических данных представлен на рис. 2 и 3.

Целью сбора статистики является определение наиболее популярных библиотек для последующей предварительной загрузки данных о них в хранилище поисковой системы, так как взаимодействие с внешними источниками данных требует значительных временных затрат.

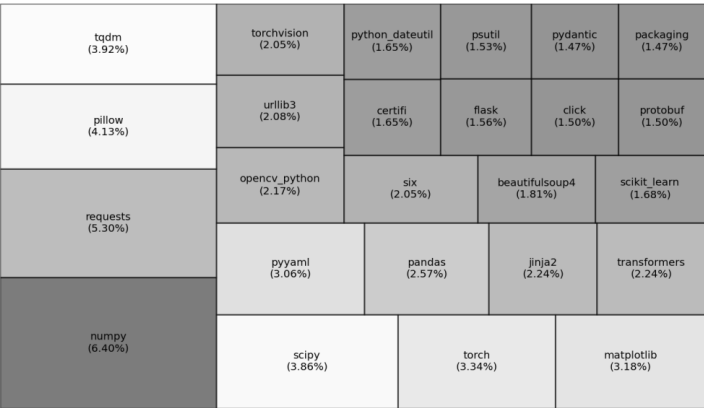


Рис. 2. Диаграмма 25 самых часто используемых компонентов

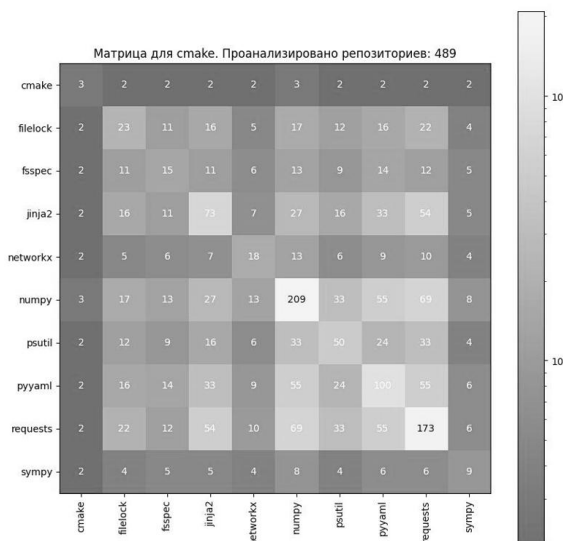


Рис. 3. Матрица комбинаций использования библиотеки «stake»

На рис. 4 представлена ER-диаграмма модели хранилища прототипа поисковой системы.

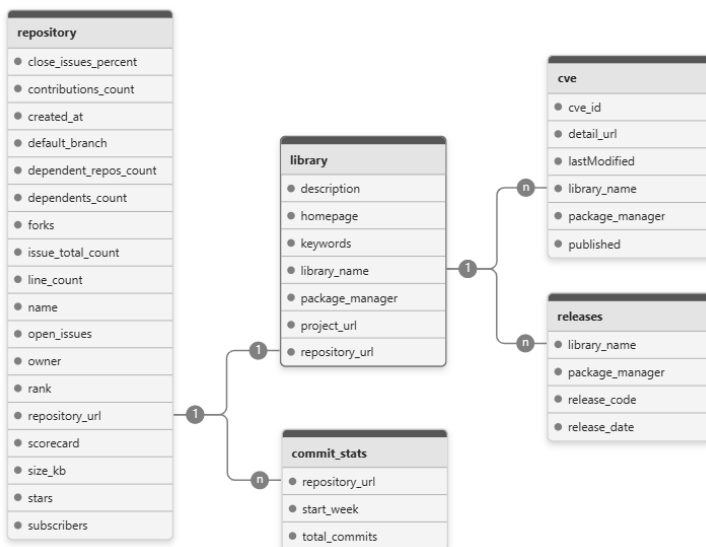


Рис. 4. ER-диаграмма модели данных хранилища поисковой системы

Как видно из рис. 4, сущность Library – основная таблица базы данных хранилища, которая содержит метаданные компонента. Включает в себя имя пакетного менеджера (pypi, pmt или maven), имя библиотеки, её описание, ссылку на проект и репозиторий проекта на IT-хостинге.

Сущность Repository содержит данные, извлеченные из репозитория проекта на IT-хостинге.

Сущность CVE содержит информацию об уязвимостях компонента.

Сущность Releases содержит информацию о динамике релизов компонента. Учитывается при оценке динамики активности процесса разработки.

В дальнейшем поисковая система будет дорабатываться. Планируется разработка модуля анализа текстов на естественном языке для автоматического извлечения функциональных и нефункциональных требований из текста технического задания.

Заключение

В рамках данной работы предложен подход к сбору данных о компонентах программных систем из различных внешних источников для формирования мультимодального представления. Представлена информационная модель компонента программной системы. Описаны шаги алгоритма поиска компонентов для использования в процессе разработки программной системы на основе функциональных требований, представленных в текстовом виде. Алгоритм поиска основан на автоматическом формировании промпта к LLM с учетом функциональных требований и языка программирования. Также в статье представлен алгоритм расчёта комплексной оценки компонента на основе анализа мультимодальных данных для последующего ранжирования компонентов. Описана архитектура текущей реализации прототипа поисковой системы. Поисковая система реализует все необходимые этапы взаимодействия с пользователем: от получения функциональных требований к компоненту, до вывода пользователю списка найденных компонентов с указанием для каждого из них значения комплексной оценки. Для повышения скорости работы поисковой системы был произведен сбор статистики для определения популярных на данный момент компонентов для проектов на языке python. Список популярных библиотек позволяет загрузить данные о них заранее. В качестве планов дальнейшей работы над проектом можно выделить:

- Разработку подсистемы анализа текста технического задания на разработку программной системы с целью получения перечня функциональных и нефункциональных требований, которые должны быть учтены при поиске компонента.
- Интеграцию с поисковыми механизмами репозитория компонентов для повышения полноты результатов и устранения проблемы галлюцинаций LLM.

Список литературы

- [Nierstrasz, 1992] Nierstrasz O., Gibbs S., Tsichritzis D. Component-oriented software development // Communications of the ACM. – 1992. – Vol. 35, Iss. 9. – P. 160-165.
- [Лаврищева, 2018] Лаврищева Е.М. Программная инженерия. Парадигмы, технологии и case-средства: учебник. 2-е изд.. – М.: Изд-во «Юрайт», 2018. – 280 с.
- [Trew, 2006] Trew T., Soepenbergh G. Identifying technical risks in third-party software for embedded products // Fifth International Conference on Commercial-off-the-Shelf (COTS)-Based Software Systems (ICCBSS'05). – IEEE, 2006.
- [Перл, 2022] Перл О.В., Перл И.А. Применение мультимодального подхода для выявления подобия в многомерных наборах данных с примером использования // International Journal of Open Information Technologies. – 2022. – Vol. 10, No. 1. – P. 41-53.
- [Nguyen, 2006] Nguyen L. A. Multimodal logic programming // Theoretical Computer Science. – 2006. – Vol. 360, Iss. 1–3. – P. 247-288.
- [Karimi, 2018] Karimi B., Bashiri M. Designing a Multi-commodity multimodal splittable supply chain network by logistic hubs for intelligent manufacturing // Procedia manufacturing. – 2018. – Vol. 17. – P. 1058-1064.
- [Kalyonova, 2017] Kalyonova O., Perl I. Revealing of entities interconnections in system dynamics modelling process by applying multimodal data analysis paradigm // 2017 21st Conference of Open Innovations Association (FRUCT). – IEEE, 2017. – P. 156-161.
- [Закалин, 2018] Закалин И.Ю. Методы поддержки принятия решений на основе информационных технологий // Вестник магистратуры. – 2018. – №. 1-1(76). – С. 23-26.

Abstracts of Sectional Reports

Section 8

Machine Learning, Neural Network Methods, Neuroinformatics

PRELIMINARY DESIGN OF AIRCRAFT: THE USE OF MACHINE LEARNING FOR CONCEPTUAL AIRCRAFT DESIGN

E.O. Adigwe (*elpromzeegm@gmail.com*)

N.M. Borgest (*mail@borgest.ru*)

Samara National Research University named after the Academician
S.P. Korolev, Samara

This paper focuses on the application of Artificial Intelligence (AI) in the field of aircraft design. Particularly, it considers the use of Machine Learning (ML) to create concept of an aircraft based on previous designs. This research uses similarity in the evolution of a living creature and artefact, precisely the initial concept of passenger airplane and fighter-jets. Also, a comprehensive survey in the field of automated design and production of aircrafts was carried out in order to see the tendencies in developing concept-projects. ML and AI are considered as an auxiliary instrument for the design of aircrafts, because they can help to accelerate the design process and make quick decision.

Keywords: Artificial Intelligence, Machine Learning, aircraft, evolution, concept-project.

IMPROVING THE EFFECTIVENESS OF THE EDUCATIONAL PROCESS USING NEURAL NETWORKS

S.V. Alubin (*sergei.alubin@yahoo.com*)

ITMO University, Saint Petersburg

The current and demanded task of improving the quality of the educational process in educational institutions is investigated. Traditional teaching methods do not consider the individual characteristics and wishes of students, which affects their academic performance. The article proposes an adaptive learning system in which researchers focused on finding solutions to the issues of organizing a more effective learning process, stimulating students' interest and academic performance using modern artificial intelligence technologies. As part of this work, an experiment was conducted and a personalized learning approach using neural networks was proposed. The results of the student survey and the tests performed by the students during the training course were used as data for the experiment. An architecture using convolutional and recurrent neural networks has

been developed. According to the results of the experiment, students' academic performance increased to 91.8% and the personalization of the educational process increased to 95.3%. This technique can be applied as part of the creation of a flexible learning system in order to improve the quality of educational services and student academic performance.

Keywords: artificial intelligence, neural networks, educational process, personalized learning.

APPLICATION OF 2D SEMANTIC SEGMENTATION METHODS BASED ON TRANSFORMER ARCHITECTURE FOR REMOTE INVESTIGATION OF ARCHAEOLOGICAL SITES STRUCTURE

A.V. Vokhmintsev (*vav@csu.ru*)^A

V.R. Abbazov (*abbavar@yandex.ru*)^B

M.A. Romanov (*std.romanov.ma@gmail.com*)^A

T.S. Vokhmintseva (*Atska85@mail.ru*)^B

M. Khater (*Mostafa.khater2024@xzhmu.edu.cn*)^{B,C}

A.V. Melnikov (*MelnikovAV@uriit.ru*)^B

^A Chelyabinsk State University, Chelyabinsk

^B Yugorsk State University, Khanty-Mansiysk

^C Suizhou University, Suizhou

This work is devoted to the development of methods for semantic segmentation of archaeological sites using machine learning methods and 2D digital models of the archaeological landscape based on aerial photography and Earth remote sensing of t The paper provides a comparative analysis of modern models for object detection and image segmentation, examines architectures based on the transformer model, such as Internimage, BEiT (Pre-Training of Image Transformers), trained models using a collection of data on archaeological sites of the Bronze Age of the Southern Urals with an assessment on a validation set in order to prevent overfitting models, The results obtained are presented and discussed.

Keywords: semantic segmentation, transformers, digital relief model, archaeological site.

ANCIENT RUSSIA HANDWRITTEN HERITAGE CORPUS BUILDING SUPPORT SYSTEM

D.V. Demidov (*dvdemidov@mephi.ru*)

NRNU MEPhI, Moscow

The paper describes an ancient Russia handwritten heritage corpus building support system and a character recognition technology for manuscripts. It presents a concept of merging optical character recognition and machine learning tools with lingware and

software tools for engineers and philologists into a single environment. The goal of collaboration is to build a digital corpus of manuscripts accessible for paleographic and textual research.

Keywords: optical recognition, corpus linguistics, ancient manuscripts, evolutionary modeling.

FINDING THE SIMILARITY OF WINNING STRATEGIES IN A MACHINE LEARNING TASKS

D.A. Dobrynin, (*rabota51@mail.ru*)

Federal Research Center “Computer Science and Control”
of the Russian Academy of Sciences, Moscow

The paper describes a method for obtaining generalized game strategies for data mining tasks using the ideas of similarity and inductive inference. The widely known tic-tac-toe game is used as a model. To obtain the data, either a complete tic-tac-toe game tree or a set of random games can be used. The principles of obtaining similarity of game strategies and an algorithm for obtaining generalized strategies are described. Some results of computer modeling are presented. It is shown that the application of this method makes it possible to obtain generalized winning strategies. The strategies obtained are well explicable and can be easily interpreted by humans.

Keywords: machine learning, tic-tac-toe game, decision making, inductive inference.

REINFORCEMENT LEARNING MODEL USING ADAPTIVE NETWORK AND FUZZY INFERENCE

A.P. Ereemeev (*eremeev@appmat.ru*)

V.S. Petrov (*PetrovVS@mpei.ru*)

NRU "MEI", Moscow

The paper proposes a deep reinforcement learning (DRL) model based on an adaptive neural network and a fuzzy inference system. The primary goal is to develop a high-performance DRL agent with enhanced explainability compared to traditional "black-box" agents. The approach is tested and comparatively analyzed using the CartPole task. Testing and comparative analysis are conducted on the CartPole task. The results confirm the potential of hybrid models combining DRL and fuzzy logic methods for creating efficient and interpretable AI systems that are critical in application areas that require trust and transparency.

Keywords: artificial intelligence, reinforcement learning, neural network, fuzzy logic, fuzzy system, real-time system.

NEURAL NETWORK MODEL OF TEMPORAL AGGREGATION OF LECTURE VIDEO FRAMES FOR IMAGE RECONSTRUCTION TASK

M.E. Ismagulov (*m_ismagulov@ugrasu.ru*)^A

A.V. Melnikov (*melnikovav@uriit.ru*)^{A,B}

^A Yugra State University, Khanty-Mansiysk

^B Yugra Research Institute of Information Technologies,
Khanty-Mansiysk

During multimodal processing of lecture videos, the task of extracting data from the video stream arises. A common challenge in this process is the occlusion of content by the lecturer. To address this issue, temporal frame aggregation methods are widely used for image reconstruction in scenarios involving moving objects in the frame. This paper presents the development of a neural network model for temporal aggregation of video frames to reconstruct areas of the board occluded by the lecturer. The development and training process of the neural network model is described, with supervised learning chosen as the training method. The core dataset consists of lecture video frames divided into samples paired with reference board images. The model is based on a hybrid architecture combining a convolutional neural network (CNN) and a recurrent layer with long short-term memory (LSTM). The machine learning library TensorFlow-Keras was used for implementation, and ONNXRuntime was employed for inference. The training results yielded metrics of accuracy: 0.78, loss: 0.092, and a PSNR of 35 dB, indicating high-quality image reconstruction. The baseline model was adopted from the paper Recurrent Temporal Aggregation Framework for Deep Video Inpainting.

Keywords: temporal video frame aggregation, convolutional neural networks, recurrent neural networks, neural network training, neural network inference, image reconstruction, multimodal lecture video processing.

AUTOMATIC THESAURUS CONSTRUCTION: EVALUATION OF CLUSTERING METHODS AND VECTOR REPRESENTATION OF WORDS

I.A. Kolomoitseva (*bolatiger@mail.ru*)

O.I. Fedyaev (*olegfyayev@mail.ru*)

Donetsk National Technical University, Donetsk

The article examines the methods of automated thesaurus construction for the CodeLex intelligent search system, with an emphasis on identifying synonyms between terms in the field of programming. The application of clustering algorithms (K-means, DBSCAN) in combination with vector representation models of words (Word2Vec, BERT) to determine the relationship of synonymy is considered. The experiments were

conducted on fragments of texts from a PHP book. The results showed that the Word2Vec+K-means bundle is the most promising for determining synonymy between function names and their descriptions. BERT has demonstrated low efficiency in this task. The results obtained can be used to improve the operation of the Codeplex system.

Keywords: information search, thesaurus, clusterization, DBSCAN, K-means, Word2Vec, BERT.

DEEP FOREST FOR SURVIVAL ANALYSIS WITHIN MULTIPLE INSTANCE LEARNING

A.V. Konstantinov (*konstantinov_av@spbstu.ru*)

L.V. Utkin (*utkin_lv@spbstu.ru*)

Peter the Great St. Petersburg Polytechnic University, Saint Petersburg

This work addresses the problem at the intersection of machine learning and survival analysis within a multiple instance learning (MIL) framework applied to tabular data. The challenge arises when each object is characterized not by a single feature vector, but rather by a set of vectors, and the event of interest occurs for the object as a whole. To address this issue, we propose a novel method based on deep forests. A key characteristic of our approach is the construction of a model that processes multiple feature vectors as input, comprising ensembles of decision trees. Numerical experiments using real-world data demonstrate the applicability and effectiveness of the proposed method.

Keywords: survival analysis, multiple instance learning, decision trees, ensembles, deep forest.

INVESTIGATION OF THE POSSIBILITIES OF COMBINING PROTECTION METHODS FOR IMAGE RECOGNITION SYSTEMS AGAINST GRADIENT ADVERSARIAL ATTACKS

I.V. Kotenko (*ivkote@comsec.spb.ru*)

I.B. Saenko (*ibsaen@comsec.spb.ru*)

V.E. Sadovnikov (*bladimir1998@mail.ru*)

St. Petersburg Federal Research Center of the Russian Academy of Sciences,
St. Petersburg

The paper discusses an approach to protecting image recognition systems from gradient adversarial attacks based on a combination of various protection methods, namely: noise reduction, compression and neural image cleaning. Combining protection methods involves finding optimal parameters that characterize these methods, at which recognition accuracy becomes maximum. The effectiveness of the considered protection methods is evaluated on the STL-10 dataset. The choice of this dataset is due to its wide

application in image classification tasks. The experimental results show that combining these protection methods makes it possible to achieve maximum image recognition accuracy when exposed to gradient adversarial attacks.

Keywords: protection from attacks, artificial intelligence, adversarial attacks, image recognition, machine learning, noise reduction, cybersecurity.

APPLYING CLASSIFICATION AND CLUSTERING METHODS IN THE PROCESS OF INVESTIGATING INFORMATION SECURITY INCIDENTS

A.V. Meshkov (*meshkov.av@itmo.ru*)

ITMO University, Saint-Petersburg

This paper explores approaches for addressing the escalating complexity of cyber threats through the development of methods and algorithms for investigating security incidents in Windows and Linux environments. Using the MITRE ATT&CK framework, the proposed approach facilitates the mapping of security events to associated tactics and techniques, enabling structured classification and clustering. Advanced pre-processing techniques and machine learning algorithms are applied to optimize event analysis, significantly reducing processing time while preserving analytical precision. Experimental evaluations validate the framework's effectiveness in enhancing incident detection and response capabilities. Future research directions include refining algorithmic performance and adapting the framework to emerging cybersecurity challenges.

Keywords: cybersecurity incidents, MITRE ATT&CK, machine learning.

APPLICATION OF REINFORCEMENT LEARNING IN MULTIPLAYER STRATEGY GAMES

A.S. Misnik (*misnik.as@phystech.edu*)

Moscow Institute of Physics and Technology, Moscow

The paper considers modern methods used for multi-agent reinforcement learning on the example of multiplayer strategy games. The model of the game is described, revealing the preferences of cooperative multi-agent interaction. The results of experiments in the model of the described game, the features of each of the considered approaches, as well as the proposed optimizations that contribute to accelerating and improving the quality of learning are presented.

Keywords: reinforcement learning, multi-agent learning, neural network methods, graph neural networks, reward decomposition.

IDENTIFYING ANOMALIES IN FINANCIAL TIME SERIES BASED ON A MULTI-SCALE AUTOENCODER

A.Yu. Puchkov (*putchkov63@mail.ru*)^A
V.S. Minin (*Mininvssm@yahoo.com*)^B

^A National Research University «Moscow Power Engineering Institute»,
(Branch) in Smolensk; Smolensk
^B LLC "VISOM", Smolensk

A method for detecting anomalies in financial time series, the data in which are displayed with different time discretization intervals, is proposed. The novelty of the method lies in the use of a multiscale autoencoder, to the input of which an extended data set is fed, including both the values of the time series itself with different time discretization, and their first and second differences. The conducted model experiments showed that such an extension increases the error of the autoencoder when restoring input data with anomalies, compared to the error of the autoencoder without such an extension.

Keywords: data anomaly detection, financial time series, artificial neural networks, autoencoder.

A METHOD FOR ENFORCING COVERAGE CONSTRAINT IN LEARNING-TO-DEFER PROBLEMS

A.V. Ponomarev (*ponomarev@iias.spb.su*)
SPC RAS, St. Petersburg

Machine learning models are frequently deployed as components within broader decision-making systems that incorporate human expertise. Learning-to-defer frameworks address this setting by jointly modeling the error distributions of both the algorithmic model and human experts, dynamically routing samples to optimize total system accuracy. A persistent challenge, however, is the inability of current methods to bound the proportion of cases referred to human experts—a practical necessity given finite expert availability and bandwidth. The paper introduces a method to enforce coverage constraints on expert deferral rates while maintaining competitive accuracy. The method's modular design allows seamless integration with diverse existing paradigms, including both learning-to-defer and rejection learning architectures.

Keywords: rejection learning, learning to defer, human-AI system, decision-making, human-AI collaboration.

AUTOMATIC TAXONOMY ENRICHMENT IN RUSSIAN LANGUAGE WITH LARGE LANGUAGE MODELS

F.A. Sadkovsky (*sadkovsky@iling-ran.ru*)^{A,B}

N.V. Loukachevitch (*louk_nat@mail.ru*)^A

I.Yu. Grishin (*grishin@sev.msu.ru*)^A

^A Lomonosov Moscow State University, Moscow

^B Institute of Linguistics RAS, Moscow

This paper presents the first investigation into the potential of large language models for the task of automatic taxonomy enrichment using Russian language material. The authors have adapted the TaxoLLaMA methodology, which had previously demonstrated high efficacy for the English language, by utilizing data from the Russian-language thesaurus, RuWordNet. As part of this research, a comparative performance analysis was conducted between multilingual models and models specifically adapted for the Russian language. The experiments confirmed the successful applicability of the method to Russian-language data and revealed a significant advantage of the Russian-specific models. After fine-tuning, the YandexGPT-5-8B-Lite model surpassed the previous best result, achieving a Mean Average Precision (MAP) score of 55.42.

Keywords: taxonomy, taxonomy enrichment, large language models, TaxoLLaMA, RuWordNet, RuAdapt.

HUMAN IMAGE RECOGNITION BASED ON INFRARED AND COLOR CAMERA DATA USING COMPARTMENT SPIKING NEURAL NETWORK

I.S. Fomin (*i.fomin@rtc.ru*)

Russian State Scientific Center for Robotics and Technical Cybernetics,
Saint-Petersburg

This study addresses the processing of images captured by different types of cameras. As a case study, human recognition is demonstrated on a subset of a publicly available dataset. We propose a training algorithm for a Siamese neural network to map images into an embedding vector space. An approach for transforming these vector representations into spiking neural network (SNN)-compatible encodings is introduced. The paper describes a classification algorithm for processing infrared (IR) and RGB images independently, as well as a method to merge them into a unified dataset for joint processing. Experimental results demonstrate the high efficacy of the proposed methods. The standalone processing achieved an F1-score of 0.891 for IR images and 0.965 for RGB images. Joint processing yielded an F1-score up to 0.957.

Keywords: person reidentification, convolutional networks, siamese networks, neural networks, IR images, RGB images.

USING ENSEMBLE METHODS TO INCREASE THE ACCURACY OF CLASSIFICATION OF NOISY OBJECTS

M.V. Fomina (*m_fomina2000@mail.ru*)

G.V. Shvets (*ShvetsGV@mpei.ru*)

National Research University “MPEI”, Moscow

The problem of intelligent analysis of noisy data obtained from real sources is considered. A method for organizing a classification system for objects affected by various types of noise is proposed. The basis is a combination of ensemble methods with a neural network approach. An architecture of a neural network that solves the problem is proposed. The results of a machine experiment are presented, and the impact of various types of noise on classification results is assessed.

Keywords: Data Classification, Noise Models, Neural Network, Ensemble method, Data Preprocessing.

Section 9

Tools and Technologies for Intelligent Systems Designing

OPTIMIZING GENERATIVE SYSTEMS WITH UNIKERNEL

E.K. Berkheev (*auzza@bk.ru*)

A.A. Romanov (*romanov73@gmail.com*)

Ulyanovsk state technical university, Ulyanovsk

The paper describes an approach that accelerates neural network models by using a unikernel environment. The existing deep learning architectures, such as CNN, ResNet, and DenseNet, are analyzed in terms of their applicability to lightweight virtualization. A methodology for optimal deployment of models in a unikernel environment is proposed, including the selection of an appropriate architecture and the implementation of an environment with minimal costs. Experimental studies are conducted demonstrating the advantages of the proposed approach, including improved response time and reduced resource consumption. Conclusions and recommendations for further research in the field of generative artificial intelligence and serverless architectures are formulated. The novelty of the approach lies in the integration of modern generative artificial intelligence methods with unikernel technology.

Keywords: unikernel, image generation, services, neural networks.

DEVELOPMENT OF A VIRTUAL HANGAR FOR DESIGNING AND STUDYING AIRCRAFT STRUCTURES

N.M. Borgest (*borgest@yandex.ru*)

S.A. Vlasov (*vlasov.ssau@mail.ru*)

Samara University (Samara National Research University
named after academician S.P. Korolev), Samara, Russia

The article is devoted to the development of a virtual hangar – an interactive environment for designing and studying aircraft structures using virtual (VR) and augmented reality (AR) technologies, standard computer-aided design (CAD) systems. An analysis of technical, ergonomic and pedagogical aspects is provided, as well as prospects for using AR/VR in engineering practice. The stages of developing a virtual hangar are described, including the creation of 3D models, object conversion algorithms and integration with an intelligent designer assistant. The solutions were tested in the educational process, which confirmed their effectiveness. Recommendations for scaling and implementing the developed technologies in aircraft manufacturing and education are proposed.

Keywords: virtual reality, augmented reality, aircraft structures, CAD, robot designer, virtual hangar.

ONTOLOGICAL PATTERN FOR INTELLECTUAL SUPPORT OF REHABILITATION TEAM

V.V. Gribova (*gribova@iacp.dvo.ru*)

E.A. Shalfeeva (*shalfe@dvo.ru*)

Institute of Automation and Control Processes FEB RAS, Vladivostok

One key area for optimizing the rehabilitation process is implementing intelligent computer systems that can support clinical decision-making with a comprehensive analysis of medical data. Ontological engineering and methods of logical inference based on formalized knowledge lay the groundwork for aggregating expert and normative knowledge and applying explainable artificial intelligence methods. This study aims to develop an ontological pattern to create intelligent computer assistants in rehabilitation that can analyze clinical documents and medical knowledge. The developed ontological pattern includes the structure of the electronic health record (EHR) with patient information, knowledge ontologies for clinical tasks, specialized solvers, a medical terminology dictionary for rehabilitation specialists, a GUI component for work coordination, and tools for filling and developing knowledge. Construction of systems using the pattern involves filling knowledge bases and integrating them with specialized solvers and GUI components. These computer assistants are designed to effectively assist with current rehabilitation issues, from diagnosing functional disorders to planning treatment programs.

Keywords: ontological pattern, rehabilitation goals, ICF codes, rehabilitation plan.

METHODS OF CONSTRUCTING AN ECOSYSTEM OF KNOWLEDGE ON THE EXAMPLE OF ENERGY

L.V. Massel (*massel@isem.irk.ru*)

A.G. Massel (*amassel@isem.irk.ru*)

V.R. Kuzmin (*kuzmin_vr@isem.irk.ru*)

T.G. Mamedov (*mamedowtymur@isem.irk.ru*)

D.A. Gaskova (*gaskovada@isem.irk.ru*)

Melentiev Energy Systems Institute SB RAS, Irkutsk

The paper discusses method proposed for integration of various services within the context of the ecosystem-based knowledge architecture in the energy sector. The scientific team represented by the authors has developed software solutions for implementing semantic modeling, a software package for predictive research of the fuel and energy complex from the point of view of energy security, as well as an information and computing system for conducting computational experiments and solving problems related to estimating pollutant emissions from energy facilities. The paper considers the design of the interaction of existing solutions as services of the developed ecosystem of knowledge in the energy sector.

Keywords: ecosystem of knowledge, energy, semantic modeling, services, knowledge processing.

TENSOR-BASED REPRESENTATION OF HETEROGENEOUS SEMANTIC NETWORKS FOR KNOWLEDGE BASE FUSION

A.I. Palchevskiy (*apalchev@gmail.com*)

A.I. Molodchenkov (*aim@tesyan.ru*)

Federal Research Center "Computer Science and Control" of the RAS,
Moscow

The paper explores a method for transforming knowledge bases, built on heterogeneous semantic networks, into a family of tensor structures. A metric for assessing the proximity of nodes in a knowledge base is introduced. The conversion of a knowledge base into a set of tensors constitutes one of the stages in developing an algorithm for automated knowledge base merging. The proposed approach can also be applied to ontologies.

Keywords: tensors, heterogeneous semantic networks, machine learning, RCAI-2025.

ONTOLOGICAL SOFTWARE ENVIRONMENT FOR INTELLIGENT TUTORING BASED ON INTEGRATED EXPERT SYSTEMS: METHODS OF CONSTRUCTING BASED ON THE SITUATIONAL CONTROL PRINCIPLES

G.V. Rybina (*GVRybina@ivanov.ru*)

A.A. Grigoryev (*grigandal625@gmail.com*)

National research nuclear university «MEPhI», Moscow

The article describes the results of research related to the construction of an adaptive ontological software environment (AOE) for intelligent tutoring based on the use of tutoring IES and web-IES developed on the basis of the problem-oriented methodology and intelligent software environment tools of the AT-TECHNOLOGY workbench. The AOE architecture has been developed as a hybrid intelligent control system (ICS) for the functioning of tutoring IES and web-IES using the principles of situational control of complex organizational and technical systems. The proposed models and methods of implementing a situational approach to solving the AOE control problem using a generalized ontology of knowledge representation about the direction/specialty of training and basic tools of inference are considered.

Keywords: intelligent tutoring systems, intelligent tutoring, integrated expert systems, IES, tutoring IES, web IES, problem-oriented methodology, AT-TECHNOLOGY workbench, adaptive ontological environment, generalized ontology model, situational control, intelligent control system, ICS.

Section 10

Applied Intelligent Systems

THE CONCEPT OF INTELLIGENT HEALTH ANALYSIS SYSTEM FOR PATIENTS WITH CHRONIC DISEASES USING FUZZY MODELS

T.V. Afanasieva (*afanaseva.tv@rea.ru*)

Plekhanov Russian University of Economics, Moscow

The paper presents a concept of spatial-temporal analytics of multidimensional numerical data of patients with chronic non-communicable diseases (CND), based on the use of fuzzy models. Due to the increase in life expectancy, the number of patients with CND will increase, which will lead to an increase in the workload of medical personnel. To support medical decision-making when managing such patients, it is important to calculate not only the assessments of the condition, but also changes in the health indicators of patients. The article formulates the problem and presents a model of an intelligent system for assessing the health of patients with chronic diseases in the spatial-temporal aspect. Application options for the proposed model are considered.

Keywords: multidimensional health indicators, CND, fuzzy inference system, fuzzy time series.

REDUCING THE COMPLEXITY OF DATA MINING BASED ON THE METHOD OF AUTOMATIC LOCALIZATION OF ANOMALIES IN MULTIDIMENSIONAL SPACES

A.Y. Buchaev (*abdulhamid0055@yandex.ru*)

I.I. Komarov (*i_krov@mail.ru*)

ITMO University, Saint Petersburg

The paper presents a method for reducing the complexity of data mining, based on the automatic localization and grouping of homogeneous and anomalous objects in multidimensional spaces. The proposed method is designed for dynamic operation with changing input data, which is achieved through the use of statistical analysis and a formal decision-making method. This also allows for easy interpretation of the grouping and localization results. Thus, the data mining process focuses on statistically unstable instances within the sample and separates them from each other, which reduces the risk of evasion attacks using visible malicious inserts. As a practical application of the obtained results, the dynamic masking of a visual transformer's attention matrix is considered. The application of the proposed solution eliminates the quadratic complexity of the attention mechanism without the need for preliminary mask adjustment and operates in real-time.

Keywords: cybersecurity, anomaly threshold determination, feature vector, statistical robustness, data mining, decision making mechanism, visual transformer.

MULTISENSORY DATA FUSION FROM MULTISTAGE PRODUCTION TO PREDICT PROCESS EQUIPMENT FAILURES

A.N. Vetrov (*vetrov_48@mail.ru*)

B.V. Paliukh (*pboris@tstu.tver.ru*)

Tver State Technical University, Tver

In this paper, we propose a new approach that combines multisensory data to detect the bifurcation of a multidimensional random process generated by the operation of technological equipment for continuous multistage production. The article describes a parametric multidimensional model for detecting time series anomalies, a method and procedure for determining the bifurcation point. The sliding window method for combining multisensory data and their aggregation for recording anomalies in real time is considered.

Keywords: multisensory data fusion, dynamic expert systems, diagnostics, bifurcation point.

THE INTERNET OF MILITARY THINGS IN THE ARMED FORCES

G.P. Vinogradov (*wgp272ng@mail.ru*)

Scientific Research Institute "Centrprogramsistem", Tver

The article is devoted to an urgent problem – the creation of the Internet of Military things. The prerequisites for relevance are considered. The article considers the status of the issue and the results of the use of smart city technology in the US armed forces. It is shown that responsive sensor networks determine the effectiveness of such technologies. A variant of the architecture of a responsive sensor network is proposed. A variant of the construction of architectural elements and the most important algorithms is presented: positioning and tracking the mobility of targets.

Keywords: Internet of Military Things, responsive sensor network, control, tracking, measurement model.

A HYBRID CASE EXTRACTION METHOD FOR DECISION SUPPORT IN LASER-BASED ADDITIVE MANUFACTURING

V.V. Gribova (*gribova@iacp.dvo.ru*)

A.I. Nikitin (*anikitin@iacp.dvo.ru*)

V.A. Timchenko (*vadim@iacp.dvo.ru*)

Institute of Automation and Control Processes FEB RAS, Vladivostok

We present a hybrid case extraction method combining domain expert knowledge with k-nearest neighbors algorithm for case-based decision support in laser-based additive manufacturing (LBAM). The method's key feature involves domain-aware similarity metrics

for heterogeneous parameters of terms of reference for the LBAM technological operations. The proposed method can be used in the development of intelligent assistants for LBAM technologists working with directed energy deposition technology.

Keywords: decision support, case-based reasoning, knowledge-based case retrieval, case similarity measures, laser processing of materials.

VIRTUAL TUTOR BASED ON INTEGRATION OF LLM WITH A COGNITIVE ARCHITECTURE

Anatoly A. Dolgikh (*tolick.dolgih2013@yandex.ru*)

Dmitry L. Khabarov (*vip.dima020210@mail.ru*)

Alexei V. Samsonovich (*avsamsonovich@mephi.ru*)

National Research Nuclear University MEPhI, Moscow

Biologically inspired cognitive architectures (BICA) are capable of performing complex human-level cognitive functions. However, for effective interaction with a social environment, they require a language interface that supports the processing of both the core meaning of utterances and their connotations, as used, for example, in the eBICA architecture. Large Language Models (LLMs) can effectively serve as a bridge between eBICA's internal representations and the connotations of speech acts. This leads to the idea of a new type of neuro-symbolic integration. To assess the viability of this concept, we implemented a prototype of a Virtual Tutor that trains students in essay writing. It combines an eBICA cognitive module, implementing the principles of self-regulated learning through moral schemas, with an interface module based on OpenAI's GPT-4o LLM. An experiment involving 48 students showed a significant and sustained improvement in essay quality, evaluated across 16 criteria. The approach is expected to be effective in other educational domains—such as mathematics, programming, and foreign languages—and more broadly, in interactive applications that require a nuanced understanding of social interaction aspects.

Keywords: Intelligent Tutoring Systems (ITS), human-computer interaction, cognitive architectures, large language models, multimodal interface, virtual intelligent agents, social interactions.

PROGNOSTIC SYSTEM FOR MONITORING PATIENTS WITH CHRONIC LYMPHOCYTIC LEUKEMIA

N.A. Ignatev (*n_ignatev@rambler.ru*)^A

E.N. Zguralskaya (*iatu@inbox.ru*)^B

M.B. Марковцева (*mmark7@yandex.ru*)^C

^A National University of Uzbekistan, Tashkent

^B Ulyanovsk State Technical University, Ulyanovsk

^C Ulyanovsk State University, Ulyanovsk

Intelligent systems for monitoring health status in clinical practice optimize the provision of medical care to patients. Such systems allow tracking changes in the values of observed parameters over time and making adjustments to the disease prognosis. The

results of intelligent analysis obtained in the form of production rules are especially useful for representing expert knowledge and automating decision-making processes. To formalize knowledge, an ontology has been developed that is used to monitor the condition of patients for 5 years from the moment of diagnosis of chronic lymphocytic leukemia (CLL). The main relationships of concepts that are sufficient to solve the problems of monitoring the condition of patients with CLL during treatment have been identified. The developed ontology is part of the prognostic system, where the monitoring parameters can change taking into account the stage of treatment, clinical, and laboratory data of patients. A knowledge extraction method based on the theory of fuzzy sets is described. A monitoring system associated with a knowledge base is described.

Keywords: ontology, production rules, knowledge base, data mining, chronic lymphocytic leukemia.

CLASSIFYING BIM MODEL ELEMENTS USING A GRAPH NEURAL NETWORK ARCHITECTURE

D.D. Kabailov (*denis2000kab@gmail.com*)

A.M. Namestnikov (*am.namestnikov@gmail.com*)

St. Petersburg State University of Architecture and Civil Engineering

This article describes a method for classifying elements of building information models based on their visibility in specific view categories, using a custom-developed graph neural network. The first part outlines an analysis of existing solutions for generating drawings and their features and disadvantages is presented. The second part outlines the core principles of constructing and structuring a disconnected graph representing a collection of building information models. Visual materials illustrating the resulting disconnected graph are also provided. The third part explains the rationale behind choosing a graph attention network (GAT) and the GATv2Conv layers. It describes the architecture of the neural network trained on the constructed graph, which is capable of predicting visibility labels for elements in different views of building information models. Preliminary test results of the trained neural network are also presented.

Keywords: graph neural network, convolution neural network, GAT, BIM, building design.

ALGORITHMS FOR CLASS IDENTIFICATION IN THE TASK OF RECOGNIZING KEYBOARD HANDWRITING USING MACHINE LEARNING

D.N. Kobzarenko (*kobzarenko_dm@mail.ru*)^{A,B}

A.G. Mustafaev (*arslan_mustafaev@mail.ru*)^{A,B}

^A Dagestan State Institute of National Economy, Makhachkala
^B MIREA Russian Technological University, Moscow

As part of scientific research on identity identification based on keyboard handwriting using modern artificial intelligence methods, experimental results were obtained with a database of 35 subjects. Based on the newly developed class identification algo-

rithms, the maximum accuracy (100%) of class recognition in test samples was obtained. The results show the high efficiency of the selected models, source processing methods, and class identification algorithms in the task of recognizing a person by keyboard handwriting. The experiments described in the paper are confirmed by Python source codes and are available on GitHub as a Google Colaboratory notebook with test arrays at the link: https://github.com/kobzarenko/KII_2025.git.

Keywords: keyboard handwriting, machine learning, classification task, CatBoost, Python.

TOWARD THE ISSUE OF PSYCHOPHYSIOLOGICAL CHARACTERISTICS IN HUMAN-MACHINE SYSTEMS

M.N. Koroleva (*maria.svyatkina@gmail.com*)

Bauman Moscow State Technical University, Moscow

Human-machine systems of semi-automatic or automatic types can and should take into account the characteristics of a person and adapt to him. The paper proposes to consider the operator in the system circuit also as a control object (for the possibilities of system adaptation). Consequently, the input information flow must undergo transformations aligned with the goals and criteria of control. These transformations should be based on a key component – the control object model (the operator model) – where psychophysiological characteristics constitute a critical element. This approach is applicable to a wide range of human-machine systems, from industrial process control systems to personalized educational trajectory design. A central focus of this paper is the translator component, which converts the model's key characteristics into the format for presenting information units (or their combinations). An example is given of changing the form of didactic unit presentation based on the consideration of learners' temperaments.

Keywords: human-machine systems, psychophysiological characteristics, world model, control object, control systems, operator.

ANOMALY DETECTION IN CONTAINER SYSTEMS BASED ON PROFILING AND NEURAL NETWORK AE-LSTM

I.V. Kotenko (*ivkote@comsec.spb.ru*)

M.V. Melnik (*mkmxvh@gmail.com*)

St. Petersburg Federal Research Center of the Russian Academy of Sciences,
Saint Petersburg

The architecture features of container systems create additional risks, allowing attackers to negatively affect not only the target container, but also other containers located on the same host, as well as the operating system itself. To improve the accuracy of anomaly detection in container systems, a technique for creating software systems for detecting anomalous behavior by assessing the state of container system elements is proposed. This technique is based on creating reference profiles of legitimate values of

container system element parameters and using the AE-LSTM neural network. Such parameters can include mount points, network interfaces, restrictions on the use of computing resources, and others. A model trained on a reference profile allows classifying the current state of a container as legitimate or anomalous based on calculating the data reconstruction error. If the reconstruction error exceeds a specified limit, an anomaly is registered. The accuracy of the presented ranges from 96.9% to 98.7% depending on the model and element of the container system.

Keywords: anomaly detection, container systems, container system elements, neural networks.

SOFTWARE PLATFORM FOR DETECTING FRAUD IN AUDIO RECORDINGS USING NERUAL NETWORK MODELS

E.S. Mytarin (*e.mytarin@ulstu.ru*)

V.S. Moshkin (*v.moshkin@ulstu.ru*)

A.Y. Zhuravlev (*zhuravlev.a@ulstu.ru*)

Ulyanovsk State Technical University, Ulyanovsk

The article presents a description of a hybrid methodology – a software platform for automated fraud detection in audio recordings of telephone conversations, based on neural network models and ontological analysis. The software platform implementing the proposed approach integrates deep learning methods such as speech recognition (Vosk, Whisper), text lemmatization, intonation analysis and comparison with an ontological database to assess the likelihood of fraudulent actions. The solution is aimed at application in the field of financial security and counteracting social engineering. The results of the experiments, in comparison with modern approaches, demonstrate the effectiveness of the platform in processing audio data, visualizing results and integrating with databases.

Keywords: fraud in audio recordings, neural network models, ontological analysis, speech recognition, machine learning, social engineering.

ESTIMATION OF FIRE AREAS BASED ON SPECTRAL INDICES AND MACHINE LEARNING METHODS

Yu.V. Pestova (*yupest@gmail.com*)

O.A. Nikolaichuk (*nikoly@icc.ru*)

Matrosov Institute for System Dynamics and Control Theory of Siberian Branch of Russian Academy of Sciences, Irkutsk

The paper presents the results of applying the Random Forest machine learning method to estimate burned woodland areas in the Kirensky Forestry of the Irkutsk region, using remote sensing data and the Google Earth Engine computing platform. To train the model, a preliminary set of spectral indices was selected, data for the study area

was prepared, and an algorithm for data processing was developed. The achieved accuracy of the developed model on the validation sample was 0.975. The results will be applied to predicting forest fire risks in the region.

Keywords: forest fires, burning, burned areas, spectral indices, machine learning methods, Random Forest.

MODELING OF IMPLICIT MEANING IN SPEECH

V.L. Stefanuk (*stefanuk@iitp.ru*) ^{A,B}

L.V. Savinitch (*savinitch@iitp.ru*) ^A

^A Kharkevich Institute for Information Transmission Problems
of the Russian Academy of Sciences, Moscow

^B People's Friendship University of Russia, Moscow

The paper is devoted to modeling the process of conveying implicit information, i.e. information containing the hidden meaning of an utterance. The language strategies used to convey, reveal and interpret the implicit meaning are demonstrated. Various prosodic means for conveying implicit information are described in detail: intonation, pitch, tempo, duration, volume, pause, etc., which are widely used in speech. They act as markers for revealing and interpreting hidden meaning in interactive communication.

The analyzed examples are accompanied by audio recordings and graphs displaying pitch fluctuations, sound intensity, modulation of tone.

Keywords: modeling, implicit, information, prosody.

DECENTRALIZED MULTI-CRITERIA ROUTING

N.P. Fomin (*heartmarshall@yandex.ru*) ^A

K.S. Yakovlev (*yakovlev@isa.ru*) ^{A,B}

^A Saint Petersburg State University, Saint Petersburg

^B Federal Research Center "Informatics and Control"
of the Russian Academy of Sciences, Moscow

The problem of decentralized multicriteria routing in computer networks is considered, where each node independently forms route solutions based on locally available information. Modern routing protocols are often limited by pathfinding algorithms that minimize the value of the scalar objective function, which can lead to suboptimal route selection in complex network scenarios where there are several equivalent quality criteria. The paper proposes two new decentralized routing methods based on modern multicriteria search algorithms. The first method, node behavior modeling, allows each node to build a local routing table based on the sending node, ensuring compliance with specified path cost constraints and determinism. The second, greedy method, decides on the next transition based only on the destination of the packet, offering a compromise

between computational complexity and strictness of restrictions. The effectiveness of both approaches has been experimentally confirmed and their comparative characteristics in terms of solution quality and resource intensity are presented.

Keywords: multi-criteria routing, Pareto optimality, pathfinding algorithms, computer networks.

INTERPRETABLE SUBSTITUTE MODELS WITH TREESHAP-ANALYSIS FOR DECISION SUPPORT SYSTEMS IN FIRE SAFETY

R.Sh. Khabibulin (*kh-r@yandex.ru*)

State Fire Service Academy of Emercom of Russia, Moscow

This paper presents a formalized approach to the creation and validation of machine learning substitute models for approximating Bayesian calculations in ranking fire protection systems. The proposed approach is based on the use of tree-like algorithms (*Random Forest*, *XGBoost*, *LightGBM*) in combination with the *TreeSHAP* method to ensure interpretability of management decisions. Using the example of oil and petroleum products storage facilities, an experimental study was conducted on a synthetic dataset of 2000 scenarios. Software has been developed that includes a comprehensive validation system with ranking correlation metrics and analysis of the impact on decision-making.

Keywords: TreeSHAP, fire safety, decision support systems, interpreted machine learning.

AN APPROACH TO FINDING THIRD-PARTY COMPONENTS FOR SOFTWARE SYSTEMS BASED ON MULTIMODAL DATA ANALYSIS

S.S. Shevaldov (*sshevaldov@yandex.ru*)

A.A. Filippov (*al.filippov@ulstu.ru*)

Ulyanovsk State Technical University, Ulyanovsk

The article explores the challenge of finding third-party components (such as dependencies and libraries) in external repositories during software development, considering the project's functional requirements and other features. The significance and complexity of this problem stem from the multimodal nature of component data. It is difficult for developers to interpret and evaluate the parameters of components, considering various project requirements and constraints. Each third-party component within a software system possesses multiple parameters that influence the developer's decision regarding its suitability for the project. A component search approach based on multimodal data analysis is proposed to address the tasks of finding appropriate components and enhancing the overall quality of the resulting software system. To validate the effectiveness of this approach, a prototype search system has been developed. The search system contains functions for collecting, processing, and analyzing multimodal component data to generate comprehensive assessments and ranging search results.

Keywords: multimodal data, components, software system, information retrieval, context.

АВТОРСКИЙ УКАЗАТЕЛЬ

Аббазов В.Р.	28	Кузьмин В.Р.	217
Адигвэ Э.О.	8	Лукашевич Н.В.	152
Алубин С.В.	17	Мамедов Т.Г.	217
Афанасьева Т.В.	250	Марковцева М.В.	318
Берхеев Э.К.	186	Массель А.Г.	217
Боргест Н.М.	8, 198	Массель Л.В.	217
Бучаев А.Я.	262	Мельник М.В.	361
Ветров А.Н.	271	Мельников А.В.	28, 74
Виноградов Г.П.	283	Мешков А.В.	115
Власов С.А.	198	Минин В.С.	129
Вохминцев А.В.	28	Мисник А.С.	122
Вохминцева Т.С.	28	Молодченков А.И.	228
Гаськова Д.А.	217	Мошкин В.С.	373
Грибова В.В.	207, 294	Мустафаев А.Г.	338
Григорьев А.А.	237	Мытарин Е.С.	373
Гришин И.Ю.	152	Наместников А.М.	327
Демидов Д.В.	40	Никитин А.И.	294
Добрынин Д.А.	52	Николайчук О.А.	384
Долгих А.А.	305	Пальчевский А.И.	228
Еремеев А.П.	62	Палюх Б.В.	271
Журавлев А.Ю.	373	Пестова Ю.В.	384
Згуральская Е.Н.	318	Петров В.С.	62
Игнатьев Н.А.	318	Пономарев А.В.	141
Исмагулов М.Е.	74	Пучков А.Ю.	129
Кабайлов Д.Д.	327	Романов А.А.	186
Кобзаренко Д.Н.	338	Романов М.А.	28
Коломойцева И.А.	85	Рыбина Г.В.	237
Комаров И.И.	262	Савинич Л.В.	396
Константинов А.В.	95	Садковский Ф.А.	152
Королева М.Н.	349	Садовников В.Е.	103
Котенко И.В.	103, 361	Саенко И.Б.	103
			455

Самсонович А.В.	305	Фомина М.В.	175
Стефанюк В.Л.	396	Хабаров Д.Л.	305
Тимченко В.А.	294	Хабибулин Р.Ш.	414
Уткин Л.В.	95	Хатер М.	28
Федяев О.И.	85	Шалфеева Е.А.	207
Филиппов А.А.	424	Швец Г.В.	175
Фомин И.С.	164	Шевалдов С.С.	424
Фомин Н.П.	403	Яковлев К.С.	403

СОДЕРЖАНИЕ

Секция 8. МАШИННОЕ ОБУЧЕНИЕ, НЕЙРОСЕТЕВЫЕ МЕТОДЫ, НЕЙРОИНФОРМАТИКА

Э.О. Адигвэ, Н.М. Боргест ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ПРИ КОНЦЕПТУАЛЬНОМ ПРОЕКТИРОВАНИИ САМОЛЕТА.....	8
С.В. Алубин ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ.....	17
А.В. Вохминцев, В.Р. Аббазов, М.А. Романов, Т.С. Вохминцева, М. Хатер, А.В. Мельников ПРИМЕНЕНИЕ МЕТОДОВ 2D СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ НА ОСНОВЕ АРХИТЕКТУРЫ ТРАНСФОРМЕРА В ЗАДАЧЕ ДИСТАНЦИОННОГО ИССЛЕДОВАНИЯ СТРУКТУРЫ АРХЕОЛОГИЧЕСКИХ ПАМЯТНИКОВ.....	28
Д.В. Демидов СИСТЕМА ПОДДЕРЖКИ СОЗДАНИЯ КОРПУСА РУКОПИСНОГО НАСЛЕДИЯ ДРЕВНЕЙ РУСИ	40
Д.А. Добрынин ПОИСК СХОДСТВА ВЫИГРЫШНЫХ СТРАТЕГИЙ В ЗАДАЧЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ	52
А.П. Еремеев, В.С. Петров МОДЕЛЬ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ С ПРИМЕНЕНИЕМ АДАПТИВНОЙ СЕТИ И НЕЧЕТКОГО ВЫВОДА	62
М.Е. Исмагулов, А.В. Мельников НЕЙРОСЕТЕВАЯ МОДЕЛЬ ТЕМПОРАЛЬНОГО ОБЪЕДИНЕНИЯ КАДРОВ ВИДЕОЛЕКЦИИ ДЛЯ РЕШЕНИЯ ЗАДАЧИ РЕКОНСТРУКЦИИ ИЗОБРАЖЕНИЯ	74
И.А. Коломойцева, О.И. Федяев АВТОМАТИЧЕСКОЕ ПОСТРОЕНИЕ ТЕЗАУРУСА: ОЦЕНКА МЕТОДОВ КЛАСТЕРИЗАЦИИ И ВЕКТОРНОГО ПРЕДСТАВЛЕНИЯ СЛОВ.....	85
А.В. Константинов, Л.В. Уткин ГЛУБОКИЙ ЛЕС ДЛЯ АНАЛИЗА ВЫЖИВАЕМОСТИ В РАМКАХ МНОГОВАРИАНТНОГО ОБУЧЕНИЯ	95
И.В. Котенко, И.Б. Саенко, В.Е. Садовников ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ КОМБИНИРОВАНИЯ МЕТОДОВ ЗАЩИТЫ СИСТЕМ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ ОТ ГРАДИЕНТНЫХ СОСТЯЗАТЕЛЬНЫХ АТАК	103

А.В. Мешков ПРИМЕНЕНИЕ МЕТОДОВ КЛАССИФИКАЦИИ И КЛАСТЕРИЗАЦИИ В ПРОЦЕССЕ РАССЛЕДОВАНИЯ ИНЦИДЕНТОВ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ.....	115
А.С. Мисник ПРИМЕНЕНИЕ ОБУЧЕНИЯ С ПОДКРЕПЛЕНИЕМ В МНОГОПОЛЬЗОВАТЕЛЬСКИХ СТРАТЕГИЧЕСКИХ ИГРАХ.....	122
А.Ю. Пучков, В.С. Минин ВЫЯВЛЕНИЕ АНОМАЛИЙ В ФИНАНСОВЫХ ВРЕМЕННЫХ РЯДАХ НА ОСНОВЕ МНОГОМАСШТАБНОГО АВТОКОДИРОВЩИКА.....	129
А.В. Пономарев МЕТОД УПРАВЛЕНИЯ ПОКРЫТИЕМ В ЗАДАЧЕ ОБУЧЕНИЯ С ДЕЛЕГИРОВАНИЕМ	141
Ф.А. Садковский, Н.В. Лукашевич, И.Ю. Гришин АВТОМАТИЧЕСКОЕ ПОПОЛНЕНИЕ ТАКСОНОМИЙ НА РУССКОМ ЯЗЫКЕ С ПОМОЩЬЮ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ	152
И.С. Фомин РАСПОЗНАВАНИЕ ИЗОБРАЖЕНИЙ ЛЮДЕЙ ПО ДАННЫМ С ИНФРАКРАСНОЙ И ЦВЕТНОЙ КАМЕРЫ С ИСПОЛЬЗОВАНИЕМ СЕГМЕНТНОЙ СПАЙКОВОЙ НЕЙРОННОЙ СЕТИ.....	164
М.В. Фомина, Г.В. Швец ИСПОЛЬЗОВАНИЕ МЕТОДОВ АНСАМБЛИРОВАНИЯ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ КЛАССИФИКАЦИИ ОБЪЕКТОВ, СОДЕРЖАЩИХ ШУМ	175

Секция 9. ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА И ТЕХНОЛОГИИ ПРОЕКТИРОВАНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Э.К. Берхеев, А.А. Романов ОПТИМИЗАЦИЯ ГЕНЕРАТИВНЫХ СИСТЕМ С ПОМОЩЬЮ UNIKERNEL	186
Н.М. Боргест, С.А. Власов РАЗРАБОТКА ВИРТУАЛЬНОГО АНГАРА ДЛЯ ПРОЕКТИРОВАНИЯ И ИЗУЧЕНИЯ АВИАЦИОННЫХ КОНСТРУКЦИЙ.....	198
В.В. Грибова, Е.А. Шалфеева ОНТОЛОГИЧЕСКИЙ ПАТТЕРН ИНТЕЛЛЕКТУАЛЬНОЙ ПОДДЕРЖКИ РЕАБИЛИТАЦИОННОЙ КОМАНДЫ.....	207

Л.В. Массель, А.Г. Массель, В.Р. Кузьмин, Т.Г. Мамедов, Д.А. Гаськова СЕРВИСЫ ДЛЯ ОБРАБОТКИ ДАННЫХ И ЗНАНИЙ В ЭКОСИСТЕМЕ ЗНАНИЙ В ЭНЕРГЕТИКЕ	217
А.И. Пальчевский, А.И. Молодченков ТЕНЗОРНОЕ ПРЕДСТАВЛЕНИЕ НЕОДНОРОДНЫХ СЕМАНТИЧЕСКИХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧИ СЛИЯНИЯ БАЗ ЗНАНИЙ	228
Г.В. Рыбина, А.А. Григорьев ОНТОЛОГИЧЕСКАЯ ПРОГРАММНАЯ СРЕДА ИНТЕЛЛЕКТУАЛЬНОГО ОБУЧЕНИЯ НА ОСНОВЕ ИНТЕГРИРОВАННЫХ ЭКСПЕРТНЫХ СИСТЕМ: МЕТОДЫ ПОСТРОЕНИЯ НА ПРИНЦИПАХ СИТУАЦИОННОГО УПРАВЛЕНИЯ	237
Секция 10. ПРИКЛАДНЫЕ ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ	
Т.В. Афанасьева КОНЦЕПЦИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ АНАЛИТИКИ ЗДОРОВЬЯ ПАЦИЕНТОВ С ХРОНИЧЕСКИМИ ЗАБОЛЕВАНИЯМИ С ИСПОЛЬЗОВАНИЕМ НЕЧЕТКИХ МОДЕЛЕЙ.....	250
А.Я. Бучаев, И.И. Комаров СНИЖЕНИЕ СЛОЖНОСТИ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ НА ОСНОВЕ МЕТОДА АВТОМАТИЧЕСКОЙ ЛОКАЛИЗАЦИИ АНОМАЛИЙ В МНОГОМЕРНЫХ ПРОСТРАНСТВАХ	262
А.Н. Ветров, Б.В. Палюх ОБЪЕДИНЕНИЕ МУЛЬТИСЕНСОРНЫХ ДАННЫХ МНОГОСТАДИЙНОГО ПРОИЗВОДСТВА ДЛЯ ПРОГНОЗИРОВАНИЯ ОТКАЗОВ ТЕХНОЛОГИЧЕСКОГО ОБОРУДОВАНИЯ	271
Г.П. Виноградов ИНТЕРНЕТ БОЕВЫХ ВЕЩЕЙ В ВООРУЖЁННЫХ СИЛАХ.....	283
В.В. Грибова, А.И. Никитин, В.А. Тимченко ГИБРИДНЫЙ МЕТОД ИЗВЛЕЧЕНИЯ ПРЕЦЕДЕНТОВ ДЛЯ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ЛАЗЕРНОМ АДДИТИВНОМ ПРОИЗВОДСТВЕ	294
А.А. Долгих, Д.Л. Хабаров, А.В. Самсонович ВИРТУАЛЬНЫЙ ТЬЮТОР НА ОСНОВЕ ИНТЕГРАЦИИ БОЛЬШОЙ ЯЗЫКОВОЙ МОДЕЛИ С КОГНИТИВНОЙ АРХИТЕКТУРОЙ	305

Н.А. Игнатъев, Е.Н. Згуральская, М.В. Марковцева ПРОГНОСТИЧЕСКАЯ СИСТЕМА МОНИТОРИНГА БОЛЬНЫХ ХРОНИЧЕСКИМ ЛИМФОЛЕЙКОЗОМ.....	318
Д.Д. Кабайлов, А.М. Наместников МАРКИРОВКА ВИДИМОСТИ ЭЛЕМЕНТОВ WIM-МОДЕЛЕЙ С ИСПОЛЬЗОВАНИЕМ ГРАФОВОЙ НЕЙРОННОЙ СЕТИ	327
Д.Н. Кобзаренко, А.Г. Мустафаев АЛГОРИТМЫ ИДЕНТИФИКАЦИИ КЛАССА В ЗАДАЧЕ РАСПОЗНАВАНИЯ КЛАВИАТУРНОГО ПОЧЕРКА С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ	338
М.Н. Королева К ВОПРОСУ О ПСИХОФИЗИОЛОГИЧЕСКИХ ХАРАКТЕРИСТИКАХ В ЧЕЛОВЕКО-МАШИННЫХ СИСТЕМАХ	349
И.В. Котенко, М.В. Мельник ОБНАРУЖЕНИЕ АНОМАЛИЙ В КОНТЕЙНЕРНЫХ СИСТЕМАХ НА ОСНОВЕ ПРОФИЛИРОВАНИЯ И НЕЙРОННОЙ СЕТИ AE-LSTM...	361
Е.С. Мытарин, В.С. Мошкин, А.Ю. Журавлев ПРОГРАММНАЯ ПЛАТФОРМА ОБНАРУЖЕНИЯ МОШЕННИЧЕСТВА В АУДИОЗАПИСЯХ С ПОМОЩЬЮ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ	373
Ю.В. Пестова, О.А. Николайчук ОЦЕНКА ПЛОЩАДЕЙ ВОЗГОРАНИЯ НА ОСНОВЕ СПЕКТРАЛЬНЫХ ИНДЕКСОВ И МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ	384
В.Л. Стефанюк, Л.В. Савинич МОДЕЛИРОВАНИЕ ПЕРЕДАЧИ ИМПЛИЦИТНОГО СМЫСЛА В УСТНОЙ РЕЧИ.....	396
Н.П. Фомин, К.С. Яковлев ДЕЦЕНТРАЛИЗОВАННАЯ МНОГОКРИТЕРИАЛЬНАЯ МАРШРУТИЗАЦИЯ.....	403
Р.Ш. Хабибуллин ИНТЕРПРЕТИРУЕМЫЕ МОДЕЛИ-ЗАМЕСТИТЕЛИ С TREESHAP-АНАЛИЗОМ ДЛЯ СИСТЕМ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В ОБЛАСТИ ПОЖАРНОЙ БЕЗОПАСНОСТИ.....	414
С.С. Шевалдов, А.А. Филиппов ПОДХОД К ПОИСКУ КОМПОНЕНТОВ ДЛЯ ПРОГРАММНЫХ СИСТЕМ НА ОСНОВЕ АНАЛИЗА МУЛЬТИМОДАЛЬНЫХ ДАННЫХ	424
Abstracts of Sectional Reports	435
АВТОРСКИЙ УКАЗАТЕЛЬ	455

Научное издание

Двадцать вторая Национальная конференция
по искусственному интеллекту с международным участием
КИИ-2025
Труды конференции в 3-х томах
Том 3

Подписано в печать 23.09.2025 г.
Формат 60x84¹/₁₆. Тираж 300 экз. Усл. печ. л. 26,8.

Издательство СПб ФИЦ РАН

ISBN 978-5-6052274-4-1

