МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Донецкий национальный технический университет»

Факультет интеллектуальных систем и программирования
Кафедра «Прикладная математика и искусственный интеллект»

СБОРНИК МАТЕРИАЛОВ

(СТУДЕНЧЕСКОЙ СЕКЦИИ)

IX ВСЕРОССИЙСКОЙ
НАУЧНО-ТЕХНИЧЕСКОЙ КОНФЕРЕНЦИИ
с международным участием

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ И НАУЧНЫХ ИССЛЕДОВАНИЯХ (СИТОНИ-2025)

20 ноября 2025 г.

Донецк
2025

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Донецкий национальный технический университет»

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ И НАУЧНЫХ ИССЛЕДОВАНИЯХ (СИТОНИ-2025 – студенческая секция)

IX Всероссийская научно-техническая конференция
с международным участием

Сборник материалов
студенческая секция

20 ноября 2025 г.

Донецк
2025

УДК 004:378+001(063)
ББК 16.2:74+72
С56

Рекомендовано к изданию
Советом факультета ИСП
ФГБОУ ВО «ДонНТУ»
(протокол № 12 от 19.12.2025 г.)

Ответственный редактор – Владимир Николаевич Павлыш

Редакционная коллегия: А. Я. Аноприенко (пред.), К. Н. Ефименко, Е. В. Прокопенко

Организационный комитет конференции:
Председатель:
Аноприенко А. Я. – канд. тех. наук, доцент, ректор ФГБОУ ВО «ДонНТУ»
Заместители председателя:
Павлыш В. Н. – д-р техн. наук, профессор, заведующий кафедрой «Прикладная математика и искусственный интеллект» ФГБОУ ВО «ДонНТУ»;
Кожухов И. Б. – д-р физ.-мат.наук, профессор кафедры «Высшая математика №1», НИУ МИЭТ (г. Москва);
Иванова С. Б. – директор ФГБНУ «ИПИИ» (г. Донецк).
Члены оргкомитета:
Ефименко К. Н. – канд. тех. наук, доцент кафедры «Прикладная математика и искусственный интеллект» ФГБОУ ВО «ДонНТУ»;
Сухорукова Е. О. – канд. физ.-мат. наук, ст. преподаватель кафедры «Прикладная математика и искусственный интеллект» ФГБОУ ВО «ДонНТУ»;
Савицкая И. В. – ассистент кафедры «Прикладная математика и искусственный интеллект» ФГБОУ ВО «ДонНТУ».
Ответственный секретарь конференции:
Прокопенко Е. В. – канд. тех. наук, доцент, кафедры «Прикладная математика и искусственный интеллект» ФГБОУ ВО «ДонНТУ».

С56

Современные информационные технологии в образовании и научных исследованиях (СИТОНИ-2025 – студенческая секция) : IX Всерос. науч.-техн. конф. с межд. участием, г. Донецк, студ. секция 20 нояб. 2025 г. / Ред. кол.: А. Я. Аноприенко (пред.), К. Н. Ефименко, Е. В. Прокопенко ; отв. ред. В. Н. Павлыш. – Донецк : ДонНТУ, 2025. – Систем. требования : просмотрщик PDF файлов. – Загл. с титул. экрана.

Материалы сборника освещают современные информационные технологии и научные исследования в сфере образования, научные достижения ученых, аспирантов, магистрантов и обучающихся высших учебных заведений из ДНР, России и Беларуссии, а также обсуждается повышение эффективности использования научного потенциала вузов, научных организаций и предприятий.

Издание предназначено для научных работников, преподавателей, аспирантов, магистрантов и обучающихся образовательных учреждений высшего образования.

Тексты докладов печатаются в авторской редакции.

УДК 004:378+001(063)
ББК 16.2:74+72

© ФГБОУ ВО «ДонНТУ», 2025 г

Организатор конференции:

ФГБОУ ВО «Донецкий национальный технический университет», НОИ «КНТ», факультет интеллектуальных систем и программирования, кафедра прикладной математики и искусственного интеллекта (г. Донецк).

Партнёры конференции:

ФГБНУ «Институт проблем искусственного интеллекта» (г. Донецк);
ФГАОУ ВО «Национальный исследовательский университет «Московский институт электронной техники» (г. Москва, г. Зеленоград);
ФГБОУ ВО «Балтийский государственный технический университет «ВОЕНМЕХ» им. Д.Ф. Устинова» (г. Санкт-Петербург);
УО «Витебский государственный университет имени П.М. Машерова» (г. Витебск, РБ);
ФГАОУ ВО «Южный федеральный университет» (г. Ростов-на-Дону, г. Таганрог);
ФГБОУ ВО «Волгоградский государственный технический университет» (г. Волгоград);
ФГБОУ ВО «Донецкий государственный университет» (г. Донецк);
Международный рецензируемый научно-теоретический журнал ФГБНУ «ИПИИ «Проблемы искусственного интеллекта» (г. Донецк);
Научный журнал ДонНТУ «Информатика и кибернетика» (г. Донецк).

Программный комитет конференции:

Павлыш Э. В. – к.э.н., доцент, зав. кафедрой информационных технологий и управления бизнесом УО «Витебский государственный университет имени П.М. Машерова» (г. Витебск, РБ);
Николаенко Д. В. – к.т.н., доц., декан факультета интеллектуальных систем и программирования ФГБОУ ВО «ДонНТУ»;
Мальчева Р. В. – к.т.н., доц., заместитель директора по науке ИКНТ ФГБОУ ВО «ДонНТУ».
Анохина И. Ю. – к.т.н., доц., доцент кафедры прикладной математики и искусственного интеллекта ФГБОУ ВО «ДонНТУ»;
Перинская Е. В. – к.т.н., доц., доцент кафедры прикладной математики и искусственного интеллекта ФГБОУ ВО «ДонНТУ»;
Тарабаева И. В.– к.т.н., доц., доцент кафедры прикладной математики и искусственного интеллекта ФГБОУ ВО «ДонНТУ».

СОДЕРЖАНИЕ

СЕКЦИЯ №1. ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА

Волга И.С., Перинская Е.В. Разработка математических моделей для предсказания течения болезни COVID 19 на основе клинических данных 8
Некрасова Д.Р., Лазебная Л.А. Обзор современных методов распознавания движений по видеоданным. 13

СЕКЦИЯ №2. ТЕХНОЛОГИИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

Агапов И. С. Перинская Е.В. Методы применения и возможного улучшения искусственного интеллекта в компьютерных играх 19
Алымов Д.А., Боднар А.В. Исследование архитектурных механизмов надёжности и масштабирования в распределённых RPA-системах 23
Белый А.А., Григорьев. А.В. Анализ архитектуры систем автоматической генерации веб-интерфейсов. 28
Бубнов Г.В., Григорьев А.В. Синтез модели принятия решений для оптимизации веб-ресурса на базе комплексной веб-аналитики 32
Будур А.Р., Зори С.А. Разработка и сравнение моделей машинного обучения для прогнозирования оттока клиентов телекоммуникационной компании 37
Винокуров К.А., Григорьев А.В. Использование предобученных моделей машинного обучения в интеллектуальных системах на основе формата ONNX 44
Гончар Д.Ю. Применение технологий обработки естественного языка для автоматизации проверки студенческих работ 48
Дегтярёв Т.Р. Новая реальность BI: обзор и сравнение российских платформ после ухода зарубежных аналогов. 52
Добренко И.С., Едемская Е.Н. Система классификации рентгеновских изображений на основе глубоких сверточных нейронных сетей 57
Дяченко Д.А., Лазебная Л.А. Исследование методов навигации в закрытых помещениях. 62
Зайцева В.А., Губенко Н.Е. Моделирование цифрового двойника университетского кампуса для поддержки учебного процесса 67
Зосимов В.В., Григорьев А.В., Филипишин Д.А. Проблемно-ориентированные редакторы онтологий для синтеза баз данных 72
Исаенко В.В., Григорьев А.В. Исследование подходов к автоматизации подбора персонала на основе технологий искусственного интеллекта 76
Кириленко Н.А., Прокопенко Е.В. Интеллектуальные системы адаптивного геймплея в компьютерных играх 80
Ковальчук А.Р., Прокопенко Е.В. Интеллектуальная платформа поддержки принятия решений в HR 85
Ковальчук А.Р., Прокопенко Е.В. HR-аналитика как инструмент образовательных программ и научных исследований. 89
Комарницкий В.Р. Исследование системы управления отделом кадров предприятия на основе системного анализа 93
Кукуруза Я.А., Тарабаева И.В. Применение методов обучения нейронных сетей в дискретных симуляционных средах 98
Куштей И.В., Анохина И.Ю. Нейросетевой интеллект для игровых персонажей 103
Лазуренко Е. А., Коломойцева И. А. Сравнительный анализ эффективности резюмирования новостных статей при помощи общедоступных LLM 107
Лободзинский В.О. Анализ современных решений для разработки виртуальных помощников с ИИ 111

Макаренко Е.А., Лазебная Л.А. Типы машинного обучения: Сравнительный анализ и области эффективного применения. 115
Панков Е.А., Григорьев А.В. Мультимодальная оценка свежести пищевых продуктов по изображению с верификацией по маркировке срока годности: обзорно-методическое исследование с применением нейронных сетей....120
Первенко Д.А. Оптимизация пользовательского опыта в веб-приложениях сервисного типа. 125
Пинчук Т.Ю., Ясинская А.А., Филипишин Д.А. Мультимодальные умные весы для самообслуживания в ритейле: edge–cloud архитектура, open-set распознавание и MLOps ...129
Рябцев М.К., Зори С.А. Риски и проблемы внедрения автоматизированного тестирования в небольших проектах....134
Сидорика М.Д., Лазебная Л.А. Оптимизация нейросетевых моделей естественно-языкового понимания (NLU) в условиях ограниченных вычислительных ресурсов. 138
Стёпушкина П.К.. Современные нейросетевые подходы к детекции объектов. 144
Стома Д.С., Лазебная Л.А. Реализация интеллектуального художника изобразительного искусства на базе генетических алгоритмов 149
Хаунов Д.И., Лазебная Л.А. Интеллектуальный подход к локальному поиску, семантической обработке и каталогизации текстовых файлов на основе NLP....154
Ястребов А.Р., Рычка О.В. Параллельная обработка транзакций в системах интернет-торговли с использованием современных архитектур баз данных....158
Буловацкий Е.А., Прокопенко Е.В. Разработка пользовательских интерфейсов с использованием нейросетей. 163

СЕКЦИЯ №3. КОМПЬЮТЕРНАЯ ИНЖЕНЕРИЯ

Дворник Н.Н.., Чередникова О.Ю. Разработка MCP-сервера для автоматизации взаимодействия с графическим интерфейсом Windows....169
Лесовой Н.М. Применение методов компьютерного зрения для автоматизированного предварительного анализа видеоконтента....174

СЕКЦИЯ №4. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ И НАУЧНЫХ ИССЛЕДОВАНИЯХ

Андросов В.А., Лазебная Л.А. Риски и вызовы генеративного искусственного интеллекта в высшем образовании....180
Белоконь Л.О., Федяев О.И. Применение архитектуры трансформеров для автоматического анализа резюме 184
Боднар А.В., Нестеренко А.Р. Аналитические информационные системы для управления цифровыми маркетинговыми коммуникациями в образовательной и научной деятельности вуза....191
Голосиенко А.И., Зори С.А. Системы управления личными задачами в образовательной среде: анализ существующих решений и возможности внедрения 196
Коротеев С.С. Применение средств анализа пользовательской активности при проектировании обучающих систем 201
Котенко А.С., Лазебная Л.А. Применение методов многокритериальной оптимизации для формирования сбалансированных групп при проектном обучении 206
Навроцкий Я. Д. Геймификация образовательного процесса: использование конструкторов игр в школьном курсе предметов 211
Тарабаева И.В., Нужная Н.А. Автоматизированное тестирование API с использованием Postman 215
Бондаренко В.В., Рычка О.В. Анализ изображений и обнаружение рака с помощью ИИ. 220
Зозуля Н.М., Федяев О.И. Экспериментальная оценка качества генерации ответов интеллектуального почтового ассистента на основе критерия адекватности....226

СЕКЦИЯ №1
ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ МАТЕМАТИКА

УДК 501

РАЗРАБОТКА МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ДЛЯ ПРЕДСКАЗАНИЯ ТЕЧЕНИЯ БОЛЕЗНИ COVID 19 НА ОСНОВЕ КЛИНИЧЕСКИХ ДАННЫХ.

Волга И.С., Перинская Е.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: volgair7@yandex.ru

Волга И.С., Перинская Е.В. Разработка математических моделей для предсказания течения болезни COVID 19 на основе клинических данных. *Пандемия COVID-19 продемонстрировала необходимость точных методов прогнозирования распространения инфекций. Математическое моделирование позволяет оценить динамику эпидемий, эффективность мер контроля и планировать ресурсу здравоохранения. В работе рассмотрены классические (SEIR, SIR) и современные подходы (АОМ, комбинированные модели). Однако вопросы для учета пространственной неоднородности и стохастических факторов остаются открытыми.*

Ключевые слова: математическое моделирование, методы прогнозирования, динамика эпидемий, пространственная неоднородность, стохастические факторы

Введение

В работе рассматриваются математические модели, которые используются для анализа и прогнозирования развития COVID-19 и основаны на трех подходах: анализ временных рядов, уравнения математической физики и имитационные модели, включая различные их комбинации и обобщения. В силу новизны и сложности заболевания COVID-19 параметры большинства математических моделей, как правильно, неизвестны, и это приводит к необходимости рассматривать обратные задачи. Основные проблемы моделирования распространения COVID-19:

1. Данные для решения обратной задачи является неполными и зашумленными, а также представляют собой большие данные (ежедневные сводки о заболевших, заразившихся, вакцинированных и т.д.).
2. Параметры меняются со временем.
3. Процесс распространения COVID-19 существенно изменяется при введении или отмене ограничительных мер (ношение масок, социальная дистанция, перевод на удаленный рабочий режим, закрытие школ, предприятий, районов и городов), появление каждого нового штамма.

Основная часть

В работе будут рассмотрены регрессионные и сетевые модели, методы фильтрации и их взаимосвязи, на основе анализе наиболее достоверных статистических данных – количество ежедневных ПЦР-тестов T(t) и индекс самоизоляции a(t). Время t во всех моделях измеряется в днях.

Особенностью функций T(t) и a(t) является повторяемость подъемов и спадов по времени с изменяющейся амплитудой. Например, T(t) возрастает по вторникам и ослабевает к понедельнику, а a(t) ослабевает в период выходных, праздников и отпусков.

Методы математической статистики и машинного обучения помогают обрабатывать, анализировать эпидемиологические данные и проводить краткосрочное прогнозирование поведение T(t) и a(t) при отсутствии резких изменений ситуации (введение ограничительных мер, мутации вируса). Приведены регрессионные, сетевые модели, а также методы фильтрации и их взаимосвязи.

Регрессионные модели

Регрессионные модели подразделяются на неадаптивные модели, для оценки параметров используются все имеющиеся данные, и адаптивные, значения параметров которых рассчитываются на основе скользящего окна наблюдений.

Вид регрессионной зависимости для неадаптивной модели выбирается исходя из свойств анализируемого временного ряда. В отдельных случаях можно применять полиномиальные или степенные функции, но чаще модель должна учитывать сезонный характер заболеваемости (например, модель Серфлинга). Неадаптивные регрессионные модели учитывают всю предысторию заболеваемости, но игнорируют локальные колебания эпидемических показателей, поэтому появление нового штамма вируса и введение ограничительных мер в регионе снижают значимость данных предыдущего периода.

Адаптивные регрессионные модели используют ограниченный отрезок временного ряда и более чувствительны к изменению ситуации. Важную роль использования адаптивных моделей играет ширина скользящего окна – число последних наблюдений, на основе которых оценивают параметры модели.

В регрессионных моделях предполагается, что невязка модели – независимая случайная величина, имеющая нормальный закон распределения с нулевым математическим ожиданием и постоянной дисперсией. Одной из проблемы прогнозирования является наличие существенной автокорреляции невязок. Регрессионная модель может быть дополнена, а прогноз – уточнен, например, с помощью авторегрессионных моделей. Диаграмма моделей прогнозирования временных рядов представлена на рисунке 1.

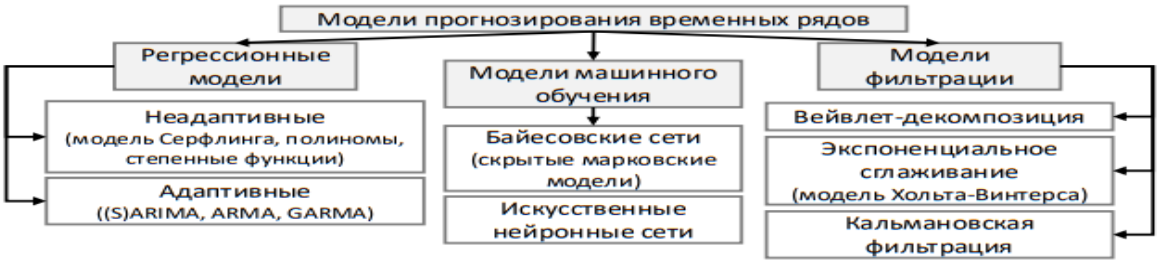

Рисунок 1 – Диаграмма моделей прогнозирования временных рядов

Авторегрессионная модель прогнозирования временного ряда

Для построения прогнозов сезонных рядов, например, количества проведённых ПЦР-тестов в регионе, использовалась авторегрессионная модель SARIMA, являющейся модификацией модели ARIMA, которая описывает одномерные временные ряда с сезонной компонентной. ARIMA является расширением моделей тип ARMA для нестационарных временных рядов, которые можно сделать стационарными взятием разностей некоторого порядка от исходного временного ряда (так называемые интегрированные или разностно-стационарные временные ряды) [1].

Модель ARIMA (*p, d, q)* для нестационарного временного ряда *T(n)* имеет вид:

$$\Delta^d T(n) = c + \sum_{i+1}^{p} a_i \Delta^d T(n-i) + \sum_{j=1}^{p} b_j \in (n-j) + \in (n) \qquad (1)$$

Здесь $\in (n)$ – стационарный временной ряд белого шума, c, a_i, b_j – параметры модели, Δ^d- оператор разности временного ряда порядка d, гарантирующий стационарность ряда (последовательное взятие d раз разностей первого порядка – сначала от временного ряда, затем от полученных разностей первого порядка – сначала от временного ряда, затем от полученных разностей первого порядка, затем от второго порядка и т.д.)

Обычно при построении модели ARIMA порядок разностей ограничивается числом d=2.

Алгоритм прогнозирования временного ряда следующий:

1. Применяем преобразование Бокса-Кокса для уменьшения дисперсии.
2. Вычисляем сезонную разность (сдвиг на 7 дней) первого порядка.
3. Вычисляем вторую разность (сдвиг на 1 день) ряда, полученного в пункте 1.
4. Проверяем стационарность ряда критерием Дики-Фуллера.
5. Передаем соответствующею проделанным действиям параметры в модель ARIMA (*p,d,q*) и подбираем остальные на основе минимизации информационного критерия Акаике.
6. Полученная модель с настроенным гиперпараметрами используется для дальнейшего прогнозирования. Полученный результат подвергается обратному преобразованию Бокса-Кокса [2].

Модели на основе машинного обучения

Машинное обучение является мощным инструментом для поиска взаимосвязи между входными и выходными данными в случаях, когда аналитические исследование затруднительно. Применение эвристических подходов для раннего обнаружения эпидемиологических рисков в некоторых случаях позволяет улучшить качества прогнозирования.

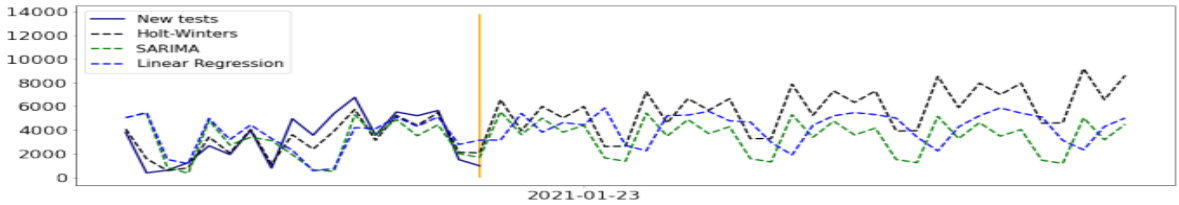

Рисунок 2 – Результаты предсказания временного ряда ежедневно проводимых ПЦР-тестов T(t) в Новосибирской области (сплошная синяя линия) на месяц вперед с 18.01.2021 (вертикальная оранжевая линия) моделями Хольта-Винтерса (черная пунктирная линия), линейной регрессии (синяя пунктирная линия) и SARIMA (зеленая пунктирная линия)

Одним из представителей моделей машинного обучения являются динамические байесовские сети – направленный граф, вершины которого соответствуют переменным модули, а ребра – вероятностным зависимостям между ними, которые заданы определенными законами распределения. После обучения как на большом, так и на малом количестве исходных данных байесовские сети позволяют оценить вероятность наступления некоторого события при наблюдаемой последовательности явления [3]. Для прогнозирования заболеваемости используется простая форма скрытых марковских моделей, основной идеей которых является сопоставление каждый случайной величины Y_t (например, количества выявленных, госпитализированных случаев с COVID-19) с ненаблюдаемой случайной величиной S_t (например, общее количества инфицированных индивидуумов), определяющей условное распределение. Таким образом, величины Y_t зависит только от значения скрытой переменной S_t в момент времени t, а последовательность S_t обладает марковским свойством, то есть величина S_t зависит только от S_{t-1} (рис. 3).

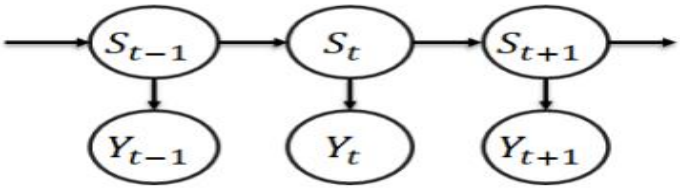

Рисунок 1.3 – Схема зависимостей в скрытой марковской модели

Искусственные нейронные сети (ИНС) представляют собой направленный взвешенный граф, вершины которого моделируют функционирование биологических нейронов (рис. 1.4). Обучение ИНС заключается в вычислении коэффициентов связей между вершинами,

определяющих силу сходящих сигналов, и выполняется на основе эмпирических данных: статистики заболеваемости и при наличии, значений факторов, ее предопределяющих. Для корректного обучения ИНС необходим большой объем исторических данных [4].

Рисунок 4 – Схема зависимостей в ИНС с одним скрытым слоем

Модели на основе фильтрации

Любые временные ряды заболеваемости можно рассматривать как случайные процесс, состоящий из сигнала, отражающего реальную эпидемическую обстановку, и высокочастотного шума. Фильтрация шума позволяет уточнить прогноз и может выполняться как в ходе предварительной обработки исходных данных, так и в составе самого алгоритма прогнозирования.

Одним из подходов является вейвлет-декомпозиция, в которой временной ряд представляется с помощью вейвлет-функций. Однако, такой поход используется совместно с другими моделями.

Одной из таких моделей является экспоненциальное сглаживание, представляющее собой частный случай взвешенного скользящего среднего, а именного значение заболеваемости y_t в момент времени t описывается взвешенной суммой последних наблюдений: $l_t = by_t(1+b)y_{t-1}$. Здесь $b \in (0,1)$ – коэффициент сглаживания, который обеспечивает уменьшение веса по мере старения данных, которое может рассматриваться как отражение естественного процесса обучения. Такой метод построения модели не подходит для радов, в поведении которых присутствуют отчетливый тренд или сезонность. Для этих целей используются обобщенные модели, например, сезонная модель Хольта-Вентерса. Результаты прогнозирования ряда ежедневных ПЦР-тестов в Новосибирской области в рамках данной модели представлены на рисунок 5.

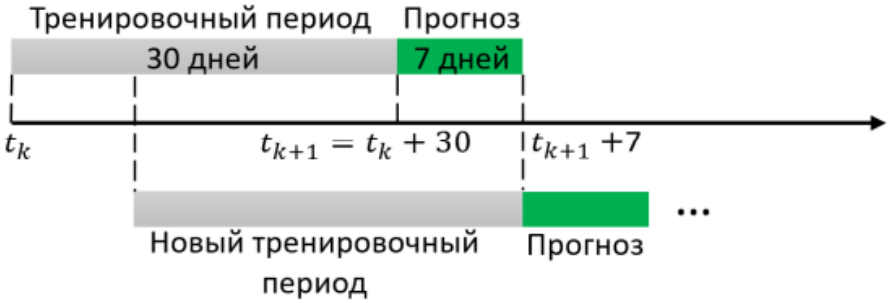

Рисунок 5 – Результаты прогнозирования ряда ежедневных ПЦР-тестов в Новосибирской области

Любые эпидемические процессы можно описать следующей системой разностных уравнений:

$$x_t = Ax_{t-1} + w_t,$$
$$y_t = Hx_t + Df_t + v_t \quad (2)$$

где x_t – вектор переменных состояний системы в момент времени t, y_t – вектор наблюдений, f_t – вектор значений внешних факторов, w_t и v_t – белый шум. Матрицы параметров A, H, D определяют модель эпидемического процесса и выбираются исходя из решаемой задачи – краткосрочного или долгосрочного прогнозирования.

Такая форма записи позволяет предложить обобщенные модели распространения заболеваемости, в частности, модели на основе калмановской фильтрации.

Выводы

Результаты, полученные на основе временных рядов, использованы нами в дифференциальных и агентных моделях. Прогнозирование ежеденых ПЦР-тестов и индекса самоизоляции позволяет строить сценарии развития COVID-19 в регионе в зависимости от введения ограничительных мер, а именно количество ожидаемых выявленных, умерших, госпитализированных, критических случаев и индекса репродукции вируса.

Литература

1. Martelloni G., Martelloni G. Analysis of the evolution of the Sars-Cov-2 in Italy, the role of the asymptomatics and the success of Logistic model // Chaos, Solitons and Fractals 140 (2020) 110150

2. Wang P., Zheng X., Li J., Zhu B. Prediction of epidemic trends in COVID19 with logistic model and machine learning technics // Chaos, Solitons and Fractals. 2020. 139, 110058.

3. Guo J. Theoretical Epidemic Laws Based on Data of COVID-19 Pandemic //medRxiv. – 2020. 5. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics // Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character 1927; 115(772):700–21.

4. Куркина Е.С., Кольцова Е.М. Математическое моделирование распространения волн эпидемии коронавируса COVID-19 в разных странах мира // Прикладная математика и информатика № 66. – М.: Изд-во факультета ВМК МГУ, 2021. С.41-46.

Volga I. S. Perinskaya E. V. Development of mathematical models for predicting the course of covid 19 disease based on clinical data. The COVID-19. *Pandemic has demonstrated the need for accurate methods for predicting the spread of infections. Mathematical modeling allows us to assess the dynamics of epidemics, the effectiveness of control measures, and plan healthcare resources. This paper discusses classical (SEIR, SIR) and modern approaches (AOM, combined models). However, the issues of accounting for spatial heterogeneity and stochastic factors remain open.*

Keywords: *mathematical modeling, forecasting methods, epidemic dynamics, spatial heterogeneity, stochastic factors*

УДК 004.932.2

ОБЗОР СОВРЕМЕННЫХ МЕТОДОВ РАСПОЗНАВАНИЯ ДВИЖЕНИЙ ПО ВИДЕОДАННЫМ

Некрасова Д.Р., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: nekrggggh@yandex.ru, l_lazebnay@mail.ru

Некрасова Д.Р., Лазебная Л.А. Обзор современных методов распознавания движений по видеоданным. *В статье рассмотрены современные методы распознавания человеческой деятельности (HAR), основанные на компьютерном зрении и глубоких нейронных сетях. Кратко описаны принципы работы 3D-CNN, RNN/LSTM, Transformer, GCN, гибридные и мультимодальные модели. Отмечено, что интеграция различных подходов повышает точность и устойчивость систем распознавания действий.*
Ключевые слова: *human action recognition (HAR), convolutional neural networks (CNN), computer vision, deep learning (DL), neural networks, recurrent neural networks (RNN, LSTM).*

Введение

Распознавание человеческой деятельности (РЧД) состоит из цепочки распознавания определенных действий человека с помощью сенсорных наблюдений и некоторых методов распознавания движений. Термин человеческая деятельность (ЧД) можно трактовать, как движение (-я) одной или нескольких частей тела человека. При этом распознавание ЧД должно определять одну и ту же деятельность разных людей одинаково. Системы, которые распознают заданный набор человеческой деятельности называются системами HAR (Human action recognition от англ. распознавание ЧД).

Системы HAR (РЧД) используют два основные направления технологий взаимодействия человека и компьютера для своей работы: контактные и бесконтактные методы. Контактные системы требуют физического взаимодействия пользователя с машиной или устройством сбора команд. Эти методы также напрямую зависят от характера данных, поступающих из различных сенсоров и источников, например, акселерометров, мультисенсорных экранов, датчиков, установленных на теле, сенсорных перчаток, прочее. Тем не менее, от контактных систем все больше и больше отказываются, потому что физический контакт требует определенных навыков и сложного оборудования, которые делают их доступными только для опытных пользователей.

Более удобной системой взаимодействия человека и компьютера являются бесконтактные методы или методы, основанные на компьютерном зрении. В отличии от контактных систем, системы на основе компьютерного зрения не требуют ношения никаких сенсоров или иных датчиков. Всю работу по распознаванию ЧД ведёт камера как источник данных и программа, которая распознает ЧД по исходным данным. В дальнейшем будем рассматривать именно бесконтактные методы распознавания движений [1].

РЧД в течение последних двух десятилетий было очень активной темой исследований в области компьютерного зрения и искусственного интеллекта (ИИ), которая фокусируется на автоматизированном анализе и понимании человеческих действий, и распознавании деятельности человека на основе движений и поз всего тела.

РЧД играет важную роль в различных областях и направлениях, таких как здравоохранение, удаленный видеонаблюдение, интеллектуальные человеко-машинные интерфейсы, развлечения, хранение и извлечение видео [2], мониторинг и оценку процессов в промышленности, управление машинами и устройствами, обнаружение мошенничества [1], а

также взаимодействие человека с компьютером [2].

Основными видами методов РЧД являются:

• методы на основе необработанных кадров (RGB): использование необработанных видеокадров в качестве входных данных для глубоких нейронных сетей;

• методы, основанные на позах: использование методов определения позы человека для извлечения данных скелета и дальнейшей классификации действий;

• гибридные методы: комбинируйте подходы, основанные на человеческих позах и необработанных кадрах, для повышения производительности.

Методы HAR варьируются от традиционных подходов с ручным выделением признаков до современных методов глубокого обучения.

Классические алгоритмы опираются на заранее определённые пространственно-временные шаблоны, что ограничивает их гибкость в сложных условиях.

Подходы на основе Deep Learning автоматически извлекают иерархические и различающие признаки из исходных данных, обеспечивая более высокую точность и устойчивость работы системы [3].

1 Распознавание движений по признакам руки и иные методы распознавания движений без использования глубокого обучения

Действия человека бывают статическими (положение конечностей) и динамическими (их движение). Распознавание может опираться на пространственные или временные признаки.

Пространственные модели:

1. Модели тела – восстанавливают позу человека по пространственной структуре действия, например, в виде 3D-кинематической модели суставов, или распознают действия напрямую из 2D-моделей.
2. Модели изображений – используют целостное представление действия в области интереса (силуэты, контуры, карты движения, оптический поток).
3. Пространственная статистика – представляет локальные действия с помощью статистики признаков из окрестных областей.

Пространственно-временные методы делятся на:

- Объёмные – представляют видео как пространственно-временной объём и опираются на признаки (текстура, цвет, поза, гистограммы направленных градиентов и оптического потока). Движения распознаются по схожестям двух объёмов. Подходы эффективны для простых действий, но плохо работают в толпе.
- Траекторные – отслеживают положение суставов или точек тела во времени, формируя траектории движений. Такие методы устойчивы к шуму и изменению освещения, применяются для сложных действий.

Временные модели:

1. Грамматики действий – описывают действие как последовательность состояний, изучая переходы между ними (Hidden Markov Models, Conditional Random Fields).
2. Шаблоны действий – типичный пример изменения признаков во времени, характеризующий конкретное действие (преобразования Фурье, вейвлеты, траектории частей тела).
3. Временная статистика – описывает распределение признаков движения во времени на основе статистических моделей.

Методы, основанные на внешнем виде, используют 2D/3D-изображения глубины и основаны на формах, движении или их комбинации. Появление камер глубины (например, Microsoft Kinect) ускорило развитие скелетных методов распознавания, в которых тело представляется как набор соединённых сегментов (суставов).

- Методы по форме – выделяют контуры, силуэты, локальные области. Например, распознавание жестов по 3D-точечному облаку.
- Методы по движению – применяют оптический поток, карты истории движе-

ния и гистограммы направленных движений.

- Гибридные методы – комбинируют признаки формы и движения.

Современные подходы используют обучение признаков, включая:

- Dictionary learning – создаёт разрежённое представление данных как линейную комбинацию базовых элементов.
- Genetic programming – эволюционный поиск функциональных зависимостей между признаками.
- Байесовские сети – вероятностные графовые модели для представления зависимостей между частями действия [1].

2 Методы распознавания движений с использованием глубокого обучения

В целом большинство систем, распознающих действия на основе глубокого обучения сталкиваются с ограниченной мощностью устройств (высокие вычислительные требования), необходимостью в хорошо подготовленном и достаточно большим датасетом, а также в ограниченной интерпретируемости результатов.

Автор работы [3] делит модели глубокого обучения на такие основные виды, как: двухпоточные сети, 3D-CHC, RNNs/LSTM, AM, Transformer, 2D-CHC, многопоточные сети, GCN, гибридные модели и мультимодальное обучение. Рассмотрим их базовые вариации.

Исходя из своего названия, двухпоточная СНС состоит из двух мало интегрированных (или вообще не интегрированных) нейронных сетей, каждая из которых выполняет свою задачу. Одна нейронная сеть называется пространственным потоком и выполняет непосредственно распознавания действия из видеокадров. Временной поток (вторая НС) распознает само движение благодаря плотному оптическому потоку. В процессе позднего слияния (могут использоваться другие варианты слияния) выполняемого в слое softmax результаты прогнозирования из двух сетей объединяются посредством усреднения или SVM.

Из плюсов данной модели можно выделить обработку временной динамики действий из оптического потока, совмещенной с обычной пространственной обработкой кадров, что значительно повышает точность по сравнению с только пространственной обработкой кадров. На стандартных датасетах показывает хорошую точность.

Две НС обычно плохо интегрированы, поэтому отсутствие связанности между ними, скорее всего, приводит к ухудшению результатов. Также вычисление оптического потока весьма накладно и может мешать распознавать системе в режиме реального времени.

Видеоданные могут быть представлены в виде 3D-тензора, где одна ось будет временной, а остальные пространственными. Для обработки видео в виде 3D-тензора были созданы 3D-СНС, которые базируются на основных базовых архитектурах и принципах 2D-СНС, но везде, где в обычных СНС были матрицы теперь находятся 3D-тензора.

Пространственно-временные особенности изучаются вместе, что значительно повышает точность краткосрочных движений, благодаря чему в коротких сценах по точности превосходят двухпоточные СНС. Однако из-за большого количества обрабатываемых данных имеет много настраиваемых параметров, которые весьма сложно настроить и добиться желаемого результата, а также большое количество данных не позволяет СНС увидеть долгосрочные закономерности и признаки, что весьма ограничивает применение 3D-СНС.

Рекуррентная нейронная сеть (РНС) и её измененная вариация LSTM при совместном распознавании временной динамики действий) показывают весьма хорошие результаты. РНС могут сохранять результаты работы одного шага работая в другом шаге в своих скрытых слоях. Хотя РНС и не могут долго удерживать значения предыдущих шагов, этого достаточно для хорошей обработки краткосрочных временных зависимостей.

Изменённая вариация LSTM содержит вентили, которые управляют попаданием значений в ячейки, которые очень долго могут хранить значения предыдущих шагов. Это позволяет игнорировать неподходящую или сохранять нужную информацию в ячейке продолжительное время, что позволяет хорошо работать с длинными последовательными данными, как, например, видео.

Системы, работающие на объединение результатов работы РНС и LSTM могут эффективно моделировать долгосрочные временные зависимости в видео, а сети LSTM решают проблему РНС с потерей длительных зависимостей (исчезающий градиент). Однако данные системы из-за рекуррентной структуры склонны к переобучению и хуже распознают пространственные характеристики кадра, нежели модели, основанные на СНС. Но в общем, их распознавание действий превосходит методы, основанные на СНС.

НС с слоем (-ями) АМ (от англ. attention mechanism – механизм внимания) позволяют обрабатывать только релевантную часть видеокадров по временным и пространственным признакам. Иначе говоря, для одного кадра обрабатывается только важная часть кадра и повторяющиеся кадры или маловажные кадры не участвуют в распознавании движений. Релевантность области или кадра определяется по похожести или корреляции данных кадра к искомым признакам, чертам, прочее. Из плюсов можно выделить концентрацию на значимых частях видео, улучшение контекстного понимания НС. К минусам можно отнести высокую ресурсоемкость, так как для каждого кадра дополнительно высчитывается карта внимания (или важность каждого пикселя). Также требуется больший датасет, чем обычно для СНС, чтобы механика внимания изучила релевантные признаки. И также производительность очень сильно зависит от той базовой модели, куда добавили слои с механизмом внимания.

Модели типа Transformer являются улучшенной версией моделей с АМ. Их отличием в распознавании видео является то, что теперь каждый кадр анализируется по своей релевантности к каждому другому кадру в видео. Таким образом, данные модели легко выделяют важные части каждого видео и имеют больше шансов точно распознать действие. Среди преимуществ модели можно назвать хорошее выявление и моделирование длительных характеристик и черт видео, что позволяет хорошо распознавать длительные действия. Также данные модели могут работать с входными последовательностями переменной длины.

Из недостатков можно указать крайне высокую ресурсоемкость, что может стать препятствием для создания хорошей модели, а также хотя данные модели хорошо выделяют общие черты видео, маленькие локальные черты могут ускользнуть от внимания модели.

Обычные 2D-СНС при распознавании видео покадрово, то есть при отдельном распознавании каждого кадра всё равно хорошо работают даже при отсутствии временного (длительного) распознавания движений. К плюсам относится более легкое обучение по отношению к 3D аналогам, хорошая производительность и качество распознавания изображений. Из минусов выделяют невозможность обработки временной информации видео, что значительно снижает точность распознавания движений на видео и приходится сочетать данную модель с другими методами, которые могут обрабатывать временную информацию.

Многопоточные сети являются улучшенной версией двухпоточных сетей. Сеть состоит из потока, обрабатывающего сами кадры (пространственный поток), потока, обрабатывающего оптический поток (временной поток), потоки создающие градиентные кадры для извлечения важных характеристик (черт). Завершает сеть слой с механизмом внимания. Так как это весьма новая модель, то у неё пока ещё малое количество исследований. Однако, данная сеть хорошо работает с динамичным фоном, и обработка большей части видео (некоторые модели обрабатывает не только кадры, оптический поток, и градиентные карты, но также аудио и иные составляющие видео) предоставляет дополнительную информацию, которая значительно улучшает точность распознавания движений. Слабыми сторонами данных сетей можно назвать чрезмерно сложные архитектуры, которые весьма сложно проектировать и обучать, разные потоки очень сложно эффективно объединять и некоторые потоки дублирую частично друг друга, поэтому модели часто работают с избыточной информацией.

Графовые сверточные сети (ГСС) – это такие модели глубокого обучения, которые специально разработанный для обработки графовых структурированных данных. Из-за того, что разные части видео связаны между собой отношениями, которые удобно представить в виде графа, то такие модели всё больше становятся актуальными. Модели ГСС хорошо моделируют зависимости между связанными узлами, что делает их очень пригодными для распознавания и моделирования человеческих поз или суставов.

GCN (ГСС) можно использовать для моделирования отношений между объектами в задаче видео или дополненной реальности, что делает их полезными для представления пространственных взаимодействий между объектами. Однако такие модели GCN можно использовать для моделирования отношений между объектами в задаче видео или дополненной реальности, что делает их полезными для представления пространственных взаимодействий между объектами. Также модели могут часто ошибаться в длительных зависимостях (чертах) и обучение таких моделей весьма ресурсоемкое.

Гибридные модели используются для того, чтобы компенсировать ограничения разных моделей, путём их совместной работы. Объединяются от двух и более моделей, чтобы улучшить качество и скорость распознавания движений. Гибридные модели могут интегрировать сильные стороны различных архитектур. Однако из-за сложности в проектировании таких моделей, они могут быть склонны к переобучению. Также чем больше моделей объединено, тем больше датасет должен быть и тем больше вычислительных ресурсов нужно.

Мультимодальное обучение является ещё более сложной версией многопоточного обучения в идейном плане. Данные модели объединяют визуальные, слуховые и текстовые данные, что повышает точность распознавания, но создает такие проблемы, как выравнивание данных и устойчивость к шумам. Интеграция информации из нескольких модальностей (например, визуальной, звуковой и текстовой) может обеспечить взаимодополняемость или усиление для более полного понимания входных данных. Однако, необходимость согласования данных из разных модальностей (например, синхронизация аудио и видео) может привести к дополнительным сложностям и ошибкам [3].

Выводы

Распознавание человеческой деятельности является одной из ключевых задач в области компьютерного зрения и искусственного интеллекта. Больше всего сейчас развиваются модели на основе глубокого изучения ведь, несмотря на высокие вычислительные требования и сложность настройки, модели на основе глубокого обучения (трансформеры, АМ и графовые сети) демонстрируют значительные успехи в распознавании движений человека. Наиболее перспективными направлениями развития являются гибридные и мультимодальные архитектуры, объединяющие визуальные, звуковые и текстовые данные. Такие подходы позволяют создавать более универсальные и точные системы распознавания, способные эффективно функционировать в реальном времени и в сложных условиях окружающей среды.

Литература

1. Beddiar D.R., Nini B., Sabokrou M., Hadid A. Vision-based human activity recognition: a survey // Multimedia Tools and Applications. 2020. Т. 79. С. 30509-30555. URL: https://link.springer.com/article/10.1007/s11042-020-09004-3.

2. Shin J., Hassan N., Miah A.S.M., Nishimura S. A Comprehensive Methodological Survey of Human Activity Recognition Across Diverse Data Modalities // Sensors. 2025. Т. 25, № 13, ст. 4028. DOI:10.3390/s25134028. URL: https://www.mdpi.com/1424-8220/25/13/4028.

3. Gong P., Luo X. A Survey of Video Action Recognition Based on Deep Learning // Knowledge-Based Systems. 2025. Т. 320. Ст. 113594. DOI:10.1016/j.knosys.2025.113594. URL: https://www.sciencedirect.com/science/article/pii/S0950705125006409.

Nekrasova D.R., Lazebnaya L.A. Review of modern methods for motion recognition from video data. *The article reviews modern methods for human action recognition (HAR) based on computer vision and deep neural networks. The principles of 3D-CNN, RNN/LSTM, Transformer, GCN, hybrid, and multimodal models are briefly described. It is noted that integrating different approaches improves the accuracy and robustness of action recognition systems.*

Keywords: *human action recognition (HAR), convolutional neural networks (CNN), computer vision, deep learning (DL), neural networks, recurrent neural networks (RNN, LSTM).*

СЕКЦИЯ №2
ТЕХНОЛОГИИ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМ

УДК 519.837.2

МЕТОДЫ ПРИМЕНЕНИЯ И ВОЗМОЖНОГО УЛУЧШЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В КОМПЬЮТЕРНЫХ ИГРАХ

Агапов И.С., ПеринскаяЕ.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: kamfer2711@gmail.com

Агапов И. С. Перинская Е.В. Методы применения и возможного улучшения искусственного интеллекта в компьютерных играх. *Контекстуальная осведомлённость – ИИ будет лучше понимать контекст игровых ситуаций и реагировать на них более эффективно.*
Ключевые слова: искусственный интеллект, генерация контента, персонализация игрового опыта, мотивация, игровые проекты.

Введение

Искусственный интеллект (ИИ) применяется в компьютерных играх для создания «умного» поведения персонажей (NPC), генерации контента и персонализации игрового опыта. Чтобы разобраться, как работает искусственный интеллект, мы должны понять, что же подразумевают под этим словосочетанием. Некоторые люди могут ответить на этот вопрос чем-то подобным: «Искусственный интеллект – это мозг, который воссоздан внутри компьютера». Отчасти они будут правы, но концепцию «ИИ как мозг» уже успели высмеять десятки раз – например, в Fallout: New Vegas. Также ИИ используется для аналитики поведения игроков, что помогает разработчикам улучшать игровые проекты.

Перспективы улучшения ИИ в играх

Динамическая адаптация – ИИ сможет более точно подстраиваться под индивидуальные предпочтения и навыки, предлагая уникальные вызовы и сценарии для каждого игрока.

Контекстуальная осведомлённость – ИИ будет лучше понимать контекст игровых ситуаций и реагировать на них более эффективно.

Поведение персонажей (NPC)

Адаптация под стиль игрока — NPC могут подстраиваться под действия и стратегию игрока.

Реактивное принятие решений — нейросети обеспечивают мгновенную реакцию NPC на изменения в игровом мире.

Обучение на основе опыта — NPC улучшают свои навыки и тактику, анализируя прошлые взаимодействия с игроком.

Генерация контента

Процедурная генерация уровней — нейросети создают уникальные игровые карты, адаптированные под стиль игры.

Генерация квестов и миссий — алгоритмы ИИ анализируют прогресс игрока, атрибуты персонажа и параметры игрового мира, создают квесты с учётом способностей и предпочтений игрока.

Генерация повествования — алгоритмы ИИ обучаются на основе существующих сюжетных структур, дуг персонажей и диалогов, генерируют разветвлённые сюжетные линии.

Персонализация опыта.

Адаптивные сценарии — ИИ анализирует действия игрока и адаптирует сценарий игры в реальном времени. Например, если игрок часто выбирает агрессивный стиль игры, ИИ может увеличить сложность боевых сцен.

Персонализированные игровые события — ИИ создаёт уникальные события, основанные на предпочтениях и действиях игрока. Например, в игре Red Dead Redemption 2 ИИ может создавать случайные события, такие как нападения бандитов или встречи с NPC, которые зависят от действий игрока.

Аналитика

Сбор данных — ИИ собирает данные о действиях игроков, таких как передвижение, взаимодействие с объектами и выборы в диалогах.

Анализ данных — ИИ анализирует собранные данные для выявления паттернов и предпочтений игроков. Это может включать анализ частоты использования определённых тактик, времени, проведённого на различных уровнях, и других аспектов игрового процесса.

Адаптация игры — на основе анализа ИИ может вносить изменения в игровой процесс, такие как изменение сложности или добавление новых элементов.

Игровой искусственный интеллект — это набор программных методов, которые используют в видеоиграх для создания иллюзии разума у NPC через поведение персонажей. Игровой ИИ включает в себя алгоритмы теории управления, робототехники, компьютерной графики и информатики в целом.

ИИ — технология, которая посредством машинного обучения позволяет системе научиться анализировать определённую информацию в виртуальной среде, чтобы получить поведение, более приближенное к человеческому. Ещё несколько десятилетий назад нечто подобное могло встретиться разве что в научной фантастике, но теперь аналогичные технологии применяют повсеместно.

Самая распространённая роль ИИ в видеоиграх – управление неигровыми персонажами, и разработчики часто используют различные трюки, чтобы NPC выглядели более умными. Один из широко используемых алгоритмов называется конечным автоматом (FSM или finite-state machine). Его ввели в разработку видеоигр в 1990-х годах. В FSM-алгоритме разработчик обобщает все возможные ситуации, с которыми может столкнуться ИИ, а затем программирует конкретную реакцию для каждой из них. Например, в шутерах искусственный интеллект атакует, когда появляется игрок, а затем отступает, когда его собственный уровень здоровья становится слишком низким.

Рисунок 1 – Пример алгоритма FSM NPC

В примере алгоритма FSM NPC может выполнять четыре основных действия в ответ на возможные ситуации: поиск помощи, уклонение, блуждание и нападение. Многие известные игры, например, Battlefield, Call of Duty и Tomb Raider, включают в себя успешные примеры искусственного интеллекта на основе FSM-алгоритма.

Более продвинутый метод, который используют разработчики для повышения персонализированного игрового опыта, — алгоритм дерева поиска Монте-Карло (MCTS или Monte Carlo Tree Search). Алгоритм MCTS был создан для предотвращения аспекта повторяемости, который присутствует в FSM-алгоритме. MCTS-алгоритм сначала обрабатывает все возможные ходы, доступные NPC в конкретный момент времени. Затем для каждого из этих возможных ходов он анализирует все действия, которыми игрок мог бы ответить. А далее — снова возвращается к оценке NPC уже на основе информации о поступках игрока.

Этот алгоритм искусственного интеллекта использовала компания IBM при создании Deep Blue — первого шахматного суперкомпьютера, который 11 мая 1997 года вошёл в историю и выиграл матч из шести партий у чемпиона мира по шахматам Гарри Каспарова.

Подобный алгоритм применяют и во многих стратегических играх. Но поскольку возможных ходов гораздо больше, чем в шахматах, рассмотреть их все попросту не получится. В таких играх алгоритм MCTS будет случайным образом выбирать некоторые из возможных ходов. Благодаря этому действия NPC становятся гораздо более непредсказуемыми для игроков.

Вспомним такую игру, как Civilization, в которой существует огромное количество вариантов событий, доступных для компьютерного противника. Постройка дерева для каждого возможного выбора и сценария заняла бы очень много времени. Именно поэтому, чтобы избежать столь огромных вычислений, алгоритм MCTS случайным образом выбирает несколько возможных вариантов. В итоге игра отнимает меньше ресурсов системы, при этом ИИ в ней всё ещё способен удивлять игроков.

Современные игры всё ещё оперируют старыми фундаментальными концепциями и методами в плане ИИ, но используют их в больших масштабах и с преимуществами вычислительной мощности компьютеров.

В будущем развитие ИИ в играх, скорее всего, не будет сосредоточено на создании более мощных NPC, чтобы они искали изощрённые способы победы над игроками. Вместо этого разработка сконцентрируется на том, как создать уникальный игровой опыт для каждого геймера.

Игроки в наши дни уделяют большое внимание деталям — это включает в себя не только внешний вид и качество графики, но и то, насколько яркой и интерактивной является игра во всех возможных отношениях. И именно ИИ может поднять игровой опыт на новый уровень. Быть может, в один прекрасный день игроки не смогут сказать, управляет персонажем в игре искусственный интеллект или другой геймер.

Выводы

Искусственный интеллект (ИИ) в играх, или игровой искусственный интеллект, – это технологии для создания «умного» поведения персонажей и элементов в видеоиграх. Его функции – взаимодействие игрового мира с действиями игрока, адаптация NPC, создание стратегий и процедурная генерация контента.

ИИ получает информацию об игроке и виртуальном мире с помощью чтения состояния мира, сенсоров и зон обнаружения, событий и триггеров, систем наблюдения, взаимодействия с объектами, реакций на команды, систем поведения и сценариев, анализа данных и статистики.

Минусы, которые есть у компьютерного искусственного интеллекта – это ограниченная гибкость, проблемы с производительностью, сложность разработки, отсутствие креативности, ошибки в поведении, а также этические и социальные вопросы**.**

Литература

1. Паласиос, Х. Unity 5.x. Программирование искусственного интеллекта в играх/ Х. Паласиос; пер. Р. Н. Рагимова — М.: ДМК Пресс, 2017 — 272 с.

2. Аттетков А. В. , Зарубин В. С., Канатников А. Н. Введение в методы оптимизации : учебное пособие. – М. : Финансы и статистика, 2014.

3. Анализ применения искусственного интеллекта в видеоиграх / URL: https://sci-article.ru/stat.php?i=1669122954

4. Имитация разума: как устроен искусственный интеллект в играх / URL: https://habr.com/ru/companies/netologyru/articles/598489/

Agapov I. S. Perinskaya E.V. Methods of application and possible improvement of artificial intelligence in computer games. *Contextual awareness — AI will better understand the context of game situations and respond to them more effectively.*

Keywords: *artificial intelligence, content generation, personalization of the gaming experience, motivation, game projects.*

УДК 004.75

ИССЛЕДОВАНИЕ АРХИТЕКТУРНЫХ МЕХАНИЗМОВ НАДЁЖНОСТИ И МАСШТАБИРОВАНИЯ В РАСПРЕДЕЛЁННЫХ RPA-СИСТЕМАХ

Алымов Д.А., Боднар А.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия им. Л.П. Фельдмана»
e-mail: dnlalmv5@mail.ru, linabykova13@ya.ru

Алымов Д.А., Боднар А.В. Исследование архитектурных механизмов надёжности и масштабирования в распределённых RPA-системах. *Выполнено исследование архитектурных механизмов, обеспечивающих надёжность и масштабируемость RPA-систем в корпоративной среде. Рассмотрены принципы организации оркестрации, взаимодействия агентов и интеграции AI-модулей. Особое внимание уделено методам репликации состояния, балансировке нагрузки и обеспечению отказоустойчивости в многоуровневых системах.*
Ключевые слова: *RPA, оркестрация, масштабирование, отказоустойчивость, корпоративные системы, архитектура автоматизации*

Введение

Современные корпоративные системы характеризуются высокой степенью распределённости и необходимостью автоматизации большого числа рутинных процессов. Технологии роботизированной автоматизации процессов (RPA) стали ключевым инструментом для реализации таких задач. В условиях комплексных архитектур и интеграции различных программных решений возникает потребность в создании модели, способной объединить UI-, API- и интеллектуальные (AI) компоненты автоматизации в единую архитектуру.

Цель данного исследования заключается в формировании архитектурной модели интеграции RPA в распределённые корпоративные системы, обеспечивающей масштабируемость, устойчивость и универсальность взаимодействия с различными уровнями автоматизации. В рамках предложенного подхода RPA рассматривается не как отдельная технология эмуляции действий пользователя, а как универсальный слой исполнения, взаимодействующий с системой посредством унифицированного оркестратора и абстрактных интерфейсов.

Оркестрация и планирование в распределённой среде

Оркестрация RPA в корпоративной архитектуре представляет собой многоуровневую систему управления, где центральный оркестратор выполняет роль координатора между роботами, процессами и внешними системами. В корпоративной системе оркестратор отвечает, например, за распределение задач между роботами, обрабатывающими обращения клиентов, обновляющими данные, связанные с клиентом и формирующими отчётность. Он контролирует версии сценариев для разных отделов, обеспечивая согласованное выполнение даже при параллельной обработке тысяч операций.

Планирование выполнения процессов реализуется с помощью различных типов планировщиков. Наиболее традиционный подход, временной, когда задачи запускаются по расписанию в определённые временные интервалы. Он используется для регламентных процедур, например, ночных выгрузок отчётов о продажах, синхронизации клиентских данных или регулярного обновления маркетинговых сегментов. Более гибкий вариант, событийный планировщик, который активирует роботов при наступлении конкретного события: изменении статуса сделки, закрытии заявки в службу поддержки или поступлении платежа. Для реактивных сценариев применяется webhook-механизм, позволяющий внешним сервисам инициировать выполнение робота в реальном времени, например, при получении новой заявки с сайта или обновлении информации в стороннем аналитическом модуле. Ключевым элемен-

том масштабируемой архитектуры являются менеджеры очередей, обеспечивающие буферизацию задач и балансировку нагрузки. Очереди используются, например, для обработки входящих обращений клиентов: при высокой нагрузке на отдел поддержки сообщения ставятся в очередь и распределяются по ботам в зависимости от заданных приоритетов.

Для обеспечения надёжности выполнения применяются архитектурные механизмы контроля целостности и восстановления. Чекпоинты позволяют сохранять промежуточное состояние процесса, что особенно важно при длительных сценариях, таких как комплексная обработка заказов или многоступенчатые воронки продаж [1,2]. Далее приведен пример оркестрации и планирования в распределённой среде абстрактной корпоративной системы (см. рис. 1).

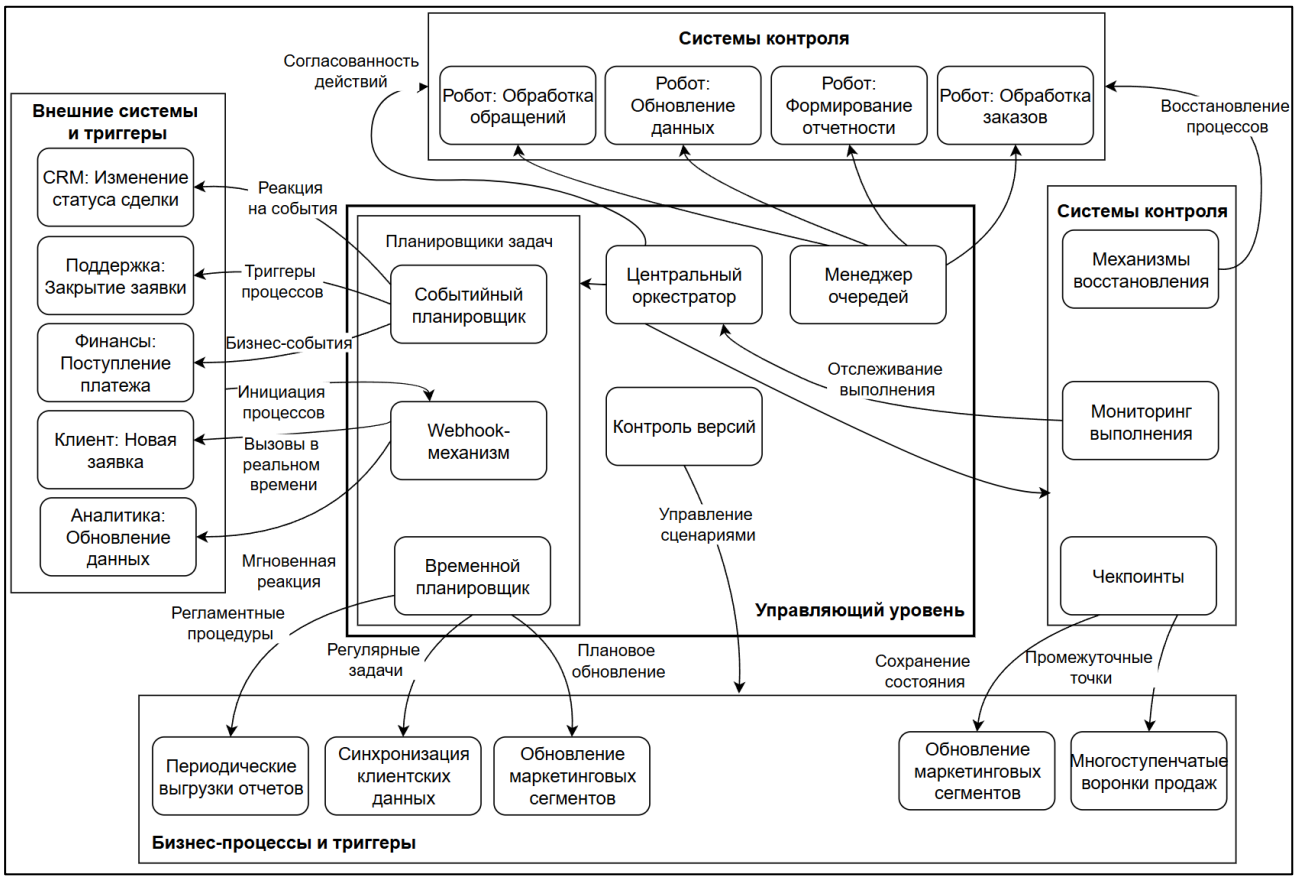

Рисунок 1 – Оркестрация и планирование

Исполнительный уровень

Исполнительный уровень RPA-системы формируется набором агентов или воркеров, выполняющих задачи, назначенные оркестратором, и представляющих собой основную вычислительную и интеграционную платформу автоматизации процессов (см. рис. 2).

На UI Automation слое агенты взаимодействуют с пользовательским интерфейсом системы через селекторы, абстрактные дескрипторы элементов и fallback-механизмы, также включающие компьютерное зрение. Такой подход особенно актуален для автоматизации действий в веб-клиентах, не имеющих открытого API: например, для автоматического заполнения данных клиентов или навигации по вкладкам. Использование компьютерного зрения позволяет распознавать динамические элементы интерфейса, а адаптивное кэширование снижает задержки при повторных операциях и повышает устойчивость ботов при изменениях в UI.

В слое API агенты взаимодействуют с корпоративными сервисами, используя прямые коннекторы. Это позволяет напрямую обновлять статусы данных, или синхронизировать клиентские данные с внешними источниками. Разделение на пакетную и транзакционную обработку позволяет гибко настраивать режимы интеграции: пакетный режим подходит для

периодической массовой загрузки лидов, тогда как транзакционные вызовы применяются для мгновенного обновления информации при изменении статуса конкретного клиента.

AI адаптеры, при необходимости включаемые в архитектуру, предоставляют интеграцию с моделями компьютерного зрения и NLP, расширяя возможности интеллектуальной автоматизации. В корпоративных системах такие адаптеры применяются для анализа входящих сообщений клиентов (распознавание текста и намерений в обращениях службы поддержки), автоматической категоризации заявок, извлечения информации из документов (паспортов, счетов, актов), а также прогнозирования вероятности закрытия сделки или оценки удовлетворённости клиентов.

Дополнительно агенты могут выполнять вспомогательные функции: транзакционную координацию процессов между системой и внешними системами, а также управление сессиями и временными токенами доступа. Например, такие модули хранят ключи интеграции с почтовыми сервисами, платежными шлюзами или системами аналитики [3].

Далее приведен пример взаимодействия слоев в абстрактной корпоративной системе (см. рис. 2).

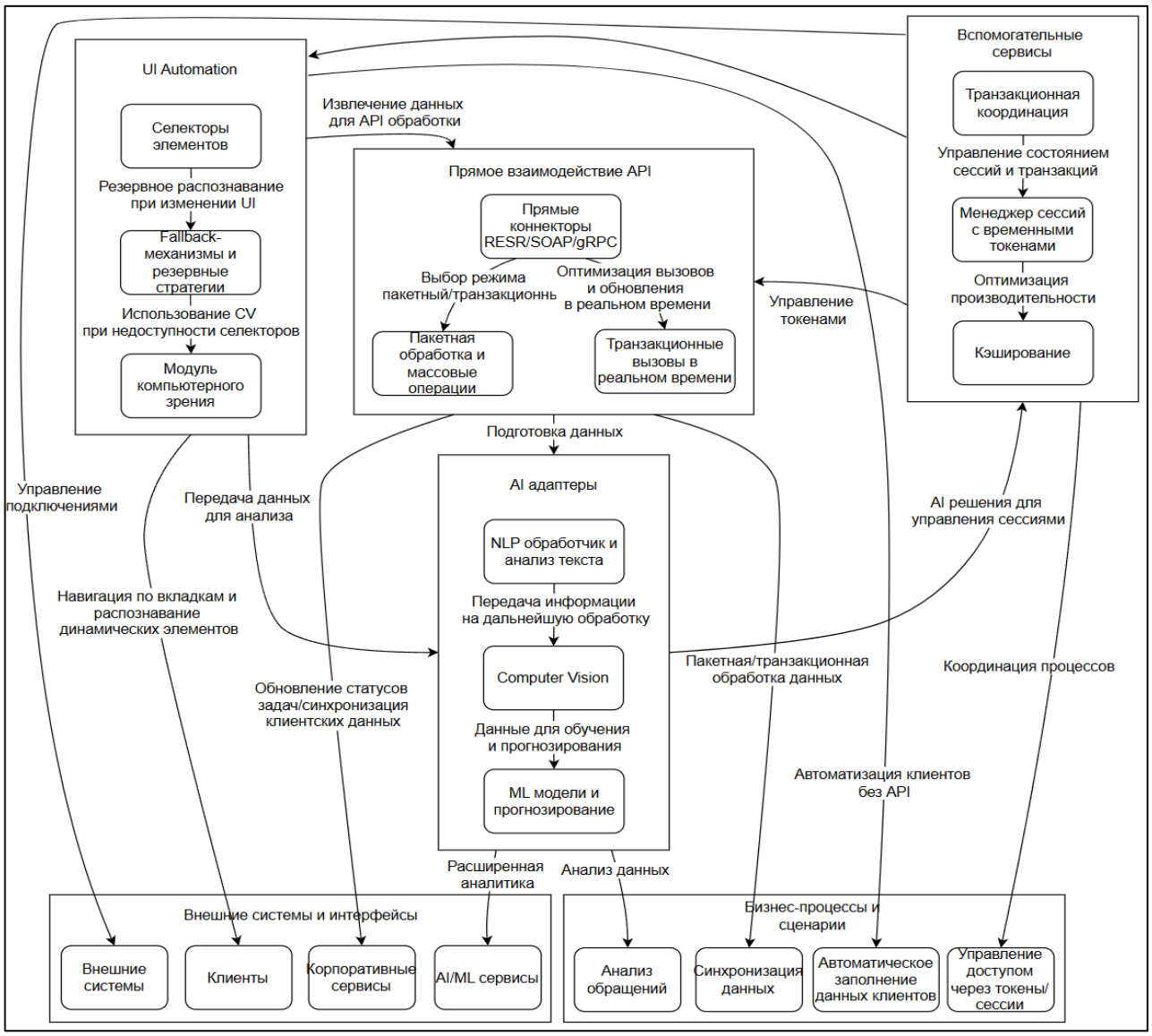

Рисунок 2 – Взаимодействие слоев RPA с системой

Надёжность, масштабирование и отказоустойчивость

Обеспечение надёжности и масштабируемости является ключевым аспектом архитектуры распределённых RPA-систем, особенно при внедрении в корпоративные экосистемы,

где роботы взаимодействуют с тысячами клиентских записей, заказов и коммуникаций одновременно. Горизонтальное масштабирование воркеров и оркестратора позволяет динамически увеличивать вычислительные ресурсы в зависимости от объёма задач. Оркестратор не хранит контекст выполнения процессов локально, а опирается на централизованное хранилище состояния, где сохраняются данные о задачах, токенах, статусах и контекстах пользователей.

Для повышения устойчивости к отказам применяются механизмы, позволяющие резервным узлам оркестратора автоматически активироваться при сбое основного. В случае частичных проблем в подсистемах применяется стратегия, когда система временно ограничивает выполнение только проблемных задач без остановки остальных роботов, тем самым поддерживая общую стабильность.

Дополнительную надёжность обеспечивает ролевое распределение оркестраторов – активные и резервные узлы координируют выполнение процессов, выбирая ведущего координатора. Такая схема критически важна для корпоративных платформ, интегрированных с внешними сервисами: рассылками, платежными шлюзами, аналитикой, документооборотом. Даже при выходе одного из оркестраторов из строя управление автоматизацией продолжается без остановки [4].

Интеграция AI-модулей

Интеграция AI-модулей в RPA-систему позволяет существенно расширить интеллектуальные возможности автоматизации. На этапе предварительной обработки данных AI-компоненты применяются для классификации и нормализации. В CRM это проявляется, например, в автоматическом анализе входящих писем клиентов, распознавании вложений, классификации обращений по категориям (жалоба, запрос, предложение), а также в извлечении ключевых данных из документов – договоров, актов, коммерческих предложений и анкет клиентов.

Во время выполнения задач AI используется для распознавания элементов интерфейса, адаптивной навигации и поддержки принятия решений в реальном времени. При автоматизации пользовательских сценариев RPA-агенты с поддержкой компьютерного зрения могут определять динамически изменяющиеся элементы интерфейса или корректировать действия робота при изменении структуры страницы.

На этапе постобработки AI играет важную роль в аналитике и прогнозировании эффективности автоматизированных процессов. Это может выражаться в построении моделей, которые оценивают результативность работы роботов, предсказывают возможные задержки, а также прогнозируют поведение клиентов. Такие модули могут использовать данные о выполненных задачах, временных затратах и истории взаимодействий для автоматического улучшения алгоритмов и сценариев автоматизации.

Для безопасности и управляемости AI-компонентов каналы данных для обучения и переобучения моделей должны быть строго контролируемыми, для этого используются логи выполнения процессов. Это позволяет не только отслеживать изменения в поведении AI, но и гарантировать воспроизводимость результатов, что критично при работе с персональными и коммерческими данными.

Важным принципом архитектуры является декомпозиция AI-компонентов, при которой модули искусственного интеллекта выступают в роли адаптеров или фасадов, подключаемых к исполнительному уровню RPA, но не являются частью ядра оркестрации. Такой подход гарантирует, что даже при отключении или замене AI-модулей базовая автоматизация продолжает функционировать без перебоев. Это особенно актуально для CRM-сред, где стабильность и предсказуемость автоматизации клиентских операций имеют приоритет над экспериментальными функциями искусственного интеллекта [5].

Выводы

Таким образом, в ходе исследования рассмотрены архитектурные принципы постро-

ния распределённых RPA-систем, направленные на обеспечение их надёжности, масштабируемости и адаптивности при интеграции в корпоративные среды. Проанализированы ключевые паттерны внедрения автоматизации, механизмы оркестрации, исполнения и взаимодействия с AI-модулями. Показано, что сочетание модульной архитектуры, горизонтального масштабирования и интеллектуальной оркестрации позволяет повысить устойчивость и эффективность автоматизации бизнес-процессов, обеспечивая гибкость системы при изменении нагрузки и инфраструктуры.

Литература

1. Фахретдинова А.И. Результаты разработки модуля оркестрации запросов для платформы RPA / Фахретдинова А.И. // XXIV Туполевские чтения (школа молодых ученых). – 2019 – №1. – Т. 5. – С. 52-55.

2. Жуков М.О. Автоматизация бизнес-процессов финансового отдела консалтинговой компании с помощью технологии RPA / Жуков М.О., Малков О.Б. // Информационные технологии и автоматизация управления. – 2021 – С. 28-42.

3. Бердюгина А.А. Применение RPA технологии для автоматизации бизнес-процессов, основанных на системах управления взаимоотношений с клиентами / Бердюгина А.А. // Молодежь и наука. – 2022 – №1. – Т. 2. – С. 29-31.

4. Сухоплюев Д.И. Мониторинг отказоустойчивости в распределенных системах / Сухоплюев Д.И., Назаров А.Н. // Computational nanotechnology. - 2024. – Т. 11. – №4. – С. 94-106.

5. Загорская К.Р. Синергия RPA и искусственного интеллекта / Загорская К.Р., Иванова И.В. // Электронный научный журнал «Дневник науки». – 2023 – №11. – Т. 1. – С. 1-11.

Alymov D.A., Bodnar A.V. Study of Reliability and Scalability Architectural Mechanisms in Distributed RPA Systems. *This research investigates architectural mechanisms that ensure the reliability and scalability of RPA systems in corporate environments. The principles of orchestration organization, agent interaction, and AI module integration are examined. Special attention is given to state replication methods, load balancing, and fault tolerance in multi-tiered systems.*

Keywords: *RPA, orchestration, scalability, fault tolerance, corporate systems, automation architecture*

УДК 004.8

АНАЛИЗ АРХИТЕКТУРЫ СИСТЕМ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ВЕБ-ИНТЕРФЕЙСОВ

Белый А.А, Григорьев А.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программной инженерии» им. Л.П. Фельдмана
e-mail: artem8belyy@yandex.ru, grigorievalvl@mail.ru

Белый А.А., Григорьев. А.В. Анализ архитектуры систем автоматической генерации веб-интерфейсов. *Выполнен краткий сравнительный анализ онтологических, DSL, ML и MLLM-подходов. Рассмотрена гибридная MLLM-система UI-UG, сочетающая визуально-языковую модель и спецификацию DSL. Проанализированы подходы, позволившие повысить качество дизайна на 14-16% по ряду метрик, сократить время отклика до 2 с. Определены ограничения и перспективы.*
Ключевые слова: *генерация веб-интерфейсов, DSL, онтологии, большие языковые модели, мультимодальные модели.*

Введение

Системы, способные генерировать веб-интерфейсы, переживают стремительное развитие. Современные библиотеки для создания сайтов в значительной степени уменьшили трудоемкость разработки веб-интерфейсов, понизили требования к квалификации программистов. Это привело к идее разработки систем, способным создать сайт намного быстрее и дешевле, чем это мог бы сделать человек. Наиболее значимым прорывом в этой области за последние несколько лет является применение искусственного интеллекта, способного понять требования пользователя по текстовому описанию и примерам уже существующих сайтов. Но искусственный интеллект все равно имеет недостатки, которые ему не позволяют полностью исключить другие подходы для синтеза веб-интерфейсов. Наилучшего результата можно добиться путем смешения подходов, взаимодополняющих друг друга. Гибридный подход способен обеспечить баланс между гибкостью системы и жесткой структурой веб-интерфейсов.

Целью данной работы является рассмотрение онтологических, DSL, ML и MLLM-подходов, определение их плюсов и недостатков, анализ мультимодальной большой языковой модели UI-UG как пример гибридного подхода, оценка и описание ограничений UI-UG.

1 Постановка задачи

Существует множество способов, которыми могут быть заданы требования к системе, генерирующей веб-интерфейс. Входные данные от пользователя зависят от архитектуры системы, которая будет синтезировать программный код. К тому же самих принципы генерации сайтов достаточно много, эффективность которых зависит от требований, поставленных перед системой.

Задачей данной работы является рассмотрение наиболее продуманных подходов для синтеза сайта, способных обеспечить быстрый отклик и эффективно использовать вычислительные ресурсы, выдавать результат, соответствующий требованиям пользователя. Также необходимо определить ситуаций, в рамках которых их использование будет оправдано.

2 Обзор подходов

2.1 Онтологии

Онтологии, в контексте генерации сайтов, являются формальными представлениями знаний о предметной области, которые описывают объекты, их свойства и отношения между ними. Такой подход требует знания предметной области [1]. Часто онтологии реализуются вручную, что является очень трудоемкой задачей, но взамен такая система является максимально предсказуемой, любое решение системы можно отследить. Недостатком являются невозможность обработать сложные и неоднозначные спецификации [2].

Для автоматизированного формирования онтологий для предметных областей могут быть использованы NLP-инструменты и языковые модели. В результате получается система, в которой онтологии генерируются автоматически, а эксперт при необходимости корректирует ошибки. Такой подход компенсирует недостаток, связанный с трудозатратностью создания онтологий вручную [3].

2.2 DSL-подходы

Одним из традиционных методов создания интерфейса является перевод требований пользователя в DSL [2]. Суть подхода заключается в применении абстракции конкретной предметной области [4]. Изначальная задача декомпозируется и описывается предметно-ориентированный язык, после чего компилятор генерирует код на целевом языке программирования [5]. Для сайтов целевыми языками являются HTML, CSS и JavaScript. Недостаток такого подхода – неспособность обрабатывать сложные задачи, необходима ручная доработка и определение правил.

2.3 Эволюционные методы

Генетический алгоритм генерирует код путем многократного отбора наиболее случайно сгенерированных вариантов. После каждого отбора наиболее качественного кода генерируются новые варианты решения задачи, логически продолжая ранее созданный код. Итерации повторяются до тех пор, пока не будет создано решение, которое полностью соответствует всем требованиям.

2.4 ML, LLM и MLLM

Наиболее распространенным подходом на текущий момент является машинное обучение (ML). Отличительной особенностью такого решения от ранее представленных является то, что алгоритм решения задан неявно. Алгоритм машинного обучения получает на вход большое количество данных, после обработки которых он способен выявить скрытые признаки и закономерности, принимать решения [6]. Явным плюсом является то, что разработчику больше не нужно вручную прописывать предметную область, достаточно подобрать наиболее эффективную модель машинного обучения и донастроить ее под бизнес-требования заказчика. Основная сложность таких систем при таком подходе – сбор и очистка данных, которые будут использованы для обучения. На это может понадобиться 80% от всего времени обучения модели.

Большие языковые модели (LLM) являются алгоритмами глубокого обучения, подмножеством машинного обучения. LLM эффективны при модификации кода или генерации относительно маленького количества кода, при попытке получить что-то сложное модель начинает совершать ошибки, теряет контекст из-за ограничений контекстного окна и накопления ошибок. Для нивелирования этого недостатка LLM используют в связке с другими подходами. Основной принцип такого подхода – разбиение пользовательской спецификации на подзадачи, генерация кусочков кода и дальнейшего объединения.

3 Архитектура MLLM UI-UG

UI-UG является примером мультимодальной большой языковой моделью (MLLM). При обучении использовался Qwen2.5-VL-72B, создающий спецификации DSL, на основе которых будут созданы страницы. Процесс обработки был оптимизирован путем отключения дообучения визуального энкодера, дообучалась только большая языковая и прослойка между визуально-языковой адаптер, который преобразует признаки изображения в токены для языковой модели, и LLM, отвечающая за то, чтобы языковая модель корректно поняла, что находится на сайте. Вследствие этого модель стала потреблять меньше ресурсов на аналогичные действия.

Следующим этапом стало обучение модели удачным и неудачным вариантам дизайна. Удачными считаются те, где хорошо сочетаются цвета, текст читабельный, элементы не залазят друг на друга. В итоге модель стала генерировать на 14-16% более качественные сайты с точки зрения дизайна.

Для улучшения качества кода, для генерации редких компонентов был использован метод GRPO. В рамках этого метода генерировали несколько сайтов по одинаковому запросу и определяли какие из них являлись удачными, а какие нет. Лучшими считались те вариации, в которых были корректно сгенерированы спецификации DSL (в этой модели это JSON) и редкие компоненты.

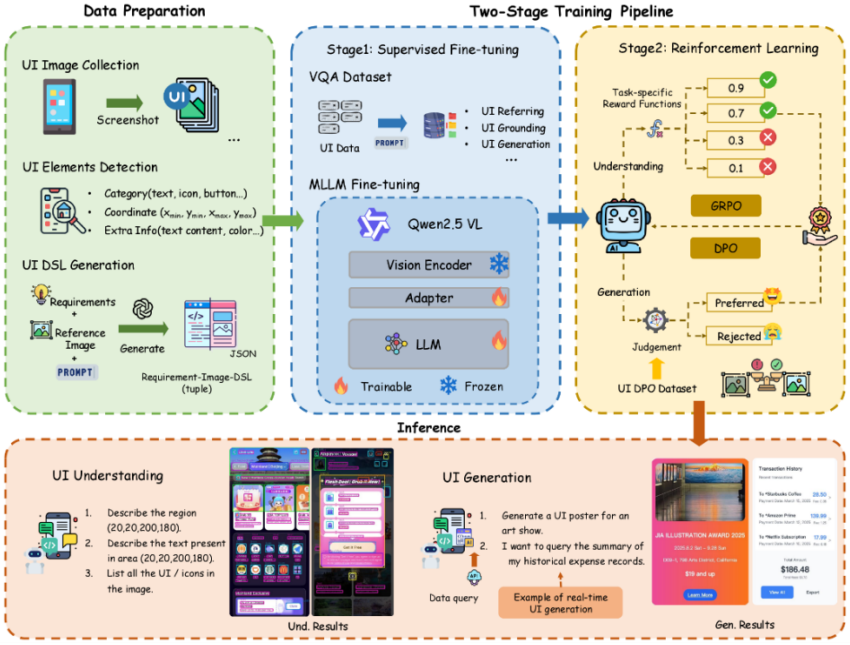

Рисунок 1 – Структура MLLM

Полученную модель сравнили актуальными на текущий момент моделями. Авторами исследования было выделено 2 группы задач: понимание интерфейса, в рамках которого сравнивалась точность цвета текста, доля элементов, у которых модель корректно определила тип, распознавание текста и генерацию интерфейса, во что вошли доля валидного JSON, сходство с примером, оценки MLLM по структуре, дизайну, содержанию, интерактивность. В понимании интерфейса UI-UG оказалась лидером, а по генерации немного отставала.

Далее уменьшили количество параметров и точность представления весов до 4 бит (INT4), сократили системные запросы. Модель стала значительно быстрее, благодаря времени на ответ равному в среднем 2 секундам стало возможно взаимодействие с моделью без долгих задержек [7].

На этом примере можно понять, что реализация, объединяющая несколько подходов одновременно, является самой эффективной на текущий момент. Недостатки DSL в виде невозможности генерировать сайты, в которых используются редкие элементы, были нивелированы гибкостью LLM. В свою очередь LLM получила ограничения по структуре в виде DSL, что помогло избежать ошибок, связанных с невозможностью одновременно учитывать все компоненты сайта.

Одним из недостатков UI-UG является невозможность обновлять определенную часть сгенерированного веб-интерфейса, в описанной версии можно только сгенерировать весь код сайта заново. При добавлении функции обновления только части сайта пользователь сможет вносить более мелкие доработки, создавать различные версии правок и выбирать наиболее подходящие под его требования. Также будет удобно функция версионирования сгенерированных сайтов, это позволит возвращаться к предыдущим состояниям сайта в случаях, когда правки были некорректно описаны или являлись избыточными.

Спецификации DSL не имеют жесткой структуры, DSL генерируется моделью, после чего проверяется на качество, но не обладает четкими правилами, способными предотвратить ряд ошибок. Добавление ограничений на размеры элементов, не дающие одному блоку оказаться поверх другого, контрастность, не позволяющие выбрать цвета для сайта, отталкивающие пользователей, поспособствует качеству генерируемых сайтов.

Для улучшения генерируемого кода можно создавать библиотеку с примерами редких и типовых решений при реализации интерфейса, подкрепив этими примерами обучение. Это позволит снизить ошибки, возникающие при генерации сложных элементов так, как модель сможет скопировать гарантированно качественную реализацию.

Выводы

В данной работе были рассмотрены онтологии, DSL, машинное обучение, используемые для генерации веб-интерфейсов. Было выявлено, что ни один из подходов не способен в полной мере выполнить задачу генерации веб-интерфейса. На практике наиболее успешным оказался гибридный подход, объединяющий в себе строгую структуру DSL подхода и гибкость мультимодальных боль-

ших языковых моделей, рассмотренный на примере модели UI-UG. За счет мультимодального подхода и генерации спецификаций DSL удалось повысить качество дизайна на 14-16% по ряду метрик и сократить отклик в среднем до 2 секунд.

В текущей версии UI-UG не реализованы строгие ограничения для DSL и невозможно обновлять только часть сгенерированного веб-интерфейса, что в значительной степени замедляет работу с системой, замедляя процесс доработки веб-интерфейса. Чтоб решить эту проблему необходимо реализовать возможность вносить точечные правки вместо того, чтобы генерировать весь интерфейс заново, создать критерии, которые будут сразу отсекать DSL с наложившимися блоками интерфейса, некорректно подобранным сочетанием цветов. Также необходимо добавить заранее подготовленные реализации сложных элементов интерфейса, позволяющие UI-UG брать уже подготовленные элементы, а не пытаться их сгенерировать. Эти улучшения понизят количество ошибок и повысят общее качество выдаваемого результата.

Таким образом, работа подтверждает ценность гибридного подхода, позволяющего компенсировать недостатки различных решений на примере модели UI-UG.

Литература

1. Грибова В.В., Паршкова С.В., Федорищев Л.А. Онтологии для разработки и генерации адаптивных пользовательских интерфейсов редакторов баз знаний // Онтология проектирования. 2022. № 2(44). URL: https://cyberleninka.ru/article/n/ontologii-dlya-razrabotki-i-generatsii-adaptivnyh-polzovatelskih-interfeysov-redaktorov-baz-znaniy (дата обращения: 29.10.2025).

2. A Comparative Review of AI Techniques for Automated Code Generation in Software Development: Advancements, Challenges, and Future Directions. ResearchGate. URL: 2. https://www.researchgate.net/publication/378542081_A_Comparative_Review_of_AI_Techniques_for_Automated_Code_Generation_in_Software_Development_Advancements_Challenges_and_Future_Directions (дата обращения: 29.10.2025).

3. Смирнова Е.В., Шаклеина Д.Д. Разработка веб-приложения автоматизированного формирования онтологии // Проблемы Науки. 2025. № 7(206). URL: https://cyberleninka.ru/article/n/razrabotka-veb-prilozheniya-avtomatizirovannogo-formirovaniya-ontologii (дата обращения: 29.10.2025).

4. Gudiev A.V., Grazhevskaya A.S. Graphic DSL for mobile development // Труды ИСП РАН. 2019. № 3. URL: https://cyberleninka.ru/article/n/graphic-dsl-for-mobile-development (дата обращения: 29.10.2025).

5. Квашенников М.С. Алгоритм формирования DSL-структуры на основе данных с рукописного рисунка-макета // Поколение будущего: Взгляд молодых ученых – 2024. Курск: ЗАО «Университетская книга», 2024. С. 92–94. EDN KLQUFZ.

6. Сметана В.В. Развитие машинного обучения: от теории к приложениям // НАУ. 2024. № 101-1. URL: https://cyberleninka.ru/article/n/razvitie-mashinnogo-obucheniya-ot-teorii-k-prilozheniyam (дата обращения: 29.10.2025).

7. UI-UG: Multimodal UI Generation with DSL and Preference Optimization // arXiv:2509.24361 (версия v1).

Belyi A.A., Grigoriev A.V. Analysis of the architecture of automatic generation of web interfaces. *A brief comparative analysis of otnological, DSL, ML and MLLM approaches is carried out. A hybrid MLM UI-UG system combining a visual-language model and a DSL specification is considered. The approaches that made it possible to improve the quality of design by 14-16% in a number of metrics, reduce the response time to 2 seconds are analyzed.*
Keywords: *generation of web interfaces, DSL, ontologies, large language models, multimodal models.*

УДК 004.5+658.8

СИНТЕЗ МОДЕЛИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ОПТИМИЗАЦИИ ВЕБ-РЕСУРСА НА БАЗЕ КОМПЛЕКСНОЙ ВЕБ-АНАЛИТИКИ

Бубнов Г.В., Григорьев А.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: gopractic@gmail.com, grigorievalvl@mail.ru

Бубнов Г.В., Григорьев А.В. Синтез модели принятия решений для оптимизации веб-ресурса на базе комплексной веб-аналитики. *В статье исследуются подходы к повышению эффективности веб-ресурсов посредством анализа пользовательского поведения, с фокусом на одностраничные сайты, где конверсия критически зависит от структурно-контентных характеристик. Предлагается модель принятия решений, основанная на интеграции данных комплексной веб-аналитики, позволяющая эмпирически обосновать оптимизацию интерфейсных и контентных элементов для повышения конверсии и достижения целевых бизнес-показателей.*
Ключевые слова: *система анализа содержания веб-сайта, информационно-аналитическая система, эффективность онлайн-коммерции, повышение конверсии.*

Введение

Актуальность исследования обусловлена возрастающими требованиями к эффективности веб-ресурсов в условиях развития цифровой экономики. Ключевым показателем результативности выступает конверсия, количественно отражающая соотношение целевых действий пользователей к общему трафику. Проблема оптимизации данного показателя особенно значима для одностраничных сайтов, где ограниченная структура предъявляет повышенные требования к каждому элементу интерфейса и контента.

Анализ существующих подходов выявил их ориентацию преимущественно на визуальные и технические аспекты веб-дизайна. В противоположность этому, в работе предлагается комплексная методика, основанная на интеграции инструментов веб-аналитики и данных о поведенческих паттернах пользователей. Такой подход позволяет рассматривать веб-ресурс как динамическую систему, где эффективность определяется корреляцией структурно-контентных характеристик и пользовательской активности.

Целью исследования является синтез обоснованной модели принятия решений для оптимизации веб-ресурсов. Предлагаемая модель обеспечивает эмпирическую базу для структурных и контентных изменений, направленных на повышение конверсии и достижение установленных бизнес-показателей.

1 Системный анализ предметной области

Наличие веб-сайта у коммерческой организации не является достаточным условием роста продаж. Критическим фактором становится способность ресурса трансформировать посетителей в клиентов. В условиях цифровой экономики, где каждое пользовательское взаимодействие представляет потенциальную коммерческую операцию, оптимизация конверсии становится стратегической задачей для обеспечения устойчивого развития компаний.

Решение данной задачи осложняется ее междисциплинарной природой. Эффективный анализ веб-ресурсов требует синтеза компетенций в области веб-разработки, обработки данных, юзабилити, поведенческой психологии и маркетинга. Особую значимость приобретает способность учитывать когнитивные особенности восприятия информации, корректно интерпретировать поведенческие метрики и идентифицировать латентные паттерны взаимодействия с интерфейсом.

Системный подход требует оценки не отдельных средств, а их совокупного эффекта в

рамках общей цифровой экосистемы. Системный подход, позволяющий рассматривать исследуемый объект как совокупность взаимосвязанных компонентов, действительно является ключевым для установления таких связей между элементами процесса [1]. Эффективность интерфейсных компонентов оценивается через метрики глубины прокрутки, времени пребывания на сайте и коэффициента отказов. Анализ пользовательских сценариев проводится с применением методов когортного анализа, A/B-тестирования и тепловых карт кликов [2], что позволяет выявлять критические точки снижения интереса и устранять барьеры в конверсии.

Инструменты повышения конверсии классифицируются на аналитические, поведенческие и технические. К аналитическим относятся системы веб-аналитики, обеспечивающие сбор и интерпретацию данных о пользовательских потоках. Поведенческие инструменты включают внедрение адаптивных интерфейсов, динамическую сегментацию аудитории, персонализацию контента и применение триггерных механизмов, нацеленных на конкретные действия пользователя. Технические решения охватывают оптимизацию скорости загрузки страниц, адаптацию под мобильные устройства, корректную работу форм обратной связи. Системный анализ подразумевает количественно оценивание вклада каждого инструмента в общую эффективность сайта.

Для обеспечения сквозной аналитики применяется системы, соединяющие данные из рекламных каналов и веб-метрик в единую модель. Такой подход позволяет выявлять причинно-следственные зависимости между поведением пользователей и результатами конверсии в режиме реального времени. Применение системного анализа в области конверсии интернет-сайтов обеспечивает переход от интуитивных решений к научно обоснованным методам управления цифровой эффективностью.

Под шаблонным интерфейсом, с которым работают посетители сайта, будем принимать сайт, включающий в себя навигационную панель, информационный блок, плиточное представление каталога товаров, панель фильтров для сортировки товаров, блок CTA – призыва к действию.

Лендинг должен не только представлять компанию, но и направлять клиента к нужному действию – покупке, заявке, звонку и т.д. Таким образом, проектируемая структура веб-сайта, ориентированная на конверсию, является фундаментом всей информационно-аналитической системы.

2 Обзор методов веб-аналитики и показателей эффективности

Современные методы веб-аналитики направлены на формирование целостного представления о взаимодействии пользователей с веб-ресурсом и последующую количественную оценку эффективности. Использование таких систем (Google Analytics, Yandex.Metrica, Plerdy, Hotjar и др.) позволяет не только отслеживать ключевые показатели эффективности (KPI) [3], но и выявлять узкие места в пользовательском пути, влияющие на достижение целевых действий.

В процессе сравнительного анализа существующих решений было установлено, что большинство традиционных инструментов аналитики веб-ресурсов ориентированы преимущественно на фиксацию статистических показателей – количество посещений, глубину прокрутки, источники трафика и коэффициент отказов. Однако такие данные отражают лишь поверхностные аспекты взаимодействия и не позволяют в полной мере оценить поведенческие закономерности пользователей в контексте конкретных интерфейсных элементов. Недостаточная детализация событийного уровня данных затрудняет принятие обоснованных решений о внесении структурных или контентных изменений в веб-ресурс.

Для преодоления указанных ограничений целесообразно применять технологию событийного анализа [4], позволяющую исследовать взаимодействие пользователей с интерфейсом на более детальном уровне. В совокупности такие инструменты обеспечивают регистрацию действий – нажатий, прокруток, наведения курсора, выбора фильтров – и формирование на их основе аналитических карт активности. Ключевыми функциями данной аналитической модели являются: фиксация параметров взаимодействия в реальном времени, агреги-

рование данных в базу событий, построение тепловых карт и построение глубины прокрутки страницы как показателей вовлечённости пользователя.

Из вышеперечисленного становится ясно, что для повышения точности оценки эффективности веб-ресурсов требуется интеграция событийных данных с традиционными метриками производительности. Комплексный подход обеспечивает переход от абстрактных статистических показателей к поведенческим индикаторам эффективности, отражающим реальные закономерности восприятия и использования контента. В результате появляется возможность построения адаптивной модели.

3 Описание модели эффективности веб-ресурсов

Для систематизации описанных процессов и отражения взаимосвязей между уровнями аналитической обработки данных разработана модель комплексного применения методов веб-аналитики (рис. 1). Данная модель визуализирует последовательность этапов сбора, обработки и интерпретации поведенческих данных посетителей веб-сайта. Событийные показатели соединены с традиционными статистическими метриками. В структуре модели выделены функциональные блоки, соответствующие ключевым стадиям аналитического цикла.

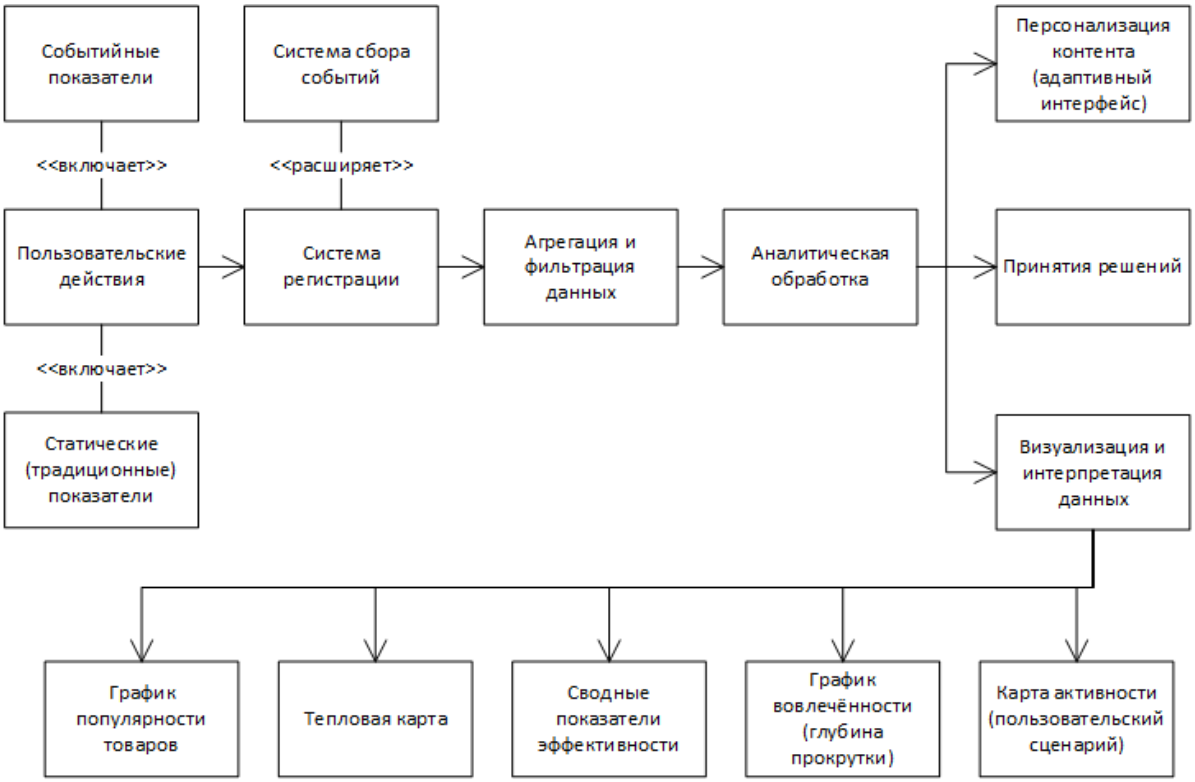

Рисунок 1 – Модель комплексного применения методов веб-аналитики

Рассмотрим детальнее интеграцию некоторых элементов информационно-аналитической системы.

Компонент каталога, отвечающий за отображение плитки с группой товаров на веб-странице, использует веб-хук, чтобы определить, когда данная секция попадает в зону видимости пользователя. Как только это происходит – вызывается функция, которая фиксирует поведенческое действие пользователя в базу данных. В результате собранные данные можно использовать для построения тепловой карты – графического инструмента, визуализирующего совокупные данные о поведении пользователей на веб-странице в виде цветовой шкалы, или анализа глубины прокрутки – метрики, выраженной в процентах, показывающей какую часть веб-страницы пользователь фактически просмотрел.

Для регистрации поведенческих действий, связанных с взаимодействием пользователя с элементами управления веб-интерфейса, применяется механизм отслеживания событий. Каждый элемент интерфейса, выполняющий функцию кнопки, снабжается обработчиком со-

бытий, фиксирующим факт нажатия и передающим соответствующие данные в информационно-аналитическую систему. Данный процесс реализуется через вызов функции. Полученные данные используются для анализа конверсии как конкретного элемента интерфейса, так и для оценки эффективности расположения элементов в общем, с последующим выявлением закономерностей в пользовательском поведении при выполнении целевых действий.

Анализ взаимодействия пользователей с системой фильтрации товарного каталога осуществляется посредством фиксации выбранных параметров фильтров и последовательности их применения. При каждом изменении состояния фильтра активируется функция, которая передает сведения о действии в базу данных. Регистрация таких аналитических данных позволяет выявить наиболее востребованные атрибуты товаров, оценить релевантность структуры фильтров и оптимизировать пользовательский сценарий навигации по каталогу. Выявленные метрики служат основой для анализа предпочтений пользователей и персонализации контента.

Анализ существующих решений показывает, что в традиционных системах оценка вовлеченности пользователей основывается на регистрации временных меток загрузки и закрытия страницы, а также событий изменения фокуса окна браузера. Последующая обработка этих данных позволяет сформировать метрики средней продолжительности сессии и времени взаимодействия с ключевыми элементами интерфейса.

Для анализа поведенческих паттернов применяется алгоритм построения графиков популярности товаров, реализующий ранжирование ассортимента по частоте просмотров карточек. Процедура включает извлечение данных о пользовательских действиях, подсчет просмотров по позициям, сортировку товаров по полученным показателям и визуализацию рейтинга в виде горизонтальной столбчатой диаграммы.

Выявленные ограничения рассмотренных подходов, в частности их ориентация на фиксацию статистических показателей без глубинного анализа поведенческих аспектов, обусловили необходимость синтеза комплексной модели, интегрирующей разрозненные данные для оптимизации управленческих решений.

4 Вопросы персональных данных в веб-аналитике

При интеграции инструментов веб-аналитики следует учесть то обстоятельство, что сбор и обработка поведенческих данных пользователей веб-сайта, таких как факты прокрутки, клики и время сессии, в соответствии со статьей 3 Федерального закона от 27.07.2006 № 152-ФЗ «О персональных данных» [5], квалифицируется как обработка персональных данных. Это связано с потенциальной возможностью идентификации субъекта через совокупность его поведенческих паттернов и предпочтений. Таким образом, осуществление описанных процедур аналитики, требует наличия легитимного основания, предусмотренного статьей 6 ФЗ-152.

Ключевым правовым основанием для обработки данных в целях веб-аналитики является получение предварительного согласия субъекта (пункт 1 части 1 статьи 6 ФЗ-152). Согласие на использование файлов cookie и сбор поведенческих метрик должно быть информированным, конкретным и сознательным, что на практике реализуется через механизм «cookie-баннеров».

Технические процедуры, такие как фиксация попадания секции каталога в область видимости с помощью хука или регистрация нажатий через функцию, направлены на сбор первичных данных о взаимодействии. Последующая обработка, включая агрегирование и обезличивание для построения тепловых карт или расчета глубины прокрутки, не отменяет необходимости первоначального правового основания. Аналогично, использование функции для ранжирования товаров на основе частоты просмотров представляет собой автоматизированное принятие решений на базе персональных данных, что требует прозрачности и раскрытия пользователю.

Следовательно, каждая из описанных аналитических операций должна осуществляться в рамках правового поля.

Выводы

Проведенное исследование демонстрирует недостаточность изолированного применения инструментов веб-аналитики для комплексной оценки эффективности веб-ресурсов. Установлено, что традиционные статистические показатели, фиксируя количественные параметры пользовательской активности, не позволяют идентифицировать причинно-следственные связи в поведенческих паттернах. В отличие от них, событийный анализ предоставляет детализированные данные о взаимодействии с интерфейсными элементами.

В результате работы синтезирована модель принятия решений, основанная на интеграции разноуровневых данных веб-аналитики. Ключевым преимуществом предложенного подхода является обеспечение целостного представления о функционировании веб-ресурса как динамической системы. Модель позволяет эмпирически обосновывать оптимизацию структурных, контентных и интерфейсных компонентов, что способствует переходу от эмпирического управления к научно обоснованному моделированию конверсионных процессов в условиях конкурентной цифровой среды.

Литература

1. Белоблоцкая Я.С., Искрова А.А., Новикова А.А. Системный подход в UX/UI дизайне / Я.С. Белоблоцкая, А.А. Искрова, А.А. Новикова // Электронные системы и технологии: материалы 58-й науч. конф. аспирантов, магистрантов и студентов БГУИР, Минск, 18–22 апр. 2022 г. / Белорус. гос. ун т информ. и радиоэлектроники; редкол.: Д. В. Лихаческий [и др.]. .– Минск, 2022. .– С. 698–700. .– Режим доступа: https://libeldoc.bsuir.by/handle/123456789/47398. .– Дата доступа: 12.10.2025.

2. Погодин, В. В. Методы и алгоритмы оценки пользовательского интерфейса (UI) и пользовательского опыта (UX) веб-сайтов / В. В. Погодин // Естественные и технические исследования в России и за рубежом: проблемы, пути совершенствования : Сборник научных статей, Краснодар, 28 апреля 2025 года. – Краснодар: ООО «КОНТУР-СП», 2025. – С. 180-186. – EDN WPXLDT.

3. Чебанов, И. М. Автоматизация ситуационного планирования KPI малого бизнеса / И. М. Чебанов // Электронный сетевой политематический журнал "Научные труды КубГТУ". – 2018. – № 3. – С. 625-634. – EDN XOPQRV.

4. Щербакова, А. А. Структурный анализ различных подходов к классификации event-мероприятий в событийном маркетинге и событийном туризме / А. А. Щербакова, И. В. Романов // Проблемы и перспективы развития спортивного ориентирования и активных видов туризма : материалы IV Всероссийской научно-практической конференции с международным участием, Смоленск, 04–05 апреля 2025 года. – Смоленск: Смоленский государственный университет спорта, 2025. – С. 144-155. – EDN UJJBCI.

5. О персональных данных : Федеральный закон от 27 июня 2006 года № 152 : [в редакции от 24 июня 2025 года]. – Доступ из справочно-правовой системы «КонсультантПлюс». – Текст : электронный.

Bubnov G.V., Grigoriev A.V. Synthesis of a decision-making model for optimizing a web resource based on integrated web analytics. *The article explores approaches to improving the efficiency of web resources through the analysis of user behavior, with a focus on single-page sites, where conversion critically depends on structural and content characteristics. A decision-making model based on the integration of complex web analytics data is proposed, which makes it possible to empirically justify the optimization of interface and content elements to increase conversion and achieve business targets.*

Keywords: *website content analysis system, information and analysis system, online commerce efficiency, conversion enhancement.*

УДК 004.855.5: 658.8.012.12

РАЗРАБОТКА И СРАВНЕНИЕ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ ОТТОКА КЛИЕНТОВ ТЕЛЕКОММУНИКАЦИОННОЙ КОМПАНИИ

Будур А.Р., Зори С.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программной инженерии» им. Л. П. Фельдмана
e-mail: anton.budur161@gmail.com, ik.ivt.rec@mail.ru

Будур А.Р., Зори С.А. Разработка и сравнение моделей машинного обучения для прогнозирования оттока клиентов телекоммуникационной компании. *В исследовании рассматривается задача разработки и сравнительного анализа моделей машинного обучения для прогнозирования оттока клиентов в телекоммуникационной отрасли. В результате проведенной работы доказано преимущество ансамблевых методов машинного обучения, в частности градиентного бустинга, перед традиционными алгоритмами для данной задачи.*
Ключевые слова: *машинное обучение, прогнозирование оттока, телекоммуникации, градиентный бустинг, анализ данных, классификация, Python.*

Введение

Современный телекоммуникационный рынок характеризуется исключительно высокой конкурентной средой и насыщенностью услуг. В таких условиях стоимость привлечения нового клиента существенно превышает расходы на удержание существующего [1]. Прогнозирование оттока клиентов становится стратегической задачей, решение которой позволяет компаниям минимизировать потери и увеличивать пожизненную ценность клиентской базы.

Задача прогнозирования оттока формализуется как проблема бинарной классификации, где на основе исторических данных необходимо определить вероятность перехода клиента в категорию ушедших. Основные методологические сложности заключаются в типичном дисбалансе классов и необходимости построения не только точной, но и интерпретируемой модели, позволяющей выявить причины ухода клиентов.

В данной работе представлено комплексное исследование, посвященное разработке и сравнению эффективности различных алгоритмов машинного обучения для решения задачи прогнозирования оттока. Использование языка Python и его библиотек для анализа данных обеспечивает воспроизводимость результатов и практическую применимость разработанного подхода.

Целью работы является создание комплексного аналитического пайплайна на языке Python, включающего генерацию синтетических данных, предобработку признаков, балансировку классов и обучение пяти различных алгоритмов классификации. Научная новизна заключается в комплексном подходе к оценке моделей с акцентом на интерпретируемость результатов и выявление ключевых факторов оттока. Практическая значимость определяется возможностью внедрения разработанного решения для своевременного выявления клиентов группы риска и разработки целевых стратегий их удержания.

1 Постановка задачи

Основной задачей настоящего исследования является разработка комплексного подхода к прогнозированию оттока клиентов телекоммуникационной компании с использованием методов машинного обучения. В рамках данной задачи формулируются следующие конкретные цели и требования.

На первом этапе требуется создать репрезентативный набор данных, адекватно отражающий специфику телекоммуникационного бизнеса. Поскольку реальные данные часто яв-

ляются коммерческой тайной и труднодоступны для исследований, необходимо разработать механизм генерации синтетических данных, сохраняющих ключевые закономерности и взаимосвязи, характерные для предметной области. Особое внимание уделяется включению именно бизнес-ориентированных признаков, таких как тип контракта, качество обслуживания и финансовые показатели, при этом демографические характеристики сознательно исключаются из анализа как менее значимые для оперативного управления.

Второй важной задачей становится проектирование и реализация сквозного аналитического пайплайна на языке Python, охватывающего все этапы работы с данными — от предобработки до построения прогнозных моделей. Пайплайн должен обеспечивать воспроизводимость результатов и возможность легкой модификации отдельных компонентов. Критически важным требованием является корректная обработка специфических проблем данных, характерных для задачи прогнозирования оттока, включая дисбаланс классов, наличие пропущенных значений и неоднородность типов признаков.

Третья задача заключается в проведении сравнительного анализа различных алгоритмов машинного обучения для выявления наиболее эффективного подхода к прогнозированию оттока. Сравнение должно проводиться по комплексному набору метрик, учитывающему как общую точность классификации, так и сбалансированность показателей precision и recall, что особенно важно в условиях неравного распределения классов.

Четвертой задачей является интерпретация результатов работы лучшей модели с идентификацией ключевых факторов, влияющих на вероятность оттока клиентов. Полученные инсайты должны иметь практическую ценность для бизнеса и служить основой для разработки конкретных мероприятий по удержанию клиентов.

Наконец, пятой задачей выступает разработка программной реализации, обеспечивающей удобство использования, модульность и возможность дальнейшего расширения функциональности. Код должен быть структурирован в виде отдельных логических блоков с четким разделением ответственности и сопровождаться подробной документацией.

Таким образом, поставленная задача носит комплексный характер и сочетает теоретические аспекты машинного обучения с практическими требованиями к разработке программного обеспечения для решения реальной бизнес-проблемы телекоммуникационной отрасли.

2 Анализ литературы

Проблема прогнозирования оттока клиентов интенсивно изучается в научной литературе. Традиционные статистические методы, такие как логистическая регрессия, долгое время оставались стандартным инструментом для решения подобных задач [2]. Однако с развитием вычислительных мощностей и алгоритмов машинного обучения все большее распространение получают ансамблевые методы.

Исследования показывают, что методы на основе деревьев решений, такие как случайный лес и градиентный бустинг, демонстрируют повышенную точность в задачах классификации с табличными данными [3]. В частности, алгоритм XGBoost, предложенный Тяньцзи Ченом и Карлосом Гестрином, зарекомендовал себя как один из наиболее эффективных инструментов в соревнованиях по анализу данных [4].

Важным аспектом при работе с данными об оттоке является проблема дисбаланса классов, когда количество ушедших клиентов значительно меньше количества лояльных. Для решения этой проблемы широко применяется метод SMOTE, который генерирует синтетические примеры миноритарного класса [5].

3 Методология исследования

Общая методология исследования строится на последовательной реализации этапов аналитического пайплайна. Первоначальный этап включает проектирование и генерацию синтетического набора данных, адекватно отражающего специфику телекоммуникационной отрасли. Синтетические данные содержат наиболее релевантные бизнес-показатели, сгруп-

пированные в тематические категории: контрактные параметры, финансовые показатели, характеристики услуг и метрики качества обслуживания. Демографические характеристики сознательно исключены из анализа в связи с их меньшей значимостью для оперативного управления бизнес-процессами.

Следующий этап предполагает тщательную предобработку данных, включая обработку пропущенных значений через импутацию медианными значениями для числовых признаков и модальными значениями для категориальных переменных. Категориальные признаки преобразуются в числовой формат с применением кодировщика LabelEncoder. Числовые признаки стандартизируются с использованием метода StandardScaler для приведения к единому масштабу.

Особое внимание уделяется проблеме дисбаланса классов. Для ее решения применяется алгоритм SMOTE, который синтезирует новые примеры миноритарного класса в пространстве признаков. Это позволяет улучшить способность моделей к обучению на недостаточно представленном классе ушедших клиентов.

Для сравнительного анализа выбраны пять алгоритмов машинного обучения: логистическая регрессия как базовый метод, случайный лес как представитель бэггинговых алгоритмов, градиентный бустинг и XGBoost как представители бустинговых подходов, а также метод опорных векторов с линейным ядром. Оценка эффективности моделей проводится с использованием кросс-валидации на пяти фолдах и тестированием на выделенной тестовой выборке. В качестве оценочных метрик используются accuracy, precision, recall, F1-мера и ROC-AUC.

4 Реализация и программное обеспечение

Разработка программного обеспечения выполнена на языке Python с использованием специализированных библиотек для анализа данных и машинного обучения. Структура проекта инкапсулирована в класс CustomerChurnPredictor, что обеспечивает модульность и повторное использование кода.

```
class CustomerChurnPredictor:
    """
    Класс для прогнозирования оттока клиентов телеком-компании
    """

    def __init__(self):...

    def _initialize_russian_names(self):...

    def analyze_business_factors(self):...

    def _create_synthetic_data(self):...

    def load_and_preprocess_data(self, data_path):...

    def _handle_missing_values(self):...

    def _encode_categorical_features(self):...

    def exploratory_data_analysis(self):...

    def prepare_data(self):...

    def initialize_models(self):...

    def train_and_evaluate_models(self, X_train, X_test, y_train, y_test):...

    def compare_models(self):...

    def feature_importance_analysis(self, feature_names):...
```

Рисунок 1 – Класс CustomerChurnPredictor

Основной функционал класса включает методы для загрузки и предобработки данных, проведения разведочного анализа, обучения моделей и оценки их эффективности. Для

генерации синтетических данных используется библиотека NumPy, обеспечивающая создание правдоподобных распределений телекоммуникационных показателей. Визуализация результатов выполняется с применением библиотек Matplotlib и Seaborn.

Процедура инициализации моделей реализована через словарную структуру, что позволяет единообразно управлять процессом обучения и прогнозирования. Для каждой модели сохраняются рассчитанные метрики и прогнозы, что способствует последующее сравнительное тестирование.

Важным аспектом реализации является обработка категориальных признаков. Метод _encode_categorical_features автоматически идентифицирует нечисловые столбцы и преобразует их с сохранением объектов кодировщика для возможного последующего использования. Метод _handle_missing_values обеспечивает надежную обработку пропущенных значений отдельно для числовых и категориальных признаков.

```
def _encode_categorical_features(self):
    """Кодирование категориальных признаков"""
    categorical_columns = self.df.select_dtypes(include=['object']).columns

    for col in categorical_columns:
        if col != 'ID_клиента':  # Пропускаем ID клиента
            self.label_encoders[col] = LabelEncoder()
            self.df[col] = self.label_encoders[col].fit_transform(
                self.df[col].astype(str)
            )
```

Рисунок 2 – Функция _encode_categorical_features

```
def _handle_missing_values(self):
    """Обработка пропущенных значений с русскими названиями"""
    print(f"\nПропущенные значения до обработки:")
    missing_before = self.df.isnull().sum()
    missing_before = missing_before[missing_before > 0]  # Только колонки с пропусками
    if len(missing_before) > 0:
        print(missing_before)
    else:
        print("Пропущенных значений не обнаружено")

    # Заполнение числовых признаков медианой
    numeric_columns = self.df.select_dtypes(include=[np.number]).columns
    self.df[numeric_columns] = self.df[numeric_columns].fillna(
        self.df[numeric_columns].median()
    )

    # Заполнение категориальных признаков модой
    categorical_columns = self.df.select_dtypes(include=['object']).columns
    for col in categorical_columns:
        self.df[col] = self.df[col].fillna(self.df[col].mode()[0])

    print(f"\nПропущенные значения после обработки:")
    missing_after = self.df.isnull().sum()
    missing_after = missing_after[missing_after > 0]  # Только колонки с пропусками
    if len(missing_after) > 0:
        print(missing_after)
    else:
        print("Пропущенных значений не обнаружено")
```

Рисунок 3 – Функция _handle_missing_values

Балансировка классов реализована с использованием библиотеки imbalanced-learn, конкретно для алгоритма SMOTE. Это позволяет улучшить качество обучения моделей на несбалансированных данных без необходимости ручной настройки весов классов.

5 Анализ результатов

Проведенный разведочный анализ данных выявил характерные закономерности, согласующиеся с известными бизнес-инсайтами телекоммуникационной отрасли. Распределение целевой переменной показало уровень оттока приблизительно восемнадцать процентов, что соответствует типичным значениям для подобных наборов данных.

Корреляционный анализ продемонстрировал выраженную положительную связь между вероятностью оттока и такими факторами как помесячный тип контракта, частотой обращения в службу поддержки и задержки платежей. Отрицательная корреляция обнаружена с наличием услуги технической поддержки, долгосрочностью обслуживания и количеством подключенных услуг.

Сравнительная оценка моделей выявила явное преимущество ансамблевых методов перед линейными подходами. Модель градиентного бустинга достигла наивысших значений по комплексной метрике F1, что свидетельствует о ее сбалансированной эффективности в отношении как точности, так и полноты прогнозирования. Модель XGBoost показала схожие результаты, незначительно уступая по большинству метрик.

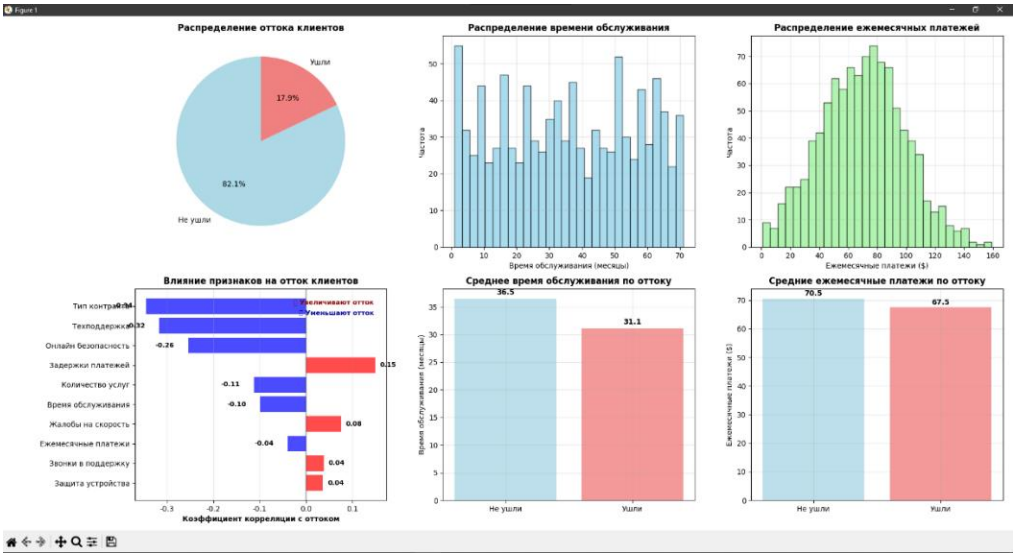

Рисунок 4 – Анализ данных телекоммуникационной отрасли

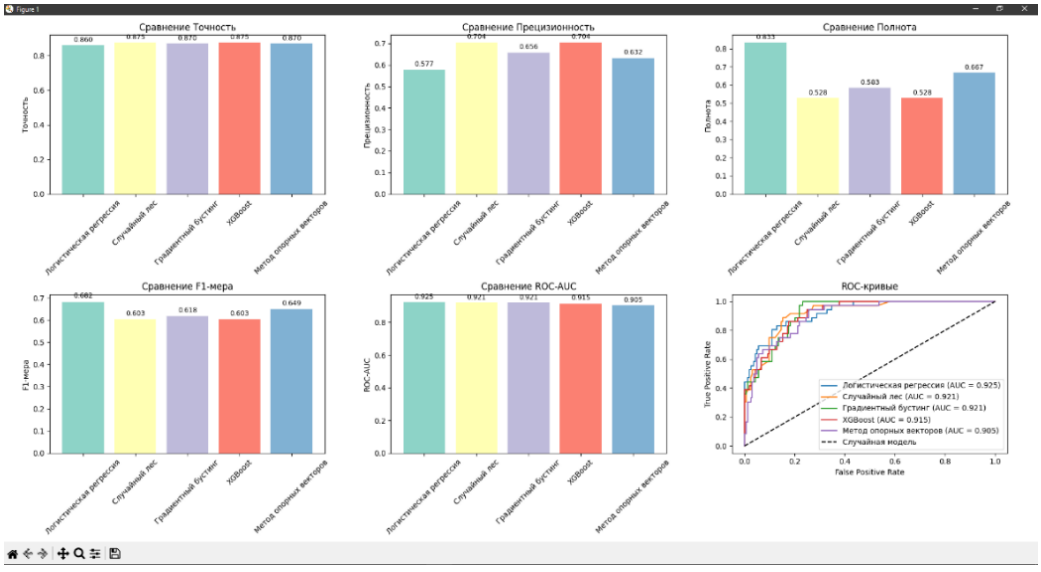

Рисунок 5 – Сравнительная оценка моделей

Логистическая регрессия и метод опорных векторов продемонстрировали ожидаемо более низкие результаты, что объясняется нелинейным характером зависимостей в данных.

При этом все модели показали достаточно высокое значение ROC-AUC, превышающее 0.85, что подтверждает общую адекватность подхода.

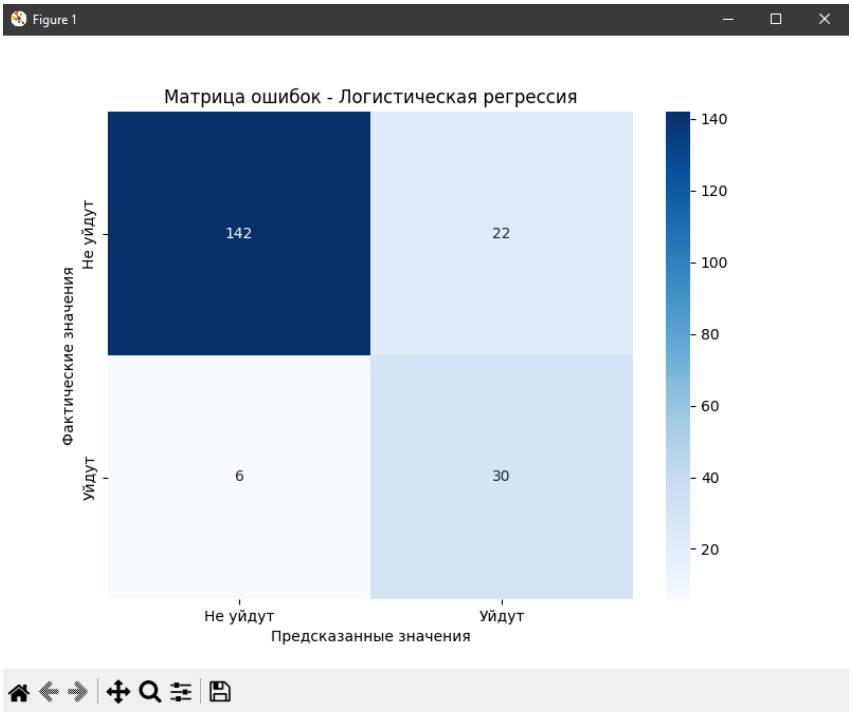

Рисунок 6 – Логическая регрессия

Выводы

В результате проведенной работы разработан и протестирован комплексный подход к прогнозированию оттока клиентов телекоммуникационных компаний. Доказано преимущество ансамблевых методов машинного обучения, в частности градиентного бустинга, перед традиционными алгоритмами для данной задачи.

Практическая значимость работы заключается в возможности непосредственного внедрения, разработанного пайплайна в бизнес-процессы телекоммуникационных компаний для своевременного выявления клиентов группы риска. Выявленные ключевые факторы оттока предоставляют менеджменту конкретные направления для улучшения сервиса и разработки нацеленных программ удержания.

Перспективы дальнейших исследований связаны с интеграцией дополнительных источников данных, включая поведенческие паттерны использования услуг и информацию о взаимодействии с каналами обслуживания. Также представляет интерес адаптация подхода для работы в режиме реального времени и разработка системы рекомендаций на основе прогнозной модели.

Литература

1. Райххелд, Ф. Ф., & Сассер, У. Э. (1990). Нулевые дефекции: Качество приходит в сферу услуг. Гарвард Бизнес Ревью, 68(5), 105-111. [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://hbr.org/1990/09/zero-defections-quality-comes-to-services.

2. Хэсти, Т., Тибширани, Р., & Фридман, Дж. (2009). Элементы статистического обучения: интеллектуальный анализ данных, вывод и прогнозирование. Серия Springer по статистике. [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://link.springer.com/book/10.1007/978-0-387-84858-7.

3. Брейман, Л. (2001). Случайные леса. Машинное обучение, 45(1), 5-32. [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://arxiv.org/abs/1407.7502.

4. Чен, Т., & Гестрин, К. (2016). XGBoost: Масштабируемая система повышения деревьев. Труды 22-й Международной конференции ACM SIGKDD по обнаружению знаний

и интеллектуальному анализу данных. [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://dl.acm.org/doi/10.1145/2939672.2939785.

5. Чавла, Н. В., Бойер, К. В., Холл, Л. О., & Кегельмайер, В. П. (2002). SMOTE: Метод передискретизации синтетического меньшинства. Журнал исследований искусственного интеллекта, 16, 321-357. [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://jair.org/index.php/jair/article/view/10302.

Budur A.R., Zori S.A. Development and Comparison of Machine Learning Models for Predicting Customer Churn in a Telecommunications Company. *This study focuses on the development and comparative analysis of machine learning models for predicting customer churn in the telecommunications industry. As a result of this work, the advantage of ensemble machine learning methods, in particular gradient boosting, over traditional algorithms for this task has been proven.*
Keywords: *machine learning, customer churn prediction, telecommunications, gradient boosting, data analysis, and classification, Python.*

УДК 004.8

ИСПОЛЬЗОВАНИЕ ПРЕДОБУЧЕННЫХ МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ НА ОСНОВЕ ФОРМАТА ONNX

Винокуров К.А., Григорьев А.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: oleksandr.vinokurov@yandex.ru, grigorievalvl@mail.com

Винокуров К.А., Григорьев А.В. Использование предобученных моделей машинного обучения в интеллектуальных системах на основе формата ONNX. *В работе рассматриваются теоретические основы применения предобученных моделей машинного обучения в интеллектуальных системах. Описан формат ONNX как универсальный стандарт обмена моделями, обеспечивающий переносимость между различными фреймворками. Проведён сравнительный анализ с другими форматами представления моделей, приведена общая архитектура среды выполнения ONNX Runtime и показаны преимущества интеграции в прикладные программные решения.*
Ключевые слова: *машинное обучение, ONNX, интеллектуальные системы, нейронные сети, переносимость, ONNX Runtime.*

Введение

Интеллектуальные системы представляют собой класс программных комплексов, использующих методы искусственного интеллекта (ИИ) и машинного обучения (МО) для анализа данных, прогнозирования и принятия решений. Их эффективность напрямую зависит от качества и производительности используемых моделей МО.

Одним из ключевых направлений развития интеллектуальных технологий является использование предобученных моделей, которые позволяют сократить время и вычислительные затраты при внедрении ИИ-алгоритмов. Вместе с тем возникает проблема совместимости и переносимости моделей между различными фреймворками.

Для решения этой задачи был разработан формат ONNX (Open Neural Network Exchange), обеспечивающий стандартный способ описания моделей нейронных сетей, независимо от среды их обучения.

Целью настоящей работы является анализ возможностей применения предобученных моделей в формате ONNX для интеграции в программное обеспечение интеллектуальных систем.

1 Постановка задачи

Современные интеллектуальные системы часто строятся как многоуровневые программные комплексы, включающие подсистемы обработки данных, нейросетевые модули и интерфейсы взаимодействия с пользователем. Используемые модели МО обучаются на специализированных платформах (PyTorch, TensorFlow, Keras), однако для их практического применения требуется перенос в производственные среды — например, в системы, написанные на C#, C++ или Java.

Основная задача заключается в том, чтобы обеспечить выполнение обученной модели вне зависимости от среды обучения и аппаратной платформы.

Формат ONNX решает эту задачу, предоставляя единое представление вычислительного графа нейронной сети. Каждый узел графа соответствует операции (оператору), а связи между ними описывают потоки данных. Это позволяет абстрагироваться от внутреннего устройства конкретного фреймворка.

2 Анализ существующих решений

До появления ONNX существовало несколько форматов представления моделей (таблица 1).

Таблица 1. Сравнение распространённых форматов представления моделей МО

Формат	Основная среда	Поддержка аппаратных ускорителей	Кроссплатформенность	Открытость стандарта
TensorFlow SavedModel	TensorFlow	Высокая (GPU, TPU)	Средняя	Да
TorchScript	PyTorch	Высокая (CUDA)	Средняя	Да
CoreML	Apple	Только устройства Apple	Низкая	Частично
ONNX	Универсальный	Высокая (CPU, GPU, DirectML, TensorRT и др.)	Высокая	Полная

Из таблицы видно, что формат ONNX обладает наибольшей универсальностью и пригоден для интеграции в программные системы любого уровня. Он поддерживается множеством фреймворков и библиотек, включая PyTorch, TensorFlow, Scikit-Learn и XGBoost.

Общий формат хранения нейросети помогает сэкономить время, лишить привязки к конкретному фреймворку и дает возможность больше экспериментировать с инфраструктурой (см. рис. 1).

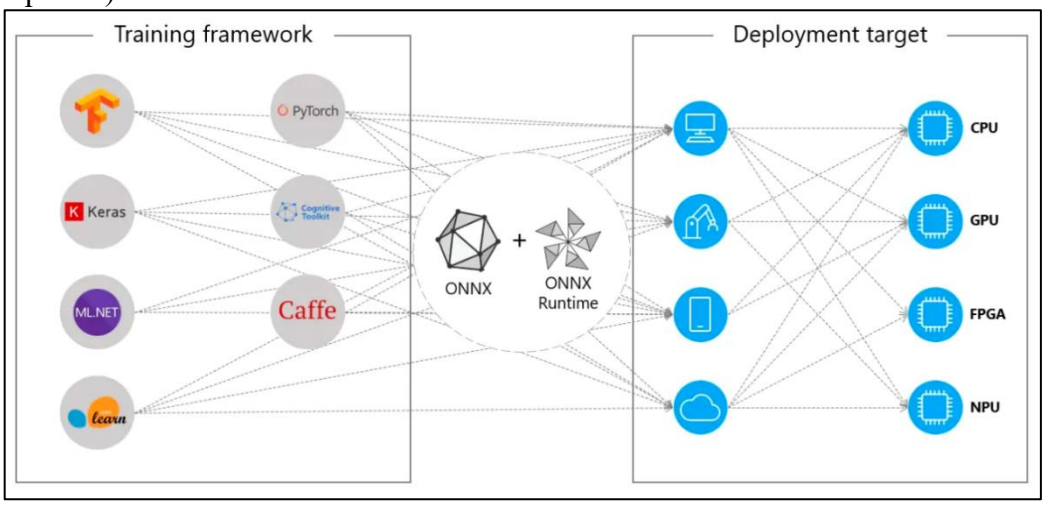

Рисунок 1 – ONNX как связь между обучением и инференсом

Также формат ONNX закрывает потребность конвертировать представления нейросетей между фреймворками (см. рис. 2), при этом разрабатывать и поддерживать по конвертеру на каждую пару фреймворков.

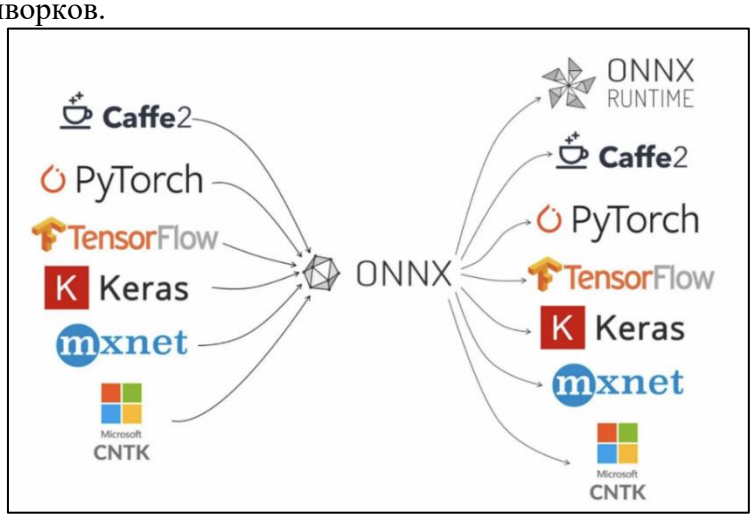

Рисунок 2 – ONNX конвертируется в обе стороны с большинством параметров

3 Архитектура и интеграция ONNX Runtime

Архитектура среды ONNX Runtime включает следующие основные компоненты:

1. Graph Optimizer – выполняет упрощение и оптимизацию вычислительного графа.
2. Execution Provider – обеспечивает взаимодействие с вычислительными устройствами (CPU, CUDA, DirectML, ROCm).
3. Inference Session – выполняет загрузку и исполнение модели.

Схематично взаимодействие можно описать выражением:

$$R = f(M, D), \tag{1}$$

где R– результат инференса (предсказание модели), M– модель в формате ONNX,D– входные данные.

Пример вызова инференса на языке C# может быть представлен следующим образом (см. рис. 3):

```
using Microsoft.ML.OnnxRuntime;
using Microsoft.ML.OnnxRuntime.Tensors;

using var session = new InferenceSession("model.onnx");
var input = NamedOnnxValue.CreateFromTensor("input", inputTensor);
using var results = session.Run(new[] { input });
```

Рисунок 3 – Пример вызова инференса на языке C#

Такой способ исполнения позволяет выполнять модели без зависимости от Python и библиотек глубокого обучения, что особенно важно для встраиваемых систем и приложений с повышенными требованиями к производительности.

4 Преимущества и ограничения подхода

Преимущества ONNX:

- независимость от среды обучения и языка реализации;
- кроссплатформенность и переносимость;
- аппаратное ускорение и оптимизация исполнения;
- поддержка большинства популярных фреймворков;
- открытая спецификация и активное сообщество.

Ограничения:

- частичная поддержка специфичных операторов некоторых фреймворков;
- сложность конвертации больших трансформерных моделей;
- необходимость настройки среды выполнения для оптимального ускорения.

Выводы

Таким образом, использование формата ONNX в интеллектуальных системах обеспечивает единый подход к интеграции предобученных моделей машинного обучения в прикладные программные продукты. Теоретически ONNX реализует концепцию инвариантности модели относительно среды исполнения, что упрощает построение распределённых и кроссплатформенных интеллектуальных решений.

В перспективе планируется расширение возможностей ONNX для моделей с гибридной структурой, развитие инструментов оптимизации и внедрение поддержки новых вычислительных платформ. Это делает ONNX одним из ключевых инструментов построения современных программных комплексов искусственного интеллекта.

Литература

1. Jajal, P.; Jiang, W.; Tewari, A. *Analysis of Failures and Risks in Deep Learning Model Converters: A Case Study in the ONNX Ecosystem* // [Электронный ресурс]. – 2023. – Режим доступа: https://arxiv.org/abs/2303.17708. – Загл. с экрана.

2. Openja, M.; Nikanjam, A.; Haj Yahmed, A.; Khomh, F.; Zhen, J. *An Empirical Study of Challenges in Converting Deep Learning Models* // [Электронный ресурс]. – 2022. – Режим доступа: https://arXiv.org/abs/2206.14322. – Загл. с экрана.

3. Microsoft Corporation. *ONNX Runtime Documentation* [Электронный ресурс]. – Режим доступа: https://onnxruntime.ai/docs. – Загл. с экрана.

4. ONNX Community / Linux Foundation. *ONNX: Open Neural Network Exchange* [Электронный ресурс]. – Режим доступа: https://onnx.ai. – Загл. с экрана.

Vinokurov K.A., Grigoriev A.V. Use of Pre-trained Machine Learning Models in Intelligent Systems Based on the ONNX Format. *The paper presents theoretical aspects of applying pre-trained machine learning models in intelligent systems. The ONNX format is described as a universal model exchange standard providing interoperability between frameworks. A comparative analysis of different model representation formats is given, and the ONNX Runtime architecture is outlined with its advantages for practical software integration.*

Keywords: *machine learning, ONNX, intelligent systems, neural networks, portability, ONNX Runtime*.

УДК 004.8

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА ДЛЯ АВТОМАТИЗАЦИИ ПРОВЕРКИ СТУДЕНЧЕСКИХ РАБОТ

Гончар Д.Ю.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: goncardaniil01@gmail.com

Гончар Д.Ю. Применение технологий обработки естественного языка для автоматизации проверки студенческих работ. *В статье рассматривается проблема автоматизации проверки текстовых работ студентов в условиях массового высшего образования. Предложен подход на основе моделей обработки естественного языка (NLP) для оценки смысловой составляющей и выявления плагиата. Разработан прототип системы, показавший эффективность при анализе реферативных работ.*
Ключевые слова: *обработка естественного языка, искусственный интеллект, автоматизация образования, проверка работ, машинное обучение.*

Введение

Рост численности студентов и объема учебной нагрузки преподавателей высшей школы актуализирует задачи автоматизации рутинных процессов [1]. Одним из таких процессов является проверка письменных работ студентов (эссе, рефератов, отчетов по практике), которая требует значительных временных затрат и подвержена фактору субъективности. Современные технологии искусственного интеллекта, в частности, направления обработки естественного языка (Natural Language Processing, NLP), предлагают инструменты для создания систем автоматизированной оценки [2].

Целью данной работы является разработка и апробация подхода к автоматической проверке студенческих работ на основе NLP-моделей для оценки релевантности содержания заданной теме и выявления текстовых заимствований.

1 Постановка задачи

Анализ современного состояния проблемы автоматизации проверки учебных работ выявил существенный разрыв между возможностями существующих систем и реальными потребностями образовательного процесса. Большинство коммерческих и академических решений, такие как «Антиплагиат», Turnitin и другие, ориентированы преимущественно на выявление текстовых заимствований и проверку формальных параметров текста [3]. При этом ключевая содержательная компонента — оценка того, насколько глубоко и полно студенческая работа раскрывает заданную тему, — остается прерогативой преподавателя и требует его непосредственного участия.

Существующие системы, основанные на анализе ключевых слов и их частотности (методы TF-IDF, Bag of Words), не способны адекватно оценивать семантическую связность и смысловую целостность текста. Они могут идентифицировать работу как релевантную при простом частом употреблении терминов, даже если эти термины используются вне контекста или сама работа не соответствует заданной теме. Обратная ситуация также возможна: работа, глубоко раскрывающая тему, но использующая сложные перифразы и синонимичные конструкции, может быть ошибочно классифицирована как нерелевантная.

Таким образом, выделяется основная проблема: отсутствие надежного инструмента для автоматизированной оценки смыслового соответствия студенческой работы заявленной тематике. Для решения данной проблемы необходимо создать систему, способную анализировать семантику текста, а не только его лексико-статистические характеристики.

В связи с этим формулируются следующие конкретные задачи исследования:

1. Провести сравнительный анализ современных NLP-моделей (включая BERT, SBERT, RoBERTa) для выявления наиболее подходящей архитектуры для задачи семантического сравнения учебных текстов.
2. Разработать алгоритм для векторного представления текстов, который бы учитывал не только отдельные семантические векторы предложений, но и общую структуру документа.
3. Реализовать модуль многокритериального сравнения векторных представлений, включающий оценку косинусного сходства, евклидова расстояния и специализированных метрик для образовательных текстов.
4. Создать прототип системы, интегрирующий разработанный алгоритм семантического анализа с традиционными методами проверки на заимствования и грамматику.
5. Провести апробацию прототипа на репрезентативной выборке студенческих работ различных жанров (рефераты, эссе, отчеты) и выполнить сравнительный анализ эффективности системы с оценками экспертов-преподавателей.

2 Основной материал

2.1 Выбор и обоснование метода семантического анализа

Для решения задачи семантического анализа рассматривались различные подходы, основанные на NLP. Классические методы, такие как TF-IDF и Bag of Words, не учитывают семантику и контекст слов, что приводит к низкой точности при оценке смысловой близости текстов [4]. В качестве базового метода был выбран подход на основе векторных представлений текста (эмбеддингов), получаемых с помощью предобученных нейросетевых моделей.

После сравнительного анализа была выбрана модель sentence-transformers/all-MiniLM-L6-v2 из библиотеки Sentence Transformers. Данная модель специально дообучалась для задач семантического сравнения предложений и коротких текстов. Она преобразует входной текст в вектор размерностью 384 в нормализованном векторном пространстве, где косинусное расстояние между векторами напрямую отражает их семантическую близость. Выбор обусловлен компромиссом между качеством (по данным разработчиков, модель показывает результаты, близкие к более крупным аналогам, таким как BERT-base) и вычислительной эффективностью, что важно для развертывания в образовательной среде [5].

2.2 Архитектура разработанной системы

Прототип системы автоматизированной проверки был реализован на языке Python и состоит из следующих взаимосвязанных модулей (рисунок 1):

Модуль предобработки данных. Выполняет очистку текста от HTML-разметки (если таковая имеется), приведение к нижнему регистру, удаление стоп-слов и знаков препинания с использованием библиотеки NLTK. Также выполняется лемматизация – приведение слов к их нормальной словарной форме.

Модуль векторного представления. На основе выбранной модели all-MiniLM-L6-v2 преобразует очищенный текст студенческой работы и корпус эталонных документов (тематические статьи, учебные пособия, конспекты лекций) в семантические векторы.

Модуль семантического анализа. Вычисляет метрику косинусного сходства между вектором студенческой работы и каждым из векторов эталонного корпуса. Итоговая оценка релевантности R вычисляется как максимальное значение косинусного сходства среди всех эталонных документов:

$$\mathrm{R} = \max(\cos(\vec{V_{work}}, \vec{V_{ref_i}})) \quad (1)$$

где $\vec{V_{work}}$ – вектор студенческой работы, $\vec{V_{ref_i}}$ – вектор i-го эталонного документа. Пороговое значение $R_{threshold} = 0.65$ было определено эмпирически в ходе предварительного тестирования на отдельной валидационной выборке.

1. Модуль проверки на заимствования. Использует комбинацию методов:
   1.1. Анализ шинглов (последовательностей из 5-7 слов) для выявления дословных совпадений внутри проверяемой выборки работ.

1.2. Сравнение с общедоступными интернет-источниками через поисковый API для выявления заимствований из открытых ресурсов.

2. Модуль формирования отчета. Генерирует итоговый отчет для преподавателя, который включает: итоговый балл релевантности, выделенные фрагменты с потенциальным плагиатом, визуализацию семантической близости к эталонным темам.

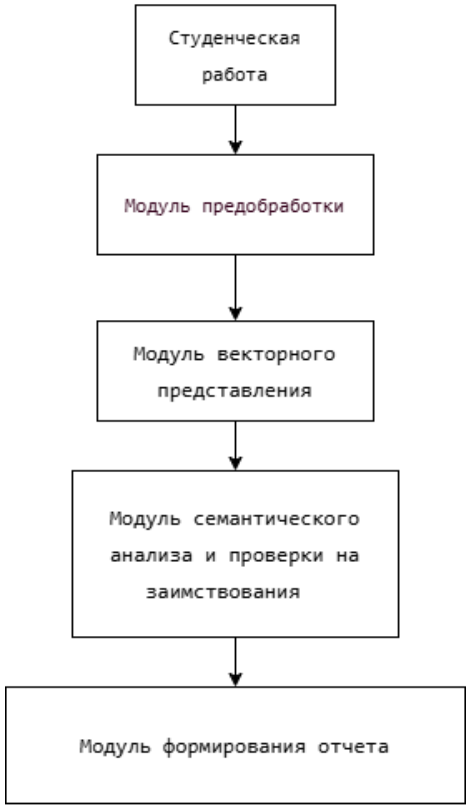

Рисунок 1 - Архитектура системы автоматизированной проверки студенческих работ

2.3 Эксперимент и анализ результатов

Для апробации системы была сформирована тестовая выборка, состоящая из 50 студенческих рефератов по теме «Машинное обучение». Объем каждого реферата составлял от 5 до 8 страниц. В качестве эталонного корпуса использовались 20 научных статей и главы из учебников, посвященных данной теме.

Результаты автоматической проверки сравнивались с экспертными оценками двух преподавателей. Консенсусная оценка экспертов считалась «золотым стандартом». Результаты эксперимента представлены в таблице 1.

Таблица 1. Результаты апробации системы автоматической проверки

Критерий оценки	**Точность системы, %**	**Полнота системы, %**	**Согласованность с экспертом (Cohen's Kappa)**
Релевантность теме (R ≥ 0.65)	88	85	0.81
Выявление дословного плагиата	94	90	0.89

Качество работы системы оценивалось по метрикам точности (precision) и полноты (recall). Точность 88% в задаче оценки релевантности означает, что из всех работ, которые система пометила как релевантные, 88% действительно были таковыми по мнению экспертов. Полнота 85% показывает, что система нашла 85% от всех релевантных работ в выборке. Высокое значение каппы Коэна (0.81) свидетельствует о почти полном согласии между системой и экспертом.

Наиболее частые ошибки системы были связаны с работами, находившимися на границе установленного порога релевантности (значения R в диапазоне 0.60–0.70). Анализ этих кейсов показал, что система испытывала трудности с работами, где использовалась сложная терминология с полным переформулированием смысла, а также с текстами, имевшими плохую структуру и логическую связность.

Выводы

Таким образом, в работе предложен и реализован подход к автоматизации проверки студенческих работ, сочетающий методы семантического анализа и выявления плагиата. Разработанный прототип системы показал высокую точность и согласованность с оценкой эксперта. Перспективы дальнейших исследований видятся в адаптации системы для проверки более сложных форматов работ (например, курсовых проектов), а также в использовании более крупных и специализированных языковых моделей для повышения точности смыслового анализа.

Литература

1. Федоров И.Б. Цифровая трансформация высшего образования: вызовы и возможности/ И.Б. Федоров, С.М. Ковалев // Университетское управление: практика и анализ. – 2021. – Т. 25. – № 1. – С. 32-45.

2. Девятков В.ВПрименение искусственного интеллекта в образовании / В.В. Девятков, А.В. Соколов // Искусственный интеллект и принятие решений. – 2022. – № 2. – С. 56-67.

3. Обзор систем проверки академических текстов [Электронный ресурс] / EdCrunch Review. – Режим доступа: https://edcrunch.review/obzor-sistem-proverki-akademicheskih-tekstov/ – Загл. с экрана.

4. Jurafsky, D. Speech and Language Processing. – 3rd ed. / D., Jurafsky, J.H. Martin, Speech – Prentice Hall, 2023. – 412 p.

5. Reimers, N. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks/ N., Reimers, I. Gurevych // Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. – 2019. – P. 3982-3992.

Gonchar D.Y. Application of Natural Language Processing Technologies for Automating the Checking of Student Papers. *The article addresses the problem of automating the checking of students' textual works in the context of mass higher education. An approach based on natural language processing (NLP) models for assessing semantic content and detecting plagiarism is proposed. A prototype system was developed, demonstrating effectiveness in analyzing abstract-type works.*

Keywords*: natural language processing, artificial intelligence, education automation, paper grading, machine learning.*

УДК 004.6

НОВАЯ РЕАЛЬНОСТЬ BI: ОБЗОР И СРАВНЕНИЕ РОССИЙСКИХ ПЛАТФОРМ ПОСЛЕ УХОДА ЗАРУБЕЖНЫХ АНАЛОГОВ

Дегтярёв Т.Р.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: degtyarevtaras3@gmail.ru

Дегтярёв Т.Р. Новая реальность BI: обзор и сравнение российских платформ после ухода зарубежных аналогов. *Данное исследование посвящено проблеме импортозамещения BI-платформ в России, актуальной в свете ухода ведущих иностранных аналогов. В статье представлен обзор наиболее популярных и функциональных российских платформ для анализа данных, выявлены их ключевые особенности и проведен сравнительный анализ характеристик.*
Ключевые слова: *анализ данных, BI-платформы, business Intelligence, сравнение BI-платформ, системы бизнес – анализа, рынок платформ анализа данных.*

Введение

В эпоху цифровой трансформации данные стали одним из наиболее ценных активов для организаций любого масштаба, а способность эффективно обрабатывать и анализировать их превратилась в ключевой фактор конкурентоспособности и успешного развития. Анализ данных позволяет выявлять скрытые закономерности, прогнозировать будущие тенденции, оптимизировать бизнес-процессы и принимать обоснованные, основанные на фактах решения.

На сегодняшний день, в условиях современных политических вызовов, российский рынок столкнулся с уходом множества ключевых зарубежных платформ, включая ведущие BI-системы, такие как Power BI, Tableau и Qlik. Прекращение их деятельности на территории России поставило перед многими компаниями задачу поиска и внедрения отечественных аналогов для обеспечения непрерывности бизнес-функций.

Целью данной работы является проведение обзора российских BI-платформ, призванных заменить ушедшие зарубежные аналоги. В рамках исследования будет произведен детальный разбор и сравнение ключевых характеристик этих отечественных решений, что позволит выявить их особенности применения и определить оптимальные сферы использования.

Постановка задачи

BI-системы - автоматизированные аналитические системы для сбора большого объема данных из разных источников с целью их дальнейшей обработки и представления в виде наглядных отчетов. BI-системы помогают определять закономерности в данных, находить слепые пятна и точки роста в бизнесе, оптимизировать рабочие процессы и принимать эффективные решения

Принцип работы систем BI разделяется на три ключевых этапа [1]:

- Сбор данных - получение информации из разнообразных источников, включая корпоративные системы и внешние ресурсы.
- Обработка и систематизация - трансформация сырых данных в структурированный и доступный формат, подготовленный для анализа.
- Визуализация: Представление обработанных данных в виде понятных и интерактивных отчетов, диаграмм и дашбордов.

При выборе BI-платформы критически важен анализ ее технологических возможностей. Хотя конкретные требования к технологическим аспектам могут варьироваться для различных ролей пользователей, фундаментальным требованием для любой платформы яв-

ляется надежная реализация всех трех вышеперечисленных этапов.

Анализ

Составим список критериев, покрывающих большую часть технического функционала, необходимого для качественной работы в программе. Критерии учитывают наличие функционала по обработке и визуализации данных, скорость разработки отчетов и сложность миграции на новую платформу. [2]

Перечислим критерии:

1. разнообразие и дизайн встроенных визуализаций
2. форматирование и кастомизация дэшборда
3. загрузка и трансформация данных
4. создание продвинутых расчетных мер и формул
5. интерактивные элементы, создание "второго слоя" данных

Основываясь на уже проведенных исследованиях специалистов в данной области, отзывах аналитиков и фирм, выделим 4 лидирующие платформы: Alpha BI, Yandex DataLens, PolyAnalyst, Форсайт.

Alpha BI – современная российская BI-система, созданная как быстрый по внедрению аналог Power BI. Разработчиком выступает компания «Барс Груп». Основная идея – облегчить переход с зарубежных BI-платформ на отечественную с минимальным дискомфортом для пользователей. [3]

Платформа Alpha BI отличается своей ориентацией на быстрое развертывание и простоту использования, что делает ее привлекательной для компаний, не располагающих обширными командами дата-аналитиков. Она стремится обеспечить интеграцию с популярными в России источниками данных, такими как 1С, а также с различными базами данных и облачными хранилищами. Ключевой особенностью является наличие гибкого конструктора дашбордов и отчетов, позволяющего пользователям без глубокого программирования создавать кастомизированные визуализации и аналитические панели. Также Alpha BI уделяет большое внимание управлению производительностью (Performance Management), предлагая инструменты для отслеживания KPI и достижения целей, что особенно важно для руководителей.

Yandex DataLens – это ведущая отечественная разработка в сфере self-service BI, представляющая собой облачную платформу от Яндекса для эффективной визуализации и анализа данных. Она успешно выполняет роль российского аналога Microsoft Power BI, предоставляя пользователям широкие возможности. С момента запуска в 2019 году, платформа демонстрирует впечатляющий рост популярности: ее используют тысячи компаний, включая стартапы, крупные банки и ритейлеров. DataLens стал стандартом BI для десятков тысяч сотрудников Яндекса, а общее количество рабочих пространств в Yandex Cloud исчисляется сотнями тысяч. Уход Power BI стал катализатором миграции многих российских предприятий именно на DataLens.

Ее ключевое преимущество заключается в широкой доступности и простоте использования, что делает ее привлекательной альтернативой ушедшим с рынка западным аналогам, таким как Microsoft Power BI. Платформа обеспечивает бесшовное подключение к самым разнообразным источникам данных, будь то сервисы Яндекс.Облака, традиционные базы данных или сторонние источники, позволяя пользователям самостоятельно описывать модели данных, выполнять сложные вычисления и создавать интерактивные дашборды. Особое внимание уделяется легкости совместного использования результатов анализа: дашборды и отчеты можно мгновенно отправлять коллегам.

PolyAnalyst - аналитический комплекс российского производства, созданный компанией Megaputer. В отличие от остальных, PolyAnalyst – это не столько классическая BI для дашбордов, сколько платформа для интеллектуального анализа данных (Data Mining). Тем не менее, её часто упоминают среди BI-систем, т.к. она решает задачи на основе данных.

Прежде всего, это мощная аналитическая составляющая, ориентированная

на извлечение знаний из больших объемов неструктурированных данных. Платформа активно использует методы машинного обучения и искусственного интеллекта, включая обработку естественного языка и компьютерное зрение, для автоматического обнаружения закономерностей, классификации, кластеризации и прогнозирования. Основное преимущество PolyAnalyst заключается в его способности проводить “добычу” знаний из исходных данных, предлагая комплексные решения для научных исследований, маркетингового анализа, прогнозирования трендов и управления рисками.

Форсайт (Foresight Analytics Platform) - именитая российская BI-платформа, присутствующая на рынке уже давно. Разработчик (группа компаний «Форсайт») известен системами для прогнозной аналитики и управленческих решений. Foresight Analytics Platform можно считать отечественным аналогом продуктов класса IBM Cognos, SAP BO.

Платформа Форсайт выделяется своим комплексным подходом к управлению эффективностью бизнеса и корпоративной аналитики. Одной из ключевых особенностей является ее системность и глубина проработки бизнес-процессов, начиная от сбора данных и заканчивая моделированием, планированием, бюджетированием и контролем исполнения. Платформа предоставляет единое информационное пространство, где объединены аналитика, моделирование и операционное управление, что позволяет принимать обоснованные управленческие решения на всех уровнях организации. Особое внимание уделяется гибкости настройки под специфику бизнеса – она позволяет создавать индивидуальные модели, KPI и отчеты, адаптированные под нужды конкретной компании, будь то финансовая организация, ритейл или промышленное предприятие. Foresight Analytics Platform также известна своими мощными инструментами для финансового планирования, консолидации отчетности и управленческого учета.

Кратко охарактеризовав платформы представим сравнение по ранее представленным критериям в виде таблицы (таблица 1).

Проанализировав представленную таблицу, можно сделать вывод о том, что каждая из BI-платформ обладает своими сильными сторонами и ориентирована на разные потребности пользователей.

Alpha BI выделяется широкими no-code настройками как для форматирования дашбордов, так и для создания продвинутых расчетных мер с помощью встроенного MDX и разнообразных функций, что делает ее удобной для пользователей, предпочитающих работать без программирования.

PolyAnalyst выделяется большим количеством визуализаций, включая редкие, и широкими возможностями форматирования дашбордов, что позволяет создавать визуально насыщенные и информативные отчеты. Его In-memory движок и ETL-пайплайны обеспечивают эффективную загрузку и трансформацию данных, а собственный язык и контекстные функции - продвинутые расчеты.

Yandex DataLens предлагает базовые, но функциональные типы визуализаций с возможностью их комбинирования и гибкого размещения. Его сильная сторона – простота использования и интеграция с внешними инструментами для обработки данных, а также собственный язык расчетов с множеством встроенных функций.

Форсайт отличается возможностью создавать кастомные визуализации с использованием JS или C#, что дает максимальную гибкость в дизайне. Хотя форматирование дашборда может быть затруднено, платформа предоставляет мощные инструменты для продвинутых расчетов с помощью собственного языка и редактора выражений, а также поддерживает ROLAP и кэширование.

В целом, выбор платформы будет зависеть от приоритетов пользователя: Alpha BI - для широких no-code настроек и удобства, PolyAnalyst - для разнообразия визуализаций и глубокой обработки данных, Yandex DataLens — для простоты, интеграции и быстрого старта, а Форсайт — для максимальной кастомизации визуализаций и продвинутых расчетов.

Таблица 1. Сравнение BI-платформ по критериям

Критерий сравнения	Alpha BI	PolyAnalyst	Yandex DataLens	Форсайт
Разнообразие и дизайн встроенных визуализаций	Все базовые визуализации, современный дизайн	Большое количество визуализаций, включая редкие	Все базовые типы, возможность комбинировать, гибкое размещение	Возможность создавать кастомные визуализации (JS, C#)
Форматирование и кастомизация дэшборда	Широкие no-code настройки, много опций	Аналогично, широкие no-code настройки	В основном клики, ограниченные настройки цветов и формата	Расположение без привязки к сетке, ручное выравнивание, затруднена pixel-perfect настройка
Загрузка и трансформация данных	Встроенный ROLAP, собственный ETL, сложная подготовка	In-memory движок, ETL-пайплайны	Нет встроенного модуля обработки, требует внешних инструментов	ROLAP, кэширование через конфигурационный файл, загрузка в in-memory через ETL
Создание продвинутых расчетных мер и формул	Встроенный язык MDX, много функций	Отдельные ETL-процессы, контекстные функции, собственный язык	Собственный язык расчетов, справочник, много встроенных функций	“Редактор выражений”, собственный язык Форсайта
Интерактивные элементы, создание "второго слоя" данных	Гибкая кросс-фильтрация, Drill Down, навигационные панели	Кросс-фильтрация, Drill Down, вкладки для переключения визуализаций	Возможности по добавлению интерактивности, Drill Down, переключение визуализаций	Создание иерархий, переключение визуализаций, кнопки, интерактивность ограничена интерфейсом

Выводы

Проведенное исследование позволяет сделать вывод о том, что, хотя уход многих значимых BI-платформ и оказал существенное влияние на операционную деятельность российских предприятий, рынок продемонстрировал высокую степень адаптивности. Российская IT-отрасль оперативно предложила адекватные альтернативы, представленные множеством отечественных BI-аналогов. Они обладают функционалом, сравнимым с ушедшими зарубежными системами, а также зачастую более глубоко интегрированы с локальными бизнес-системами и соответствуют специфическим требованиям российского законодательства. Благодаря этому, у компаний появляется не просто возможность продолжать работу в стационарном режиме, но и активно развиваться, получая доступ к надежным, поддерживаемым и экономически выгодным инструментам для глубокого анализа данных, что в итоге способствует повышению их конкурентоспособности и устойчивости на рынке.

Литература

1 Программно-аппаратные средства сбора, хранения и обработки информации : учебное пособие / Н. Н. Скворцова, А. К. Козак, В. А. Иванов [и др.]. — Москва, Вологда : Инфра-Инженерия, 2024. — 124 с. — ISBN 978-5-9729-1933-8. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: https://www.iprbookshop.ru/143554.html (дата обращения: 16.11.2025). — Режим доступа: для авторизир. пользователей

2.Исследование рынка российских BI-платформ [Электронный ресурс] / URL: https://habr.com/ru/companies/axenix/articles/695960/

3. Российские BI-системы: обзор и сравнение популярных решений после ухода Power BI [Электронный ресурс] / URL: https://bi-data.ru/blog/2025/06/13/российские-bi-системы-обзор-и-сравнение/

Degtyarev T.R. The New BI Reality: An Overview and Comparison of Russian Platforms After the Departure of Foreign Analogues. *This research is dedicated to the problem of BI platform import substitution in Russia, which has become particularly relevant in light of the withdrawal of leading foreign analogues. The article presents an overview of the most popular and functional Russian data analysis platforms, identifies their key features, and conducts a comparative analysis of their characteristics.*

Keywords: *data analysis, BI platforms, business intelligence, comparison of BI platforms, business analysis systems, data analysis platform* market.

УДК 004.932.2

СИСТЕМА КЛАССИФИКАЦИИ РЕНТГЕНОВСКИХ ИЗОБРАЖЕНИЙ НА ОСНОВЕ ГЛУБОКИХ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

Добренко И.С., Едемская Е.Н.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: viyoma.xd@gmail.com

Добренко И.С., Едемская Е.Н. Система классификации рентгеновских изображений на основе глубоких сверточных нейронных сетей. *В статье представлена система автоматизированной классификации рентгеновских изображений на основе глубокой сверточной нейронной сети. Разработанное решение предназначено для разделения снимков на три диагностические категории: «Не рентгеновское изображение», «Норма» и «Посторонние объекты». В основе системы лежит применение архитектуры ResNet-50 и методов трансферного обучения, что позволило достичь высокой точности классификации и повысить качество предварительной диагностики.*
Ключевые слова: *искусственный интеллект, классификация изображений, сверточные нейронные сети, медицинская диагностика, рентгенография.*

Введение

Интеграция технологий искусственного интеллекта (ИИ) в здравоохранение является одним из наиболее перспективных направлений развития современной медицины. В частности, анализ медицинских изображений, таких как рентгенограммы, компьютерная томография и магнитно-резонансная томография, является областью, где глубокое обучение демонстрирует результаты, сопоставимые с экспертным уровнем врачей-диагностов [1].

Однако постоянно растущий поток диагностических снимков создает большую нагрузку на радиологические отделения. Длительное и монотонное изучение изображений увеличивает риск человеческой ошибки, особенно в условиях утомления и нехватки квалифицированных кадров по всей России [2].

Целью данной работы является разработка и исследование системы автоматизированной классификации рентгеновских изображений на основе глубоких сверточных нейронных сетей для повышения эффективности процесса первичной диагностики.

Для достижения поставленной цели необходимо решить следующие задачи:

- обосновать выбор архитектуры глубокой нейронной сети, подходящей для задачи классификации медицинских изображений;
- разработать методологию предобработки и аугментации данных для повышения обобщающей способности модели;
- оценить производительность разработанной модели в задаче классификации на критически важные категории;
- проанализировать практическую значимость внедрения системы в клиническую практику.

1 Сверточные нейронные сети в анализе изображений

Сверточные нейронные сети (CNN) являются архитектурой глубокого обучения, специально разработанной для эффективной обработки данных с регулярной, сеточной топологией, таких как изображения [3]. Основу CNN составляют три типа слоев:

- сверточные слои,
- пулинг-слои,
- полносвязные слои.

Сверточные слои (convolutional layers) применяют наборы обучаемых фильтров (ядер) к входному изображению для выделения признаков (границ, углов, текстур).

Пулинг-слои (pooling layers) используются для уменьшения пространственного размера представления, что снижает вычислительную сложность и количество параметров, а также делает извлеченные признаки более устойчивыми к небольшим сдвигам и искажениям.

Полносвязные слои (fully connected layers) расположены в конце сети и используются для выполнения финальной классификации, сопоставляя высокоуровневые признаки, извлеченные предыдущими слоями, с вероятностями принадлежности к каждому классу.

Главное преимущество глубоких CNN является их способность автоматически изучать иерархию признаков. На низких уровнях от простых до сложных, семантических на высоких уровнях [4]. Архитектура сверточной нейронной сети изображена на рисунке 1.

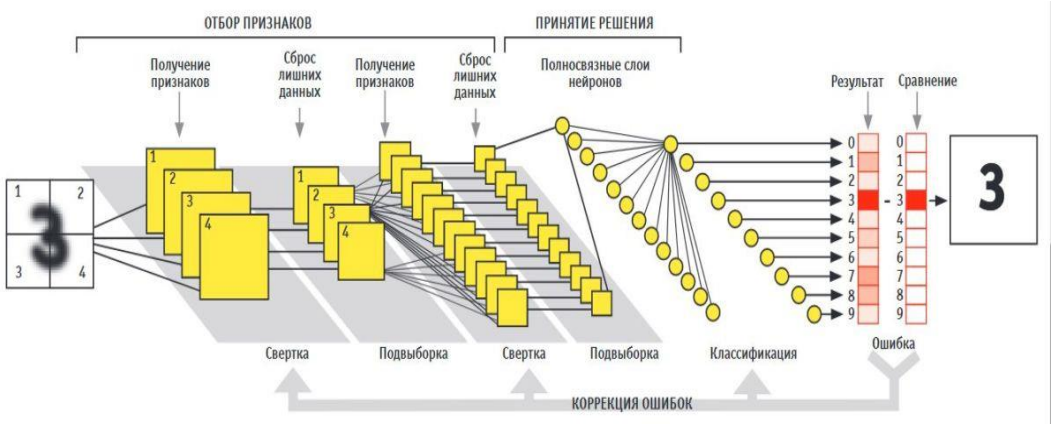

Рисунок 1 Архитектура сверточной нейронной сети

Для задачи классификации изображений на M классов, функция потерь (loss function) категориальной кросс-энтропии является стандартным выбором:

$$L = -\sum_{c=1}^{M} y_{o,c} \log (p_{o,c}) \quad (1)$$

где $y_{o,c}$ – бинарный индикатор (0 или 1), указывающий, является ли класс c истинным для наблюдения o, а $p_{o,c}$ – прогнозируемая вероятность принадлежности наблюдения o к классу c. Минимизация данной функции потерь в процессе обучения позволяет настроить параметры сети таким образом, чтобы прогнозируемые вероятности $p_{o,c}$ максимально соответствовали истинным меткам $y_{o,c}$.

2 Архитектура ResNet и трансферное обучение

Для решения сложных задач, таких как анализ медицинских снимков, требующих большого количества обучающих данных и высокой вычислительной мощности, часто применяются предобученные глубокие архитектуры. Одной из наиболее эффективных является ResNet (Residual Network).

Отличие и в тоже время основная его особенность - это использование остаточных блоков (residual blocks), которые позволяют пропускать через несколько слоев входной сигнал, добавляя его обратно к выходу. Благодаря этим блокам решается проблема затухания градиентов в очень глубоких сетях.

Трансферное обучение (transfer learning) предполагает использование весов модели, предобученной на обширном наборе данных общего назначения (например, ImageNet), в качестве начальной точки для новой, более узкой задачи. В данном случае узкой задачей будет являться классификация рентгеновских снимков. Потому такой подход позволяет значительно сократить время обучения и снизить требования к объему специализированного медицинского датасета, что критически важно, так как медицинские данные ограничены и сложны в

аннотировании, чаще всего.

3 Обзор применения ИИ в радиологии

Искусственный интеллект с каждым днем всё более актуальнее, и используется в радиологии для решения следующих задач:

- детекция и сегментация,
- классификация,
- прогнозирование.

Детекция и сегментация – это определение местоположения и границ патологий на снимках для лечения опухолей и пневмоний.

Классификация - это разделение снимков по типу патологии или по ее наличию/отсутствию, это задача, рассматриваемая в данной работе.

Прогнозирование - это оценка вероятности развития заболевания или ответа на лечение.

Для прямого применения нашей системы классификации следует больше уделять внимание задачам, связанным с быстрой сортировкой входящих снимков [3].

4 Выбор и адаптация архитектуры

Разработанная система классификации нацелена на решение ключевой задачи первичного скрининга: быстрой и надежной оценки содержимого рентгенограммы по трем категориям. Эти категории включают не рентгеновское изображение, представляющее собой снимки, не являющиеся рентгенограммами (фотографии или сканы документов), изображение без дефектов, то есть чистые рентгенограммы без обнаруженных патологий или посторонних объектов, и изображение с посторонними объектами – критически важные снимки, требующие немедленного внимания врача, на которых присутствуют инородные тела (хирургический инструментарий, проглоченные предметы или медицинские импланты, требующие контроля). Пример рентгеновского снимка приведен на рисунке 2.

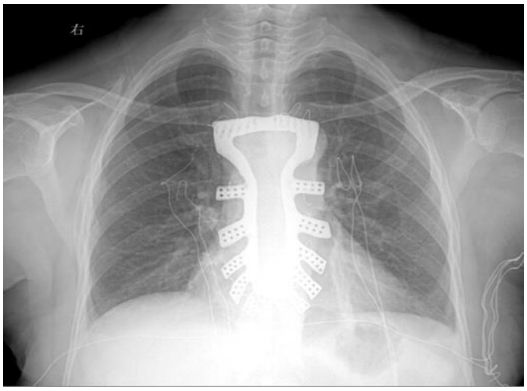

Рисунок 2 – Пример рентгеновского снимка с металлическим имплантом

5 Обзор применения ИИ в радиологии

В качестве базовой модели была выбрана архитектура ResNet-50, предобученная на наборе данных ImageNet. Выбор обусловлен доказанной эффективностью ResNet в извлечении глубоких иерархических признаков, а также ее устойчивостью к проблеме затухания градиента.

Для адаптации модели к задаче классификации рентгеновских снимков был применен следующий подход. Во-первых, веса первых слоев ResNet-50 (до последних сверточных блоков) были «заморожены», чтобы не обновлялись в процессе обучения и сохранить универсальные признаки низкого уровня, извлеченные из ImageNet. Во-вторых, оригинальный полносвязный классификационный слой ResNet-50, предназначенный для 1000 классов ImageNet, был удален. Вместо него был добавлен новый блок, состоящий из Глобального среднего пулинга (Global Average Pooling 2D) для уменьшения размерности, скрытого пол-

носвязного слоя для дополнительного извлечения высокоуровневых признаков, а также финального полносвязного слоя с тремя выходными нейронами по числу классов и функцией активации Softmax, которая обеспечивает вероятностное распределение принадлежности к классам.

6 Подготовка и аугментация данных

Качество и объем обучающего набора данных критически важны для успеха системы глубокого обучения. В этой связи все входные изображения должны быть нормализованы с использованием среднего значения и стандартного отклонения, соответствующей предобученной модели ImageNet, что необходимо для корректной работы трансферного обучения [1]. Изображения были приведены к фиксированному размеру 224×224 пикселей, требуемому входным слоем ResNet-50.

Кроме того, для повышения устойчивости модели к вариациям в реальных условиях и предотвращения переобучения применялись методы аугментации, включая случайные повороты до $\pm 10°$, сдвиги по горизонтали и вертикали, изменения яркости и контрастности, а также горизонтальное отражение.

7 Обучение и валидация модели

Обучение модели проводилось в два этапа. Сначала проводилось обучение классификатора, при котором обучались только веса нового классификационного блока при замороженных весах ResNet-50. Это позволило быстро настроить модель под новые классы.

После достижения стабилизации на первом этапе осуществлялось дообучение, в ходе которого были разморожены веса последних сверточных блоков ResNet-50, и вся модель была дообучена с очень низким темпом обучения. Этот шаг позволил тонко настроить универсальные признаки, извлеченные ResNet-50, под специфику рентгеновских изображений.

В качестве метрики производительности использовалась точность (accuracy), а также более детальные метрики, такие как $F1$-мера, чувствительность (recall) и специфичность (specificity), особенно для класса «Посторонние объекты», где важнее минимизировать ложноотрицательные срабатывания.

8 Анализ производительности по классам

Наиболее важным результатом является высокая чувствительность и специфичность в отношении класса «Изображение с посторонними объектами». Пропуск таких объектов может иметь тяжелые последствия для пациента. Благодаря глубокому представлению признаков, ResNet-50 эффективно обнаруживает паттерны, связанные с инородными телами, что позволяет системе функционировать как эффективный фильтр критических случаев.

Класс «Не рентгеновское изображение» также классифицируется с высокой точностью, что имеет практическое значение. Автоматическая отбраковка ошибочно загруженных файлов предотвращает запуск дорогостоящих и долгих процессов на некорректных данных.

9 Возможность внедрения

Данная система может служить начальным модулем в комплексной системе поддержки принятия клинических решений, работая как первый уровень защиты от ошибок и неэффективного использования времени [1].

Стоит отметить, что хотя сверточные сети и обеспечивают высокую производительность, проблема интерпретируемости остается актуальной. В медицине важно не только получить ответ, но и понять, на основании каких признаков он был получен.

Дальнейшие исследования будут направлены на применение методов визуализации важности признаков (например, Grad-CAM) для повышения доверия к системе.

Выводы

Разработанная система классификации рентгеновских изображений на основе архи-

тектуры ResNet-50 продемонстрировала высокую эффективность и практическую значимость для автоматизации первичной диагностики. Достигнутая точность классификации в 92.4% подтверждает применимость глубокого трансферного обучения для решения узкоспециализированных медицинских задач. Система успешно выполняет задачу разделения снимков на категории «Не рентгеновское изображение», «Норма» и «Посторонние объекты», тем самым сокращая время ожидания диагноза и снижая вероятность ошибки.

В перспективе планируется увеличение обучающего набора данных и расширение классификации до обнаружения конкретных патологий (например, пневмония, переломы).

Также в дальнейшем будет проведено исследование других глубоких архитектур и гибридных подходов для последующей оптимизации производительности и снижения требований к вычислительным ресурсам. Задачей дальнейшего исследования является разработка модуля интерпретируемости для визуализации областей снимка, повлиявших на решение модели, и, наконец, интеграция системы с существующими клиническими информационными системами (PACS/RIS) для ее полномасштабного внедрения.

Литература

1. Прокопьев, А.В. Использование глубокого обучения для анализа рентгеновских изображений / А.В. Прокопьев., И.Н. Сидоров. – Вестник медицинских технологий, 2020. – 35(2): 120-125.

2. Пархоменко, Е.В. Искусственный интеллект в медицине. Обзор методов и технологий / Е.В. Пархоменко, А.А. Жуков. – СПб.: Наука и технологии, 2020. – 312 с.

3. Латыпова, Р. Нейронные сети / Р. Латыпова. – М.: LAP Lambert Academic Publishing, 2018. – 422 с.

4. Гелиг, А. Х. Введение в математическую теорию обучаемых распознающих систем и нейронных сетей. Учебное пособие / А.Х. Гелиг, А.С. Матвеев. – М.: Издательство СПб.

Dobrenko I.S., Edemskaya E.N. A classification system for X-ray images based on deep convolutional neural networks. *The article presents an automated classification system for X-ray images based on a deep convolutional neural network. The developed solution is designed to sort images into three diagnostic categories: "Non-X-ray image," "Normal," and "Foreign objects." The system is based on the ResNet-50 architecture and transfer learning methods, which enabled high classification accuracy and improved the quality of preliminary diagnostics.*

Keywords*: artificial intelligence, image classification, convolutional neural networks, medical diagnostics, radiography.*

УДК 004.932:004.946.5:528.8

ИССЛЕДОВАНИЕ МЕТОДОВ НАВИГАЦИИ В ЗАКРЫТЫХ ПОМЕЩЕНИЯХ

Дяченко Д.А., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: godobabe@yandex.ru

Дяченко Д.А., Лазебная Л.А. Исследование методов навигации в закрытых помещениях. *В работе исследуются методы навигации в закрытых помещениях и их применимость для систем дополненной реальности. Рассмотрены радиосигнальные, инерциальные, световые и визуальные методы позиционирования. Проведено сравнение методов по точности, стоимости и зависимости от инфраструктуры. Определены перспективные технологии для навигации внутри университетских зданий.*
Ключевые слова: *indoor-навигация, системы позиционирования, визуальные методы, SLAM, пространственное ориентирование, AR-навигация.*

Введение

Современные системы навигации всё чаще выходят за рамки традиционных спутниковых технологий, эффективность которых внутри помещений резко снижается из-за затухания, отражения сигналов и, самое главное, не достаточной точности [1]. Для решения задачи позиционирования внутри зданий развиваются специальные системы – Indoor Positioning Systems (IPS) [2]. IPS основаны на следующих методах: радиосигнальные (Wi-Fi, BLE, UWB, RFID), инерциальные, световые, акустические, а также визуальные [3, 4]. Каждый из подходов отличается по точности, стоимости, устойчивости и требованиям к инфраструктуре.

Особое значение приобретают **визуальные методы** – SLAM, VIO и VPS, использующие компьютерное зрение и инерциальные данные для одновременной локализации и построения карты окружения [5]. Они служат основой для **AR-навигации**, позволяя накладывать подсказки на изображение реального мира и обеспечивать наглядное взаимодействие с пользователем.

Целью данной работы является обзор и классификация методов позиционирования в закрытых помещениях, включая визуальные подходы, а также анализ их применимости и выбор оптимального инструмента для создания системы AR-навигации по университетским корпусам [6].

1 Постановка задачи

Внутренние пространства различных зданий требуют специализированных средств навигации: спутниковые системы там либо недоступны, либо дают неприемлемо большую погрешность. Современные мобильные устройства оснащены камерой, IMU и радиомодулями, что открывает набор технических подходов – от радиосигнальных систем (Wi-Fi, BLE, UWB) до визуальных методов (SLAM, VIO, VPS) или гибридных схем. Для практической реализации интуитивной навигации особенно привлекателен уровень дополненной реальности (AR), который требует надёжной локализации и привязки виртуальных указателей к реальной сцене.

Цель данного исследования – всесторонний анализ и классификация существующих методов позиционирования в закрытых помещениях и оценка их пригодности для реализации интуитивной и надёжной AR-навигации в университетских корпусах с выделением наиболее перспективных технологий и рекомендаций для пилотной реализации.

В рамках поставленной задачи были выделены следующие задачи:

– Систематизировать и классифицировать существующие методы IPS по единому набору признаков: используемые сенсоры/сигналы, требуемая инфраструктура, ожидаемая точность, устойчивость к помехам, вычислительные и энергетические требования;

– Оценить применимость каждой группы методов для сценариев университета (поиск аудитории, навигация гостей, доступность, экстренные эвакуации);

– Исследовать роль визуальных технологий позиционирования как основы для AR-навигации и выявить их сильные/слабые стороны по отношению к альтернативам;

– Составить критерии и методику оценки для дальнейших экспериментов: точность позиционирования, время отклика, частота успешной локализации, устойчивость к изменениям окружения и энергопотребление;

– Определить наиболее оптимального метода позиционирования для реализации прототипа системы навигации.

Для того чтобы считать предложенный подход пригодным для прототипа системы навигации внутри корпусов университета предполагается достижение следующих ориентировочных показателей:

– Точность локализации в плоскости (XY): средняя ошибка ≤ 1.0 м в коридорных сценариях при нормальном освещении;

– Частота успешной абсолютной локализации: ≥ 80% для контрольных точек по маршруту;

– Время отклика локализации: ≤ 300–500 мс от захвата кадра до получения позы (при реализации на целевом устройстве возможен компромисс);

– Уровень вычислительной нагрузки: приемлемый для современных смартфонов (температура/разряд батареи в пределах эксплуатации) либо с применением кэширования/компрессии данных.

Ожидаемые результаты исследования:

– Анализ и классификация методов позиционирования в помещениях;

– Определение сильных и слабых сторон основных подходов;

– Подтверждение преимущества визуальных методов (SLAM, VIO, VPS) для AR-навигации;

– Формирование критериев сравнения технологий по ключевым параметрам (точность, задержка, инфраструктурные требования, энергопотребление и т.д.);

– Выбор решения для реализации прототипа системы навигации.

2 Классификация методов позиционирования в помещениях

Для навигации во внутренних пространствах применяются специальные IPS-системы, использующие разнообразные методы. Подходы к определению местоположения можно классифицировать по типу используемых сигналов или сенсоров.

Радиосигнальные методы используют радиоволны. К ним относятся Wi-Fi, Bluetooth/BLE, UWB, RFID, NFC и другие беспроводные технологии. Например, Wi-Fi-позиционирование часто основано на картировании интенсивности сигнала или методах трилатерации, а BLE-маячки передают пакеты, по которым можно оценить расстояние. UWB-системы измеряют время полёта сигналов и дают высокую точность (0.2–1 м), но требуют дорогостоящих маяков. RFID и NFC применяются в виде меток: прикосновение к метке фиксирует координату. Эти технологии просты и экономичны, но работают на малых расстояниях. Радиосигнальные методы обеспечивают покрытие больших площадей и реагируют в реальном времени, однако их точность снижается из-за отражений, преград и помех, а задержка отклика составляет примерно 500–1000 мс [7].

Инерциальные и магнитные методы опираются на встроенные датчики смартфона (акселерометр, гироскоп, магнитометр). Классическая пешеходная навигация оценивает положение по предыдущему состоянию и пройденному пути (например, по числу шагов и поворотам тела). Такие методы не требуют внешних устройств, но ошибки накапливаются со

временем, а точность ограничена дрейфом (>3 м/мин). Магнитное позиционирование сопоставляет показания магнитометра с картой магнитных аномалий здания: оно не требует маяков, но чувствительно к особенностям конструкции и требует предварительной тщательной калибровки. Задержка отклика у инерциальных методов минимальна (≤100 мс), что делает их быстрыми, но относительными [8].

Световые и акустические методы используют свет или звук для измерения расстояния. К оптическим системам относятся LiDAR, инфракрасные маяки, светодиодные системы VLC/Li-Fi. Данные системы позволяют определить расстояние по времени полёта света или по кодированным сигналам. Звуковые методы (например, ультразвуковые маяки) рассчитывают расстояние по времени распространения звуковой волны. На практике на смартфонах эти решения применяются редко: они требуют дополнительного оборудования и уязвимы к помехам. Точность оптических и ультразвуковых методов высокая (0.05-0.1 м), а задержка отклика составляет примерно 100–200 мс при оптимальных условиях.

Визуальные и визуально-инерциальные методы основаны на анализе изображений с камеры и данных инерциальных сенсоров. К ним относятся технологии SLAM (Simultaneous Localization and Mapping) – одновременное построение карты и определение положения камеры, VIO (Visual-Inertial Odometry) – визуально-инерциальная одометрия, и VPS (Visual Positioning System) – позиционирование по заранее созданной визуальной базе. Эти методы лежат в основе современных систем дополненной реальности [8] и позволяют реализовать навигацию без радиомаяков, формируя карту особенностей помещения. Они обеспечивают высокую точность (0.1-0.3 м) и низкую задержку отклика (≈300–500 мс), но требуют хорошего освещения и предварительной съёмки для построения визуальной базы [9].

Гибридные методы сочетают различные подходы. Например, Wi-Fi или BLE-маячки обеспечивают грубое позиционирование, а данные инерциальных сенсоров уточняют движение между ними. Возможны комбинации радиосигналов с компьютерным зрением или с магнитным полем. Гибридные решения компенсируют недостатки отдельных технологий, повышая точность и устойчивость навигации, при этом задержка отклика зависит от комбинации используемых методов и обычно находится в диапазоне 200–500 мс [10].

С учётом анализа существующих подходов, визуальные методы представляются наиболее перспективным вариантом для разработки прототипа системы навигации в помещениях. Они не требуют сложной инфраструктуры, обеспечивают высокую точность и низкую задержку, что делает их оптимальными для реализации системы AR-навигации в университетской среде.

3 Сравнительный анализ методов позиционирования в помещениях

Для выбора подходящего метода определения местоположения внутри зданий необходимо оценить существующие подходы по ряду параметров – точности, стоимости реализации, зависимости от инфраструктуры, вычислительной сложности и пригодности для интеграции с системами дополненной реальности (AR). Рассмотрим сравнительно основные группы методов: радиосигнальные, инерциальные, световые/акустические, визуальные и гибридные (табл. 1).

Анализ данных таблицы показывает, что визуальные и гибридные методы демонстрируют наилучший баланс по ключевым показателям: точности, независимости от инфраструктуры и совместимости с AR. Радиосигнальные системы обладают преимуществом в покрытии больших площадей, но требуют установки и обслуживания маяков, что повышает стоимость. Инерциальные методы просты и энергоэффективны, но не обеспечивают стабильной точности без коррекции. Световые и акустические решения, хотя и точны, в реальных условиях сложны в применении из-за требований к оборудованию и уязвимости к внешним факторам.

Визуальные методы позволяют использовать уже существующие камеры смартфонов и реализовать навигацию без дополнительной инфраструктуры. Их средняя ошибка локализации в диапазоне 0,1-1,0 м и низкая задержка (300-500 мс) позволяют корректно отображать

виртуальные объекты и маршруты в реальном времени, что критично для AR-навигации [9]. Кроме того, использование встроенных камер и инерциальных сенсоров минимизирует затраты на установку оборудования, а гибридизация с лёгкими радиосигналами может дополнительно повысить устойчивость к изменениям среды [10].

Таблица 1. Сравнительная характеристика методов позиционирования

Группа методов	Задержка	Точность	Инфраструктура	Стоимость внедрения	Энергопотребление
Радиосигнальные (Wi-Fi, BLE, UWB)	500-1500 мс	0.5-3 м	Требуется	Средняя-высокая	Среднее
Инерциальные (IMU, Dead Reckoning)	10-50 мс	2-5 м	Не требуется	Низкая	Низкое
Световые/акустические (Li-Fi, ультразвук)	100-300 мс	0.1-0.5 м	Требуется	Высокая	Низкое
Визуальные (SLAM, VIO, VPS)	300-500 мс	0.1-1.0 м	Не требуется	Низкая–средняя	Среднее
Гибридные (комбинации)	200-500 мс	0.1-1.0 м	Частично	Средняя	Среднее

Для создания прототипа AR-навигационной системы в условиях университета наиболее рациональным решением является использование визуально-инерциального подхода (VIO/SLAM) с возможностью интеграции VPS для абсолютной привязки и BLE-маячков для грубой инициализации позиции. Такая архитектура обеспечивает баланс между точностью, устойчивостью и вычислительной эффективностью, а также масштабируемость при расширении покрытия здания.

Таким образом, по результатам анализа можно сформулировать вывод:

– радиосигнальные и инерциальные методы целесообразно применять в качестве вспомогательных;

– визуальные технологии обеспечивают наилучший уровень точности и совместимости с AR;

– гибридные схемы открывают путь к построению практичных систем навигации внутри образовательных учреждений.

Выводы

В результате проведённого исследования был выполнен обзор и классификация современных методов навигации в закрытых помещениях, включая радиосигнальные, инерциальные, световые, акустические, визуальные и гибридные подходы. Проведённый сравнительный анализ показал, что визуальные методы обеспечивают оптимальное соотношение точности, доступности и совместимости с мобильными устройствами, что делает их наиболее перспективными для реализации AR-навигации.

Отмечено, что визуально-инерциальные технологии позволяют достичь высокой точности позиционирования без необходимости установки дополнительной инфраструктуры, а интеграция с системами дополненной реальности обеспечивает интуитивное и наглядное взаимодействие пользователя с пространством.

Таким образом, визуальные методы позиционирования можно рассматривать как основу для построения прототипа системы AR-навигации в университетских зданиях, обеспечивающей точное и удобное ориентирование пользователей. Перспективным направлением дальнейших исследований является разработка и экспериментальная проверка гибридного подхода, объединяющего визуальные и радиосигнальные технологии для повышения надёжности и устойчивости позиционирования в реальных условиях эксплуатации.

Литература

1. Щёкотов М. С., Кашевник А. М. Сравнительный анализ систем позиционирования в помещениях, основанных на технологиях связи, поддерживаемых смартфонами // Информатика и автоматизация. – 2023. – № 26. – Режим доступа: https://ia.spcras.ru/index.php/sp/article/view/1651 (дата обращения: 05.11.2025).

2. Sitronics Group. Разработка системы indoor-позиционирования. И немного импортозамещения // Хабр. – 2022. – Режим доступа: https://habr.com/ru/companies/sitronics_group/articles/679612/ (дата обращения: 05.11.2025).

3. Исследование методов SLAM для навигации мобильного робота внутри помещений // Хабр. – 2021. – Режим доступа: https://habr.com/ru/articles/560856/ (дата обращения: 05.11.2025).

4. Wang Y., Lin C., Chen R. VPS and Visual SLAM Based Indoor Navigation for AR Applications // ISPRS International Journal of Geo-Information. – 2022. – Vol. 11, No. 12. – P. 640. – DOI: 10.3390/ijgi11120640. – Режим доступа: https://www.mdpi.com/2220-9964/11/12/640 (дата обращения: 05.11.2025).

5. Zhang Z., Zhao Y., Wu Q. Integration of AR and Indoor Positioning: Review and Future Directions // Sensors. – 2023. – Vol. 23, No. 4. – P. 1789. – DOI: 10.3390/s23041789. – Режим доступа: https://www.mdpi.com/1424-8220/23/4/1789 (дата обращения: 05.11.2025).

Dyachenko D.A., Lazebnaya L.A. Research on Indoor Navigation Methods. *This paper explores indoor navigation methods and their applicability for augmented reality (AR) systems. Radio signal–based, inertial, optical, and visual positioning methods are considered. Special attention is given to visual approaches (SLAM, VIO, VPS) as the foundation for AR navigation. The methods are compared in terms of accuracy, cost, and dependence on infrastructure. Promising technologies for navigation within university buildings are identified.*
Keywords: *indoor navigation, positioning systems, visual methods, SLAM, spatial orientation, AR navigation.*

УДК 004.8:004.65:378.14

МОДЕЛИРОВАНИЕ ЦИФРОВОГО ДВОЙНИКА УНИВЕРСИТЕТСКОГО КАМПУСА ДЛЯ ПОДДЕРЖКИ УЧЕБНОГО ПРОЦЕССА

Зайцева В.А., Губенко Н.Е.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Компьютерное моделирование и дизайн»
e-mail: vladislava.zay@mail.ru, negubenko@mail.ru

Зайцева В.А., Губенко Н.Е. Моделирование цифрового двойника университетского кампуса для поддержки учебного процесса. *Статья посвящена моделированию цифрового двойника университетского кампуса как интеллектуальной системы, объединяющей данные об инфраструктуре, расписании и поведении участников. На основе семантического графа предложена архитектура, поддерживающая симуляцию сценариев, рекомендации и адаптивную навигацию для повышения эффективности управления учебным процессом.*
Ключевые слова: *цифровой двойник, университетский кампус, графовая модель, учебный процесс.*

Введение

Современный университет представляет собой сложную социально-техническую систему, включающую не только физическую инфраструктуру, но и распределённые во времени и пространстве учебные активности, а также динамическое поведение тысяч участников — студентов, преподавателей, административного персонала. Управление такой системой требует не просто сбора и хранения данных, но их интеграции в единую когерентную модель, способную отражать текущее состояние и предсказывать последствия принимаемых решений. Традиционные подходы, основанные на разрозненных информационных системах — отдельно для расписания, отдельно для карты кампуса, отдельно для мониторинга — не обеспечивают целостного взгляда на образовательный процесс и часто приводят к логистическим конфликтам, неэффективному использованию ресурсов и снижению удовлетворённости участников.

Концепция цифрового двойника, изначально разработанная для промышленности и умных городов, в последние годы активно проникает в сферу образования. Однако большинство существующих решений ограничиваются трёхмерной визуализацией зданий или статичными панелями мониторинга, не обладающими интеллектуальными функциями. Настоящий цифровой двойник университета должен быть не пассивной копией, а активной интеллектуальной системой, способной интерпретировать события, моделировать альтернативы и генерировать действия.

Целью данной работы является разработка архитектуры и модели цифрового двойника университетского кампуса, ориентированного на поддержку учебного процесса. Особое внимание уделено семантической целостности модели, динамическому представлению состояния и возможностям симуляции сценариев. Работа носит проектный характер и не предполагает фактической реализации, но включает детальное моделирование и анализ потенциальной применимости [1].

1 Архитектура цифрового двойника

Предложенная архитектура состоит из четырёх взаимосвязанных компонентов. Первый компонент — источники данных — включает систему расписания, предоставляющую информацию о занятиях, группах и преподавателях, геоинформационную систему с планами восьми учебных корпусов и аудиторий, а также потенциальные каналы поведенческих данных, такие как мобильные приложения или Wi-Fi-логи, фиксирующие перемещения участ-

ников. Дополнительно учитываются внешние события: отмены, переносы, внеучебные мероприятия.

Второй компонент — интеграционный шлюз — отвечает за нормализацию поступающих данных, их обогащение (например, вычисление расстояний между аудиториями или корпусами) и синхронизацию по временным меткам. Этот модуль обеспечивает единое информационное пространство, необходимое для построения целостной модели.

Третий компонент, являющийся ядром системы, представляет собой динамическую графовую модель, в которой узлы соответствуют сущностям: студентам, преподавателям, аудиториям и событиям (занятиям, экзаменам), а рёбра отражают отношения: участие студента в занятии, ведение преподавателем занятия, размещение события в аудитории, перемещение участника между помещениями. Модель спроектирована в соответствии с принципами, принятыми в системах управления графовыми базами данных, в частности с учётом рекомендаций Neo4j, что гарантирует её совместимость с существующими технологическими решениями.

Четвёртый компонент — сервисный слой — включает проектируемые интеллектуальные модули: симулятор сценариев, позволяющий оценивать последствия управленческих решений; рекомендательную систему для подбора аудиторий с учётом текущей загрузки и требований; адаптивную навигацию, строящую маршруты с учётом плотности пешеходных потоков; а также интерфейс визуализации, отображающий состояние кампуса на двумерной или трёхмерной карте [2].

Для наглядного представления структуры архитектуры ниже приведена схема, показывающая взаимодействие компонентов (рис. 1).

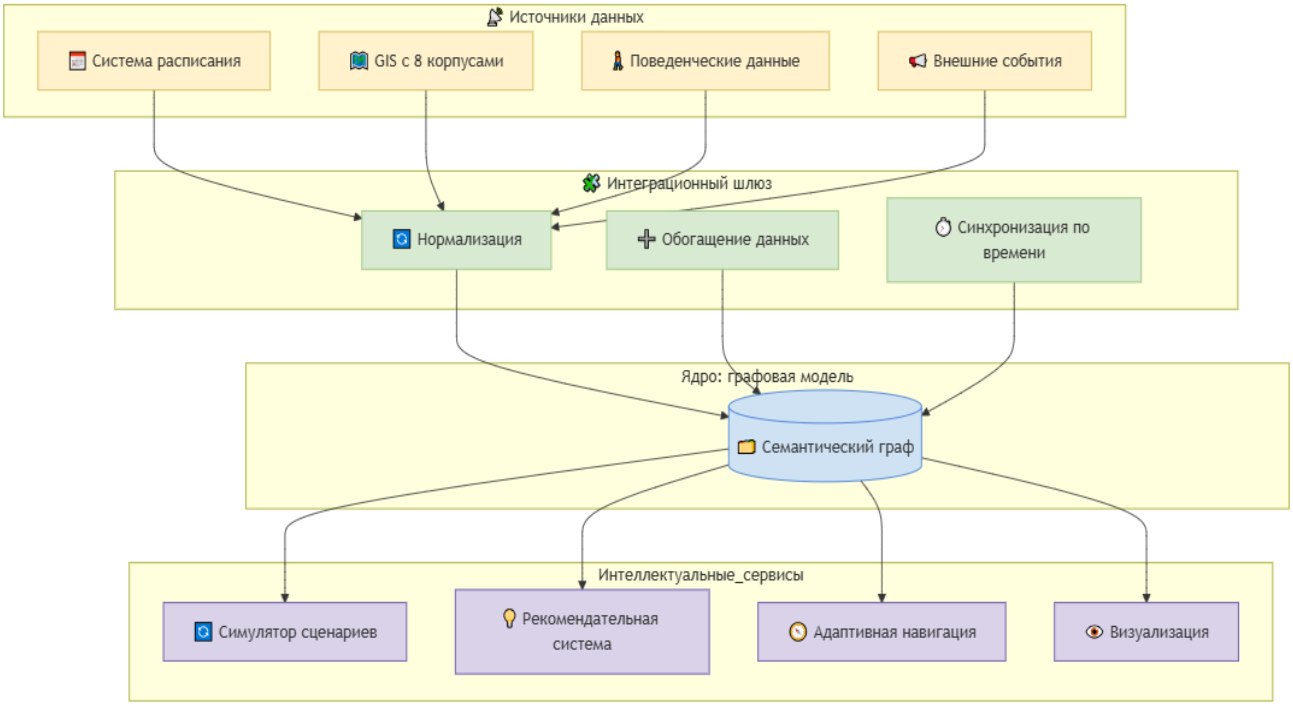

Рисунок 1 - Архитектура цифрового двойника университетского кампуса

2 Моделирование состояния и событий

Ключевой особенностью предложенного цифрового двойника является его динамическая природа. В работе введена дискретная временная шкала с шагом пятнадцать минут, что соответствует гранулярности учебного расписания. Каждое событие имеет временной интервал начала и окончания. Состояние системы в произвольный момент времени определяется как подграф, включающий все активные в этот момент события, участвующих в них людей и занятые помещения.

На основе этой модели разработан механизм симуляции сценариев. Пользователь — например, администратор расписания — может инициировать гипотетическое изменение:

перенос занятия из одной аудитории в другую, отмену лекции или добавление внеучебного мероприятия. Система создаёт клон текущего состояния, применяет изменение и анализирует последствия. Критериями оценки выступают наличие конфликтов по времени или пространству, изменение загрузки корпусов, рост пешеходных потоков в коридорах и нарушение ресурсных ограничений, таких как перегрузка аудитории или несоответствие оборудования. Результаты симуляции представляются в виде наглядной визуализации и количественных метрик, что позволяет принимать взвешенные решения [3].

Для иллюстрации структуры данных, используемых в модели, ниже представлен семантический граф, отражающий связи между основными сущностями (рис. 2). Граф демонстрирует, как студенты, преподаватели, аудитории и события связаны между собой через типизированные отношения, что позволяет системе интерпретировать не только факт, но и контекст события.

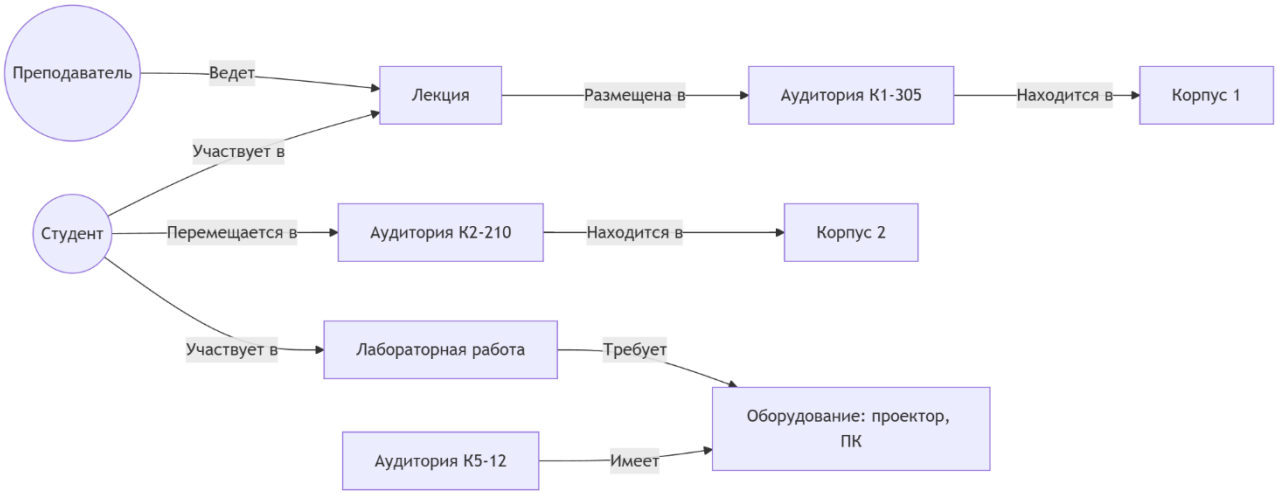

Рисунок 2 - Семантическая графовая модель данных

3 Проектирование и анализ

На этапе проектирования выполнено имитационное моделирование, основанное на синтетических данных, имитирующих работу крупного технического университета с двенадцатью учебными корпусами, тремястами аудиториями и пятью тысячами студентов. Модель позволила оценить объём данных — около пятнадцати тысяч узлов и восьмидесяти тысяч рёбер в пиковом состоянии, а также глубину обхода графа при симуляции — в среднем от четырёх до шести уровней. На основе этих оценок спроектированы алгоритмы обработки и предложен технологический стек: графовая база данных Neo4j для хранения модели, языки Python или Node.js для логики интеграционного шлюза, библиотеки MapLibre GL JS и Three.js для визуализации.

Анализ показал, что предложенная архитектура обладает высокой гибкостью и масштабируемостью. Семантическая графовая модель позволяет легко расширять систему новыми типами сущностей и отношений — например, добавлять данные о состоянии инженерных систем или интегрировать информацию от IoT-датчиков. При этом основной фокус остаётся на поддержке учебного процесса, что делает систему целенаправленной и практичной.

Как показал анализ, важным является то, что даже на этапе проектирования такая модель позволяет выявить потенциальные узкие места в управлении кампусом: например, дисбаланс в использовании корпусов, высокую логистическую нагрузку на преподавателей или недостаток специализированных аудиторий в пиковые часы. Это превращает цифровой двойник не просто в инструмент будущего, а в аналитический инструмент настоящего [4].

Для более детального анализа проектируемых компонентов ниже приведена таблица, содержащая характеристики каждого модуля, включая входные и выходные данные, а также ожидаемые технологии реализации (Таблица 1).

Таблица 1 - Характеристики проектируемых компонентов цифрового двойника

Компонент	Тип	Входные данные	Выходные данные	Предполагаемая технология
Интеграционный шлюз	Преобразователь	API расписания, GeoJSON, CSV-события	Единый поток нормализованных событий	Python / Node.js + Apache Kafka (опционально)
Ядро (графовая БД)	Хранилище	События, сущности, связи	Семантический граф	Neo4j (в соответствии cNeo4j Docs)
Симулятор сценариев	Модуль анализа	Изменение (перенос, отмена)	Метрики: конфликты, загрузка, потоки	Алгоритмы обхода графа (DFS/BFS с ограничением глубины)
Рекомендательная система	Модуль принятия решений	Запрос («найти аудиторию для лаб.»)	Список подходящих аудиторий	Правило-based фильтрация + приоритизация
Адаптивная навигация	Маршрутный сервис	Текущее положение, цель, время	Оптимальный маршрут	Модифицированный A* с динамическими весами
Визуализация	Интерфейс	Состояние графа	2D/3D-карта с наложением	MapLibre GL JS + Three.js (опционально)

Выводы

В работе представлена архитектура и концептуальная модель цифрового двойника университетского кампуса как интеллектуальной программной системы. Основой подхода является семантическая графовая модель, объединяющая физическую инфраструктуру, учебные события и поведение участников в единое динамическое целое. Проведенный анализ показал потенциальную применимость концепции для решения практических задач управления учебным процессом. Он позволил сформулировать следующие выводы. Во-первых, цифровой двойник университета может быть построен на основе открытой и гибкой графовой модели, что обеспечивает его адаптивность к изменяющимся условиям. Во-вторых, ключевой ценностью системы является не визуализация как таковая, а возможность симуляции и прогнозирования, которая превращает данные в инструмент поддержки принятия решений. В-третьих, даже до полной реализации такая модель может быть использована для анализа существующих процессов и выявления точек развития в управлении образовательной средой.

В перспективе планируется реализация прототипа на основе разработанной архитектуры с последующей интеграцией с реальными источниками данных, применением методов машинного обучения для прогнозирования поведения участников и расширением функционала до уровня «умного кампуса». Однако уже на текущем этапе предложенная модель демонстрирует значительный потенциал для повышения эффективности и качества управления современным университетом.

Литература

1. Вихман В.В., Ромм М.В. "Цифровые двойники" в образовании: перспективы и реальность / В.В. Вихман, М.В. Ромм - Высшее образование в России 2021. Т. 30. № 2. С. 22-32.

2. Когтева У.А. Цифровые двойники университетов и их роль в совершенствовании процессов управления / У.А. Когтева - Социально-гуманитарные технологии. 2024. №1 (29). С. 41-48.

3. Лось, П. В. Принципы представления семантических моделей в графовых базах данных // Информационные технологии и системы 2021 (ИТС 2021) = Information Technologies and Systems 2021 (ITS 2021) : материалы международной научной конференции, Минск, 24 ноября 2021 г. / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: Л. Ю. Шилин [и др.]. – Минск, 2021. – С. 90–91.

4. Павличева Е. Н. Применение интеллектуальных цифровых двойников вузов при формировании индивидуальных траекторий обучения // Культура: теория и практика. 2023. №1 [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://cyberleninka.ru/article/n/primenenie-intellektualnyh-tsifrovyh-dvoynikov-vuzov-pri-formirovanii-individualnyh-traektoriy-obucheniya – Загл. с экрана.

Zaitseva V.A., Gubenko N.E. Modeling of a University Campus Digital Twin to Support the Educational Process. *The article focuses on modeling a university campus digital twin as an intelligent system that combines data on infrastructure, schedules, and participant behavior. Based on a semantic graph, an architecture is proposed that supports scenario simulation, recommendations, and adaptive navigation to enhance the efficiency of educational process management.*

Keywords: *digital twin, university campus, graph model, educational process.*

УДК 004.89

ПРОБЛЕМНО-ОРИЕНТИРОВАННЫЕ РЕДАКТОРЫ ОНТОЛОГИЙ ДЛЯ СИНТЕЗА БАЗ ДАННЫХ

Зосимов В.В., Григорьев А.В., Филипишин Д. А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия»
e-mail: vlad.zosimov.12anarhist@gmail.com, grigorievalvl@mail.ru, domaco@mail.ru

Зосимов В.В., Григорьев А.В., Филипишин Д.А. Проблемно-ориентированные редакторы онтологий для синтеза баз данных**.** *В тексте данной статьи представлен краткий обзор существующих редакторов онтологий. После их анализа рассматривается разработка проблемно-ориентированного редактора онтологий, реализованного с использованием языка программирования Python и системы управления базами данных PostgreSQL.*
Ключевые слова: *онтология, дерево, интерфейс, база данных, СУБД PostgreSQL, алгоритмы.*

Введение

Проблема автоматизации создания баз данных (БД) достаточно актуальна. Современные информационные технологии требуют эффективных инструментов для структурирования знаний и автоматизации создания специализированных баз данных. Онтологии, как формальные модели предметных областей, позволяют описывать концепты и связи между ними, что облегчает обработку и анализ информации [1, 2]. В данной работе рассматривается разработка проблемно-ориентированного редактора онтологий, который обеспечивает создание и управление онтологиями, а также их применение для синтеза баз данных посредством выполнения SQL-запросов, связанных с выбранными ответами.

Цель работы — анализ аналогичных систем, разработка требований, проектирование архитектуры и базы данных, реализация алгоритмов для создания и прохождения онтологий для синтеза БД на основе результатов прохождения, а также тестирование функциональности.

Онтологии применяются в поисковых системах, инженерии знаний и семантической интеграции. Аналогичные системы включают Protégé (открытый редактор с OWL-поддержкой и плагинами)[3], WebProtege (веб-версия для совместной работы), TopBraid Composer (коммерческий инструмент с SPARQL), NeOn Toolkit (модульная среда), OntoGraf (плагин для Protégé), Owlready2 (библиотека Python) и GraphDB (семантическая база данных).

После анализа существующих проблемно-ориентированных редакторов (ПОР) онтологий, а также изучения, анализа и обзора литературы по онтологиям, было принято решение о разработке ПОР онтологий. Данный редактор онтологий сфокусируется на синтезе баз данных через логический вывод, с поддержкой И-ИЛИ-деревьев и SQL-запросов.

1 Проектирование ПОР онтологий

В последние годы сформировалась парадигма компьютерных онтологий, основными признаками которых являются:

– иерархическая структура конечного множества понятий, описывающих заданную предметную область;

– структура онтологии может быть представлена онтографом, вершинами которого являются понятия (концепты), а дугами – семантические отношения (связи) между ними;

– концепты и связи между ними интерпретируются в соответствии с результатами;

– анализа электронных источников знаний (например, корпуса текстов) заданной

предметной области;

– онтограф должен быть представлен формально на одном из языков описания онтологий [3].

Онтология проектирования — это также формализованное описание знаний субъектов проектирования о процессе проектирования новых или модернизаций уже известных артефактов, включая знания о самом объекте проектирования и близких к нему по свойствам артефактов, а также тезаурус предметной области [4].

После анализа теории и существующих редакторов онтологий, была разработана диаграмма прецедентов для проектируемого ПО редактора онтологий (см. рис. 1).

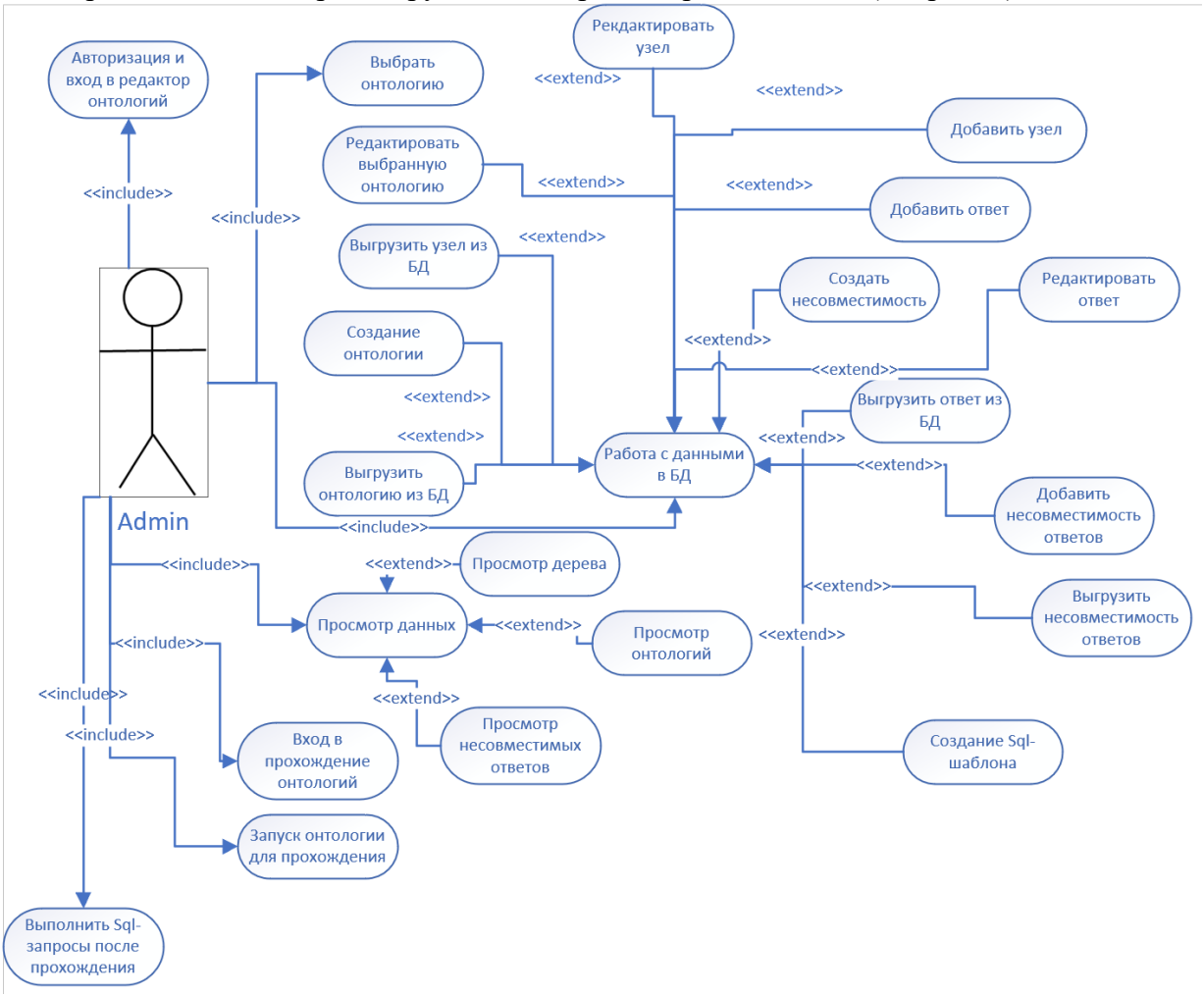

Рисунок 1 – Диаграмма прецедентов редактора онтологий

Диаграмма прецедентов UML – это диаграмма, показывающая связи между актёрами и действиями (прецедентами), которые они могут выполнять. Эта диаграмма позволяет точно описать полную функциональность программной системы [5].

Как видно из рис. 1, администратор может:

– создать новую онтологию, задав название и описание;

– редактировать существующие онтологии, включая изменение названия и описания;

– удалить онтологию и все связанные с ней узлы, и ответы;

– добавлять новые узлы (вопросы) к существующим онтологиям, задавая тип узла (AND/OR);

– редактировать текст узлов и их типы;

– удалять узлы и все связанные с ними ответы;

– добавлять ответы к узлам, включая текст ответа и связанные SQL-запросы;

- изменять текст ответов и связанные SQL-запросы;
- удалять ответы и узлы из системы;
- администратор может добавлять правила несовместимости между ответами;
- просматривать, редактировать и удалять существующие несовместимости;
- добавлять онтологии, узлы, ответы, несовместимости, шаблоны;
- запускать прохождение онтологии, и управлять процессом прохождения;
- выполнять собранные SQL-запросы после прохождения онтологии.

Пользователь может только использовать онтологии и синтезировать БД из собранных SQL-запросов.

Программа будет состоять из двух частей, в первой будет редактор онтологий, с которым работает администратор после авторизации, а во второй может так же работать пользователь, и через применение онтологий он может синтезировать БД.

2 Алгоритмы редактирования проблемно-ориентированных онтологий

Эффективность программы для редактирования и применения онтологий зависит от её алгоритмов, которые позволяют работать с деревом И-ИЛИ узлов и их ответов. Работа всей программы тесно связана с БД, на которой хранятся все данные.

2.1 Общий алгоритм работы загрузки и обновления дерева

Необходимо отметить, что основным алгоритмом программы как для редактирования, так и для прохождения онтологий, является алгоритм, в котором происходит структурирование и визуализация загруженной или созданной онтологии в виде И-ИЛИ дерева.

2.2 Алгоритм работы с корнем дерева

Корень дерева – это самый верхний уровень структуры онтологии, представляет ее как контейнер, хранит название и описание онтологии. К корню прикрепляется и создается вместе с ним первый ИЛИ-узел дерева, к которому крепятся остальные ответы и узлы. Это обеспечивает И-ИЛИ логику дерева

2.3 Алгоритмы работы с узлами и ответами

При работе с данными алгоритмами следует учесть, что к одному ответу можно прикрепить только один узел. В то время как для одного узла ограничение ответов не задано. Они обеспечивают гибкость и точность в моделировании логических связей, что критично для предотвращения несоответствий в онтологии и автоматизации синтеза БД.

2.4 Алгоритм удаления выбранного узла или ответа

Этот алгоритм важен для сохранения целостности онтологии, так как позволяет безопасно очищать устаревшие или ошибочные элементы без риска нарушения зависимостей при помощи каскадного удаления узла или ответа, включая все связанные элементы, с подтверждением действия. При проверке зависимости и обновлении дерева после удаления, предотвращает повреждение структуры онтологии.

2.5 Алгоритмы работы с SQL шаблонами и запросами

В данных алгоритмах присутствуют произвольные, пользовательские SQL-запросы, которые администратор может ввести напрямую для каждого ответа. Они более гибкие и позволяют реализовать нестандартные логики SQL-шаблоны – параметризованные структуры, которые генерируют стандартные SQL-команды на основе выбранных ответов в онтологии. Они прикрепляются к ответам как готовые блоки, обеспечивая автоматизацию и стандартизацию операций синтеза БД.

2.6 Алгоритм добавления несовместимости ответов

Этот алгоритм ключевой для поддержания логической корректности онтологии,

предотвращая выборы, ведущие к конфликтам и некорректным БД. Если несовместимый ответ будет выбран в будущем, то он блокирует другой ответ для выбора в будущем.

2.7 Алгоритмы работы применения выбранной онтологии

Данный алгоритм обеспечивает интерактивную и надежную обработку пользовательских выборов, что важно для точного сбора данных и переходов в И-ИЛИ дереве.

2.8 Алгоритм синтеза собранной БД

Этот алгоритм завершает цикл, обеспечивая последовательное и безопасное выполнение SQL-запросов для создания БД, что критично для целостности и работоспособности результата. Его значимость в обработке ошибок и порядке операций минимизирует риски неудач, повышая автоматизацию и надежность синтеза, что позволяет эффективно применять онтологии в реальных задачах инженерии знаний.

3 Результат разработки ПОР онтологий

После разработки алгоритмов, они были реализованы на языке Python, и были объединены в две основные подпрограммы для работы и применения онтологий.

Выводы

В ходе выполнения данной работы был проведен анализ существующих редакторов онтологий, таких как Protégé, Owlready2 и другие, после чего было решено разработать и реализовать ПОР онтологий для их создания и последующего их применения для синтеза БД.

Дальнейшие разработки и реализация алгоритмов для создания, редактирования и применения онтологий являются перспективными. Они могут включать интеграцию с другими СУБД (MySQL, Oracle). Также могут быть добавлены модули машинного обучения для автоматической генерации узлов на основе анализа текстов, улучшение пользовательского интерфейса для веб-версии (с использованием фреймворков типа Flask или Django). В дополнение, можно реализовать внедрение механизмов совместной работы в реальном времени.

Литература

1. Мошаров, А. Онтология и семантика // Habr. – UML: https://habr.com/ru/articles/848274/, свободный. – Загл. с экрана.

2. Чалая, Л. Э. Метод автоматического построения онтологических моделей с древовидной структурой концептов // CyberLeninka. – UML: https://cyberleninka.ru/article/n/metod-avtomaticheskogo-postroeniya-ontologicheskih-modeley-s-drevovidnoy-strukturoy-kontseptov/viewer, свободный. – Загл. с экрана.

3. Музен, М. А. Проект «Протеже»: взгляд в прошлое и в будущее // AI Matters. – 2015. – Т. 1, № 4. – Июнь. – UML: https://protege.stanford.edu/, свободный. – DOI: 10.1145/2557001.25757003.

4. Смирнов С. В. Онтология проектирования/ С.В. Смирнов, Н.М. Боргест, Е. А. Климакова // Новая техника. – 2011. – № 0. – С. 9–24. – ISSN 2223-9537. – Загл. с экрана.

5. Диаграммы прецедентов: крупным планом [электронный ресурс]. – Режим доступа: http://www.intuit.ru/studies/courses/1007/229/lecture/5962. – Загл. с экрана.

Zosimov V.V., Grigoriev A.V., Filipishin D.A. Problem-oriented editors of ontologies for database synthesis. *The text of this article provides a brief overview of existing ontology editors. After analyzing them, we consider the development of a problem-oriented ontology editor implemented using the Python programming language and the PostgreSQL database management system.*

Keywords*: ontology, tree, interface, database, PostgreSQL database management system, algorithms.*

УДК 004.8

ИССЛЕДОВАНИЕ ПОДХОДОВ К АВТОМАТИЗАЦИИ ПОДБОРА ПЕРСОНАЛА НА ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Исаенко В.В., Григорьев А.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра программной инженерии им. Л.П. Фельдмана
e-mail: provcom11@gmail.com, grigorievalvl@mail.ru

Исаенко В.В., Григорьев А.В. Исследование подходов к автоматизации подбора персонала на основе технологий искусственного интеллекта. *Рассмотрены современные подходы и методологические приёмы автоматизации процессов подбора персонала с использованием технологий искусственного интеллекта. Проанализированы последние российские публикации за 2024–2025 гг., выявлены нерешённые проблемы валидации моделей, обеспечение прозрачности решений и интеграции. На основе анализа сформулированы рекомендации по выбору методологий и направлений дальнейших исследований.*
Ключевые слова: автоматизация подбора персонала, искусственный интеллект, рекрутинг, системы рекомендаций.

Введение

Автоматизация подбора персонала становится одной из ключевых задач в управлении человеческими ресурсами ввиду роста объёмов входящих заявок, необходимости ускорения найма и повышения качества совпадения компетенций кандидатов с требованиями вакансий.

Современные предприятия и государственные организации стремятся внедрять интеллектуальные средства, которые помогают сократить время рекрутинга, снизить нагрузку на HR-специалистов и повысить объективность первичного отбора. Вместе с тем при практическом внедрении возникают методологические проблемы: выбор адекватных моделей для скрининга резюме, обеспечение интерпретируемости алгоритмов, корректная подготовка и аннотирование корпоративных данных, а также соблюдение этических и правовых норм при автоматической обработке персональных данных.

Целью работы является проведение методологического обзора подходов к автоматизации подбора персонала на базе технологий искусственного интеллекта, выделение основных архитектурных решений и оценка их пригодности для практической имплементации в российских условиях.

Анализ современных исследований и публикаций

За последние годы в российской литературе наблюдается заметный рост работ, посвящённых цифровизации HR-процессов и применению ИИ в рекрутменте.

Так, исследования Ляпунцовой (2024) рассматривают внедрение интеллектуальных инструментов на этапах отбора и оценки кандидатов, в том числе автоматизированных систем анализа видеособеседований и предиктивной оценки соответствия кандидата вакансии [1].

Работы Руденко (2024) дают обзор применения NLP-методов для автоматической обработки текстов вакансий и резюме, что позволяет строить более точные профили компетенций и автоматизировать этап ранжирования кандидатов [2].

Иванова (2024) и другие авторы анализируют влияние цифровых инструментов на процесс адаптации и удержания персонала, обращая внимание на необходимость интеграции HR-систем с корпоративной аналитикой и бизнес-процессами [3].

Доброчинский (2024) проводит детальный анализ возможностей ИИ для автоматизации HR-скрининга кандидатов и отмечает проблемы, связанные с верификацией качества автоматического отбора и риском предвзятости моделей [4].

Наконец, Наумов (2025) описывает практические сценарии применения голосовых ботов и транскрипционных сервисов в первичной фильтрации, подчёркивая экономический эффект и ограничения по качеству данных [5].

Нерешённые части общей проблемы

Несмотря на активное количество публикаций, ряд важных методологических аспектов остаётся недостаточно проработанным.

Во-первых, вопрос стандартизации представления данных о кандидатах и вакансии: в различных организациях форматы резюме и профилей существенно отличаются, что затрудняет переносимость обученных моделей.

Во-вторых, проблема интерпретируемости и объяснимости решений ИИ — для HR-менеджера важна не только рекомендация, но и аргументация, на основе каких признаков система отобрала кандидата.

В-третьих, верификация и валидация моделей на реальных корпоративных данных нередко опирается на неполные метрики (например, только точность ранжирования), тогда как важны бизнес-метрики: время закрытия вакансии, долгосрочная успешность найма и текучесть.

Наконец, вопросы этики и соответствия требованиям законодательства о персональных данных требуют разработки методик анонимизации и управления согласиями при обучении моделей.

Основные подходы и методология реализации

В литературе можно выделить несколько ключевых методологических направлений при построении систем автоматизации подбора персонала.

1. Сбор и предобработка данных. На первом этапе формируются корпуса резюме, вакансий и данных о предыдущих наймах. Практика российских исследователей подчёркивает важность нормализации полей резюме (опыт, образование, навыки), извлечения сущностей и маппинга на унифицированные онтологии компетенций [2-3]. Методологически это достигается с помощью правил-парсеров, комбинируемых с моделями распознавания сущностей (NER) и лексическими словарями отраслевых терминов.

2. Модели ранжирования и классификации. Для первичного скрининга используют ансамбли методов: классические алгоритмы машинного обучения (логистическая регрессия, SVM) в сочетании с методами градиентного бустинга и нейронными сетями для извлечения сложных признаков. Российские работы показывают, что комбинированные подходы дают лучший компромисс между объяснимостью и точностью: простые модели используются для контроля важных признаков, сложные — для захвата нелинейных зависимостей [1, 4].

3. Обработка естественного языка (NLP). Технологии обработки текста применяются для анализа текста резюме и вакансий, определения синонимичных навыков и соответствия терминологии. В российских исследованиях 2024 года отмечается успешное использование трансформерных архитектур для задач семантического поиска и сопоставления вакансий — при этом подчеркивается потребность в дообучении на специализированных корпоративных корпусах [2].

4. Системы рекомендаций и профильный матчинг. Рекомендательные механизмы используют коллаборативные и контентные модели, а также гибридные решения, которые учитывают характер вакансии, профиль команды и исторические данные найма. Методологически важна калибровка модели под бизнес-цели: например, при дефиците квалифицированных кадров предпочтительны решения с более «широким» подбором кандидатов, тогда как для узких вакансий — более строгие фильтры [3].

5. Интеллектуальные чат-боты и мультимодальные сервисы. Голосовые и текстовые боты применяются для первичного взаимодействия, проведения скрининговых интервью и сбора дополнительной информации. Практические исследования демонстрируют экономию времени рекрутеров, однако указывают на необходимость качественной интеграции с CRM и системами учёта вакансий [5].

Методологические рекомендации по валидации и оценке

Ключевая методологическая задача — корректная оценка качества ИИ-решений в HR. Авторы отечественных публикаций предлагают сочетать традиционные ML-метрики (precision/recall, AUC) с бизнес-ориентированными показателями (среднее время закрытия вакансии, retention hire). Для оценки устойчивости моделей рекомендуются методы перекрёстной валидации на стратифицированных подвыборках и тестирование на отложенных временных срезах (time-split), что имитирует реальную эксплуатацию модели. Важным шагом является проведение A/B-тестирования при развёртывании, с мониторингом влияния автоматизации на качество найма и опыт кандидатов.

Практические аспекты внедрения

При внедрении следует учитывать организационные и технические моменты: интеграция с текущими HRIS/ATS, обеспечение конвейера данных (ETL), настройки прав доступа и логирования. Рекомендуется поэтапный подход: пилотный проект на одном подразделении, оценка эффекта, доработка и масштабирование. Также важно иметь механизм «человека в цикле» для критических решений, чтобы сочетать скорость автоматизации и экспертную оценку HR-специалиста.

Этические и правовые соображения

Применение ИИ в рекрутменте затрагивает персональные данные и риски дискриминации. Методологически необходимо применять техники анонимизации, проводить аудит данных на предмет присутствия предвзятости, а также использовать методы объяснимости (explainable AI) для демонстрации причин отбора. В российских публикациях указывается на потребность в локальных регламентах и политике обработки данных, учитывающей действующее законодательство и внутренние корпоративные требования [4].

Выводы

Исследование показало, что в российской науке сформировались базовые методологические подходы к автоматизации подбора персонала: предобработка данных, комбинированные модели ранжирования, NLP-модули и рекомендательные системы. Основные проблемы связаны с практической валидацией моделей на корпоративных данных, интерпретируемостью решений и этическими аспектами. Перспективными направлениями являются разработка стандартных онтологий компетенций, методы контрфактической оценки предвзятости, а также исследования по интеграции мультимодальных данных (текст, аудио, видео) при уважении прав кандидатов. Также важна разработка методик оценки долгосрочного эффекта автоматизации на качество персонала и экономическую эффективность.

Литература

1. Ляпунцова Е.В. Оптимизация рекрутинговых процессов путём внедрения интеллектуальных технологий анализа данных [Электронный ресурс] / Вестник управления персоналом. 2024. – Режим доступа: https://cyberleninka.ru/article/n/optimizatsiya-rekrutingovyh-protsessov-putem-vnedreniya-intellektualnyh-tehnologiy-analiza-dannyh - Загл. с экрана..

2. Руденко Л.Г. Тренды применения искусственного интеллекта в HR-менеджменте [Электронный ресурс] / Журнал современных исследований. 2024. – Режим доступа: https://cyberleninka.ru/article/n/trendy-primeneniya-iskusstvennogo-intellekta-v-hr-menedzhmente - Загл. с экрана.

3. Иванова И.Г. Цифровые технологии в HR: современные тренды управления персоналом в России [Электронный ресурс] / Экономика и менеджмент. 2024. – Режим доступа: https://cyberleninka.ru/article/n/tsifrovye-tehnologii-v-hr-sovremennye-trendy-upravleniya-personalom-v-rossii - Загл. с экрана..

4. Доброчинский В.Б. Искусственный интеллект в рекрутменте: проблема анализа экономической эффективности [Электронный ресурс] / Информационные технологии в экономике. 2024. – Режим доступа: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-rekrutmente-problema-analiza-ekonomicheskoy-effektivnosti - Загл. с экрана.

5. Наумов В.А. Анализ возможностей искусственного интеллекта для автоматизации HR-скрининга кандидатов [Электронный ресурс] / Прикладные исследования. 2025. – Режим доступа: https://cyberleninka.ru/article/n/analiz-vozmozhnostey-iskusstvennogo-intellekta-dlya-avtomatizatsii-hr-skrininga-kandidatov - Загл. с экрана.

Isaenko V.V., Grigoriev A.V. Research on Approaches to the Automation of Personnel Recruitment Based on Artificial Intelligence Technologies. *The paper examines modern approaches and methodological techniques for automating personnel recruitment processes using artificial intelligence technologies. Recent Russian publications from 2024–2025 are analyzed, revealing unresolved issues of model validation, decision transparency, and system integration. Based on the analysis, recommendations are formulated for the selection of methodologies and directions for further research.*

Keywords: *recruitment automation, artificial intelligence, HR, recommendation systems.*

УДК 004.89

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ АДАПТИВНО ГЕЙМПЛЕЯ В КОМПЬЮТЕРНЫХ ИГРАХ

Кириленко Н.А., Прокопенко Е.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: kola22022003@bk.ru, prokopenko1515@rambler.ru

Кириленко Н.А., Прокопенко Е.В. Интеллектуальные системы адаптивного геймплея в компьютерных играх. *Выполнен анализ современных методов построения адаптивного геймплея с применением технологий искусственного интеллекта. Рассмотрены подходы к динамическому изменению сложности и поведения игровых агентов на основе анализа действий игрока. Особое внимание уделено использованию алгоритмов обучения с подкреплением и нейросетевых моделей для индивидуализации игрового процесса. Проанализированы инструменты реализации адаптивных систем в игровых движках Unreal Engine и Unity. Отмечены отечественные исследования в области интеллектуальных алгоритмов и их применения в интерактивных и обучающих системах.*

Ключевые слова: *геймплей, адаптация, интеллект, обучение, поведение, игра.*

Введение

Современные компьютерные игры становятся всё более сложными и интерактивными, а требования пользователей к качеству игрового опыта постоянно растут. Одним из ключевых направлений развития индустрии является адаптивный геймплей — способность игры подстраиваться под индивидуальные особенности, поведение и уровень навыков игрока [1]. Такой подход позволяет поддерживать интерес пользователя, избегать фрустрации и обеспечивать оптимальный баланс между вызовом и достижением.

Традиционные системы сложности, основанные на фиксированных настройках, не способны учитывать стиль игры конкретного пользователя. В результате одни игроки сталкиваются с чрезмерной сложностью, а другим, напротив, становится слишком легко. Эта проблема особенно заметна в проектах с нелинейной структурой и открытым миром, где поведение пользователя трудно предсказать заранее. Решение подобных задач стало возможным благодаря развитию искусственного интеллекта и методов анализа данных.

Интеллектуальные системы адаптивного геймплея используют машинное обучение, обучение с подкреплением и поведенческий анализ для динамического изменения игровых параметров. Такие системы могут автоматически регулировать количество противников, изменять стратегию поведения врагов, варьировать частоту появления ресурсов и даже перестраивать сценарий миссий в зависимости от успехов игрока.

Применение адаптивных моделей повышает вовлечённость, делает игровой процесс индивидуальным и способствует формированию устойчивого интереса. Кроме того, подобные технологии находят применение не только в развлечениях, но и в обучающих симуляторах и системах профессиональной подготовки, где интеллектуальная адаптация сложности позволяет учитывать уровень знаний и реакцию обучаемого [4].

Цель данной работы – рассмотреть принципы построения интеллектуальных систем адаптивного геймплея, классифицировать существующие подходы и определить возможности их реализации в современных игровых движках, таких как Unreal Engine и Unity.

1 Постановка задачи

Современные компьютерные игры стремятся обеспечить игроку максимально комфортный и интересный процесс взаимодействия. Однако при использовании традиционных

подходов к управлению сложностью игровой процесс нередко оказывается либо чрезмерно простым, либо, наоборот, слишком трудным. Это объясняется тем, что классические системы опираются на фиксированные уровни сложности, выбираемые пользователем до начала игры. Такой подход не учитывает индивидуальные особенности игрока, его стиль поведения, скорость реакции и опыт.

В результате создаётся противоречие между заранее установленными параметрами сложности и реальными возможностями пользователя. Игрок с низким уровнем навыков может быстро потерять интерес из-за частых неудач, а опытному пользователю, напротив, становится скучно из-за отсутствия вызова. Для устранения этой проблемы требуется создание интеллектуальной системы адаптации, способной автоматически анализировать действия игрока и подстраивать параметры игры в реальном времени.

Задача таких систем заключается в обеспечении динамического баланса между сложностью и удовлетворением от игрового процесса. Модель должна не просто изменять количественные параметры, например здоровье противников или частоту появления ресурсов, но и учитывать более сложные поведенческие факторы — стратегию действий, темп прохождения, стиль боя или склонность к исследованию игрового мира.

Решение этой задачи возможно благодаря применению методов искусственного интеллекта и машинного обучения, в частности алгоритмов обучения с подкреплением и анализа поведения. Такие методы позволяют игре «понимать» игрока, прогнозировать его действия и адаптировать сценарий или механику под конкретную ситуацию. [3] В результате создаётся система, в которой каждый игровой опыт становится индивидуальным, а уровень вовлечённости и удовлетворённости пользователя — максимально высоким.

2 Традиционные системы управления сложностью

Проблема управления сложностью в компьютерных играх решалась ещё задолго до появления интеллектуальных систем. На ранних этапах развития игровой индустрии разработчики использовали простейшие механизмы регулирования — фиксированные уровни сложности, выбираемые пользователем перед началом игры. Такой подход оставался стандартом в течение десятилетий и применялся в большинстве коммерческих проектов.

Традиционные системы сложности строятся на заранее определённых наборах параметров: скорости реакции противников, количестве здоровья, объёме наносимого урона и частоте появления игровых ресурсов. При этом поведение противников и структура уровней остаются неизменными независимо от стиля игры пользователя. Основное преимущество такого метода заключается в простоте реализации и предсказуемости результата, однако он полностью лишён гибкости и не способен адаптироваться под конкретного игрока.

Попытки улучшить классические системы сводились к добавлению простых динамических элементов, например увеличению или уменьшению сложности в зависимости от числа поражений игрока. Однако подобные механизмы имели ограниченный эффект и нередко воспринимались как искусственные. Игроки могли замечать внезапное упрощение или усложнение уровней, что негативно влияло на восприятие процесса. Таким образом, статическая модель регулирования сложности не обеспечивает достаточной глубины взаимодействия и не способна поддерживать оптимальный уровень вовлечённости в долгосрочной перспективе.

Переход к более гибким и адаптивным решениям стал возможен лишь с развитием методов искусственного интеллекта, которые позволили анализировать действия игрока и динамически изменять параметры игрового мира. Это направление стало логическим развитием традиционных подходов и основой современных интеллектуальных систем адаптивного геймплея.

3 Интеллектуальные методы адаптации сложности

Развитие технологий искусственного интеллекта открыло новые возможности для организации адаптивного геймплея. В отличие от статических систем, интеллектуальные моде-

ли способны анализировать действия игрока и изменять параметры игрового процесса в реальном времени. Основная идея заключается в том, чтобы поддерживать оптимальный уровень вызова — не допуская ни излишней простоты, ни чрезмерной сложности, тем самым обеспечивая максимальное вовлечение игрока. [5]

Одним из ключевых направлений является использование обучения с подкреплением (Reinforcement Learning), при котором алгоритм рассматривает игровой процесс как последовательность действий и реакций. Система получает «награду» за действия, которые поддерживают интерес пользователя, и «штраф» за ситуации, приводящие к разочарованию или потере вовлечённости. Со временем агент обучается выстраивать баланс между успехом и неудачей игрока, корректируя сложность заданий, поведение противников или распределение ресурсов.

Не менее важным направлением является применение нейросетевых моделей, способных анализировать стиль игры пользователя. Такие сети обучаются на статистике игровых сессий, выявляют закономерности в поведении игроков и формируют индивидуальные профили. На основе этих данных система может прогнозировать уровень навыков и предпочтения пользователя, адаптируя механику, сюжет или интерфейс. Применение подобных технологий наблюдается в современных играх с нелинейной структурой, где сценарий формируется динамически и отличается для каждого игрока.

Перспективным направлением также считается интеграция интеллектуальных алгоритмов в поведенческие деревья (Behavior Trees) и системы принятия решений (Decision Trees), используемые в игровых движках. Это позволяет противникам и союзным персонажам не просто реагировать на события, но и корректировать стратегию в зависимости от успехов или ошибок игрока. Подобные решения активно применяются в таких проектах, как Left 4 Dead, Resident Evil 4 и Alien: Isolation, где система динамически регулирует интенсивность действий, сохраняя баланс напряжения и интереса.

Таким образом, интеллектуальные методы адаптации сложности обеспечивают новый уровень взаимодействия между игроком и виртуальной средой. Они позволяют создавать игры, способные обучаться вместе с пользователем, изменяя сценарий и механику под индивидуальные особенности, что делает игровой опыт по-настоящему уникальным.

4 Реализация адаптивных систем в игровых движках

Современные игровые движки предоставляют разработчикам широкие возможности для создания интеллектуальных систем адаптации геймплея. Наиболее активно такие технологии развиваются в Unreal Engine и Unity, которые стали стандартом для построения интерактивных и обучающихся игровых систем.

В Unreal Engine ключевую роль в реализации адаптивного поведения играют модули Behavior Tree и Blackboard. Они позволяют формировать сложные схемы принятия решений, где поведение неигровых персонажей динамически изменяется в зависимости от состояния игрока и ситуации в игровом мире. Благодаря этому враги могут реагировать на уровень угрозы, нехватку ресурсов или успешные действия пользователя. Дополнительно разработчики могут использовать систему Environment Query System (EQS) для анализа пространства и выбора оптимальных решений агентом, что создаёт эффект «осознанного» поведения.

Кроме встроенных средств, Unreal Engine поддерживает интеграцию моделей машинного обучения через Python API и внешние библиотеки, включая TensorFlow и PyTorch. Это позволяет создавать адаптивные системы, в которых параметры сложности, сценарии и поведение противников регулируются на основе данных, собранных в ходе игровых сессий. Таким образом, игра может анализировать действия пользователей, выявлять закономерности и изменять баланс в реальном времени.

В Unity основным инструментом разработки интеллектуальных игровых систем является пакет ML-Agents Toolkit, предоставляющий возможность обучать агентов с использованием методов обучения с подкреплением. Разработчик может определить, какие действия агента считаются «успешными», и система самостоятельно обучает виртуального персонажа

оптимальной стратегии взаимодействия. Подобный подход используется для обучения моделей поведения врагов, балансировки уровней, а также адаптации механик под стиль игрока.

Обе платформы поддерживают применение аналитических систем для сбора и обработки данных. Это даёт возможность реализовывать адаптивный дизайн, основанный на статистике прохождений и поведении реальных пользователей. В результате игровой процесс становится более естественным и динамичным, а сложность — сбалансированной в зависимости от способностей игрока.

Таким образом, современные игровые движки позволяют не только использовать готовые алгоритмы адаптации, но и интегрировать собственные нейросетевые модели, что делает возможным создание полностью интеллектуальных игровых систем.

5 Отечественные разработки и исследования в области адаптивных игровых систем

В отечественной научной среде интерес к адаптивным системам в интерактивных приложениях активно растёт. Несмотря на то что основная масса прикладных решений разрабатывается зарубежными компаниями, российские исследователи вносят существенный вклад в развитие теоретических и алгоритмических основ адаптивного геймплея.

Так, в работах В.В. Глушкова и А.И. Касьянова рассматриваются вопросы оптимизации поведения виртуальных агентов на основе анализа игровых событий и вероятностных моделей. Учёные отмечают, что применение адаптивных алгоритмов позволяет создавать более устойчивые и реалистичные модели поведения, а также формировать динамическую сложность, зависящую от индивидуальных действий пользователя. [6]

И.И. Литвинова в своих публикациях исследует возможности внедрения элементов искусственного интеллекта в игровые и обучающие системы. В частности, ею предложены методы адаптации обучающих симуляторов под уровень подготовки пользователя, основанные на принципах машинного обучения и анализе игровых данных. [7] Аналогичные принципы применимы и к адаптивным играм, где корректировка сложности осуществляется через статистическую обработку игровых параметров и реакций игрока.

Кроме того, в российских журналах «Программные продукты и системы», «Вестник компьютерных и информационных технологий» и «Информатика и системы управления» регулярно публикуются статьи, посвящённые моделированию виртуальных сред, адаптивным алгоритмам и обучающим симуляторам. Эти работы демонстрируют потенциал применения интеллектуальных технологий не только в развлекательных, но и в образовательных проектах, где персонализация и адаптация играют ключевую роль.

В целом, отечественные исследования в области адаптивных игровых систем направлены на интеграцию алгоритмов машинного обучения и когнитивных моделей в структуры принятия решений. Несмотря на ограниченные ресурсы и меньшую коммерческую направленность, такие работы создают фундамент для будущих практических решений и внедрения интеллектуальных технологий в российский игровой рынок.

Выводы

Таким образом, развитие интеллектуальных систем адаптивного геймплея представляет собой одно из наиболее перспективных направлений современной игровой индустрии. В отличие от традиционных методов регулирования сложности, основанных на фиксированных параметрах, интеллектуальные модели позволяют формировать гибкую и динамичную среду, реагирующую на действия и особенности игрока.

Использование алгоритмов машинного обучения, обучения с подкреплением и поведенческого анализа обеспечивает высокий уровень персонализации игрового процесса. Такие системы способны не только поддерживать баланс сложности, но и формировать индивидуальные сценарии, что существенно повышает вовлечённость и интерес пользователей.

Отечественные исследования в этой области подтверждают значимость интеллектуальных методов и демонстрируют потенциал их применения не только в игровых проектах, но и в образовательных и симуляционных системах. В перспективе развитие адаптивного

геймплея будет способствовать появлению полностью персонализированных интерактивных сред, где каждый игрок получает уникальный опыт, соответствующий его стилю, интересам и уровню подготовки.

Литература

1. Ханик, Р. Динамическое регулирование сложности в компьютерных играх [Текст] / Р. Ханик // Материалы семинара AAAI по проблемам игрового ИИ. – 2005. – С. 45–51.

2. Тогелиус, Дж., Шакер, Н., Нельсон, М. Искусственный интеллект и игры [Текст] / Дж. Тогелиус, Н. Шакер, М. Нельсон. – Лондон: Springer, 2016. – 312 с.

3. Саттон, Р., Барто, Э. Обучение с подкреплением: введение [Текст] / Р. Саттон, Э. Барто. – Москва : ДМК Пресс, 2020. – 548 с.

4. Эпик Геймс. Документация Unreal Engine 5: искусственный интеллект и поведенческие деревья [Электронный ресурс]. – Режим доступа: https://docs.unrealengine.com/5.0/en-US/ai-and-behavior-trees-in-unreal-engine/. – Загл. с экрана.

5. Юнити Текнолоджиз. ML-Agents Toolkit Documentation [Электронный ресурс]. – Режим доступа: https://unity.com/products/machine-learning-agents. – Загл. с экрана.

6. Глушков, В. В., Касьянов, А. И. Алгоритмы процедурной генерации игровых пространств [Текст] / В. В. Глушков, А. И. Касьянов // Вестник компьютерных и информационных технологий. – 2020. – № 7. – 52–59 с.

7. Литвинова, И. И. Интеллектуальные системы генерации виртуальных миров [Текст] / И. И. Литвинова // Программные продукты и системы. – 2019. – № 3. – 48–55 с.

Kirilenko N.A., Prokopenko E.V. Intelligent Systems of Adaptive Gameplay in Computer Games. *The paper analyzes modern methods of building adaptive gameplay using artificial intelligence technologies. Various approaches to dynamically adjusting game difficulty and agent behavior based on player actions are considered. Special attention is paid to the use of reinforcement learning algorithms and neural network models for personalizing the gaming experience. Tools for implementing adaptive systems in game engines such as Unreal Engine and Unity are analyzed. Russian research in the field of intelligent algorithms and their application in interactive and educational systems is also highlighted.*
Keywords: *gameplay, adaptation, intelligence, learning, behavior, game.*

УДК 004.89

ИНТЕЛЛЕКТУАЛЬНАЯ ПЛАТФОРМА ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ В HR

Ковальчук А.Р., Прокопенко Е.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: MrPanda20000@mail.ru, prokopenko1515@rambler.ru

Ковальчук А.Р., Прокопенко Е.В. Интеллектуальная платформа поддержки принятия решений в HR. *Выполнен анализ современных методов построения адаптивных HR-систем с применением методов анализа данных, машинного обучения и оптимизации. Рассмотрены подходы к прогнозированию кадрового спроса, оценке риска текучести и динамическому формированию расписаний на основе поведенческих и операционных метрик сотрудников. Особое внимание уделено интеграции предиктивных моделей и оптимизационных движков в корпоративные и облачные платформы (Workday, SAP SuccessFactors, Odoo/1C), а также использованию event-driven архитектур и ML-фреймворков для онлайн-адаптации. Проанализированы отечественные разработки и практические сценарии применения в корпоративном управлении, образовании и службах занятости.*

Ключевые слова: *HR-ИС, адаптация, прогнозирование, оптимизация, машинное обучение, управление персоналом.*

Введение

Современные организации опираются на данные и информационные технологии для принятия управленческих решений, а требования к эффективности и адаптивности HR-процессов постоянно растут. Информационные системы управления персоналом (HR-ИС) уже давно перестали быть простыми реестрами работников: они должны поддерживать подбор, планирование, оценку и развитие сотрудников в условиях быстро меняющейся среды. Для этого необходимы не только надёжные программные решения, но и формализованные принципы построения систем, способных учитывать бизнес-цели, ограниченные ресурсы и человеческий фактор.

Традиционные HR-ИС часто опираются на статические правила и ручные процедуры, что приводит к медленной реакции на изменения спроса, неверной оценке потребности в кадрах и неэффективному распределению ресурсов. В результате организации сталкиваются с избыточными затратами, нехваткой компетенций в критические периоды или низкой мотивацией персонала. Проблема усугубляется при работе с распределёнными командами и гибридными моделями занятости, где предсказуемость процессов невысока.

Решение таких задач становится возможным благодаря интеграции методов анализа данных, математического моделирования и интеллектуальных алгоритмов в архитектуру HR-ИС. Модели прогнозирования, оптимизационные методы, алгоритмы машинного обучения и сценарное моделирование позволяют автоматически оценивать потребности, формировать расписания, подбирать кандидатов и прогнозировать риски текучести с учётом стохастических факторов и ограничений бизнеса.[2][4]

Применение интеллектуальных компонентов повышает адаптивность системы, снижает операционные затраты и улучшает качество управленческих решений. Такие подходы актуальны не только для корпоративного управления персоналом, но и для образовательных учреждений, служб занятости и сервисов профессиональной переподготовки, где требуется индивидуализация планов развития и эффективное распределение ресурсов.

Цель данной работы — сформулировать принципы построения информационной системы управления персоналом с интегрированными математическими и интеллектуальными модулями, классифицировать подходы к решению ключевых задач и оценить возможности их внедрения в современных программных и облачных платформах.

1 Постановка задачи

Современные организации требуют гибких и быстрых решений в области управления персоналом, однако многие существующие HR-системы опираются на статические правила и шаблоны, задаваемые вручную. Это приводит к несоответствию между текущими потребностями бизнеса и реальным состоянием кадров: одни подразделения остаются перегружены, другие — недозагружены; приоритеты подбора и развития часто отстают от изменений в спросе на компетенции. Классические подходы не учитывают динамики спроса, индивидуальных характеристик сотрудников и неопределённости внешних факторов (сезонность, проектные пики, текучесть).

В результате возникает противоречие между жёстко заданными кадровыми политиками и переменчивыми организационными условиями. Менеджеры получают либо избыточные кадровые расходы при избытке персонала, либо снижение качества услуг и выгорание сотрудников при дефиците кадров. Слабая адаптивность систем особенно заметна в распределённых командах и гибридных графиках, где поведение и доступность сотрудников меняются динамично и предсказать их заранее сложно. Для устранения этой проблемы необходима интеллектуальная система управления персоналом, способная в реальном времени анализировать данные, предсказывать потребности и автоматически предлагать оптимальные решения.[1]

Задача такой системы заключается в обеспечении динамического баланса между требованиями бизнеса, доступными ресурсами и удовлетворённостью сотрудников. Модель должна не ограничиваться простыми правиловыми корректировками (например, смена количества людей в смене), но и учитывать комплексные поведенческие и организационные факторы — квалификацию и многофункциональность сотрудников, мотивацию и риск увольнения, предпочтения по графику, влияние перекрестных навыков в командах и критичность задач. Кроме того, система должна работать с неполными, шумными и имеющимися в разное время данными, адаптируясь к изменениям в режиме онлайн или периодических пересчётов.

Решение этой задачи возможно при сочетании методов анализа данных, прогнозирования и оптимизации: статистические и машинно-обучающие модели для предсказания спроса и риска текучести, целочисленные и стохастические оптимизационные модели для планирования и распределения ресурсов, а также эвристики и алгоритмы автоматической корректировки для работы в реальном времени. Интеграция таких модулей позволит системе «понимать» текущую кадровую ситуацию, прогнозировать её развитие и предлагать решения — от перераспределения нагрузки и временного найма до программ развития и удержания сотрудников.

В результате должна появиться адаптивная HR-ИС, которая снижает операционные затраты, уменьшает риски нехватки кадров, повышает удовлетворённость и эффективность сотрудников, а также поддерживает принятие обоснованных управленческих решений в условиях неопределённости.

2 Традиционные системы управления персоналом

Традиционные HR-системы опираются на жёстко заданные правила и шаблоны: штатные расписания, фиксированные регламенты подбора, заранее установленные KPI и периодические ручные пересмотры. Такие подходы просты в внедрении и предсказуемы, но не учитывают динамику спроса, индивидуальные особенности сотрудников и непредвиденные события (пиковая нагрузка, текучесть, удалённая работа).

В результате возникают типичные проблемы: избыточные затраты при простое сотрудников, перегрузка и выгорание при дефиците ресурсов, низкая точность подбора компе-

тенций и медленная реакция управления на изменения. Попытки улучшения через простые автоправила (увеличение смен при росте нагрузки и т.п.) дают ограниченный эффект и часто воспринимаются как грубые корректировки.

Необходим переход к адаптивным решениям: система должна в реальном времени анализировать доступные данные, прогнозировать потребности и автоматически предлагать оптимальные действия — перераспределение задач, гибкие графики, временный найм или программы удержания.[3]

Задача — разработать HR-ИС, способную динамически сопоставлять требования бизнеса, доступные ресурсы и предпочтения сотрудников, работая с шумными и неполными данными и обеспечивая цикл «данные → аналитика → решение → обратная связь».

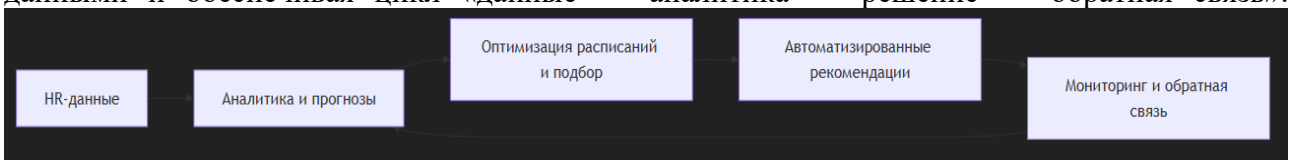

Рисунок 1 Постановка задачи

3 Реализация адаптивных систем в управлении персоналом

Современные HR-платформы и сервисы предоставляют богатые возможности для внедрения адаптивных решений управления персоналом. Крупные облачные системы (Workday, SAP SuccessFactors, Oracle HCM) обеспечивают расширяемые API, коннекторы для данных и встроенные аналитические модули, что позволяет интегрировать модели прогнозирования и оптимизационные движки прямо в бизнес-процессы. Открытые и локальные решения (Odoo, 1C, OrangeHRM) дают гибкость для разработки собственных модулей планирования и рекомендаций.

Ключевые компоненты реализации — сбор и консолидация данных (временные метки присутствия, эффективность, компетенции, вакансии), пайплайн обработки (ETL/streaming), аналитические модели (регрессии, классификация, модели времени до события) и движок принятия решений (оптимизация расписаний, подбор кандидатов, план развития). Интеграция с ML-фреймворками (scikit-learn, TensorFlow, PyTorch) и оптимизаторами (CP-SAT, CPLEX/Gurobi или эвристики) позволяет автоматически формировать расписания, предсказывать риск текучести и рекомендовать мероприятия по удержанию.

Для реального времени применяют event-driven архитектуры (Webhook/Kafka), микросервисы и облачные функции, что обеспечивает быстрый отклик на изменения — например, перераспределение смен при неожиданном снижении числа доступных сотрудников. Визуализация и панели управления дают менеджерам прозрачные рекомендации и сценарии «что-если», а механизмы A/B-тестирования и непрерывной валидации позволяют оценивать эффективность адаптивных правил.

Практическое применение включает: динамическое составление графиков с учётом навыков и предпочтений, прогноз кадровой потребности и автоматизированный рекрутинг, персонализированные планы обучения и карьеры, а также автоматические триггеры для удержания сотрудников. При этом важно учитывать вопросы безопасности данных, соответствие GDPR/локальным регуляциям и контроль доступа к аналитике.

В результате современные платформы позволяют не только встраивать готовые алгоритмы, но и развивать собственные интеллектуальные модули, обеспечивая адаптивность HR-процессов и повышение оперативности управленческих решений.[5]

Выводы

Таким образом, развитие интеллектуальных адаптивных информационных систем управления персоналом представляет собой одно из перспективных направлений современного управления и HR-технологий. В отличие от традиционных статических правил и ручных процедур, интеллектуальные модели на основе анализа данных и оптимизации обеспе-

чивают гибкую реакцию на изменения спроса, доступность сотрудников и уровень компетенций.

Применение методов машинного обучения, прогнозной аналитики и оптимизационных алгоритмов позволяет не только прогнозировать потребности и риски (например, текучесть), но и автоматически генерировать практические решения — динамические графики, рекомендации по найму и программам удержания, персонализированные планы обучения. Это повышает оперативность управленческих решений, снижает операционные затраты и улучшает удовлетворённость персонала.

Современные платформы и облачные архитектуры дают технические возможности для интеграции таких моделей в бизнес-процессы и реализации циклов «данные → аналитика → действие → валидация». Отечественные и международные исследования подтверждают эффективность подходов, а практические внедрения демонстрируют экономический и организационный эффект в реальных компаниях.

В перспективе развитие адаптивных HR-ИС приведёт к появлению более устойчивых и персонализированных рабочих сред: автоматизированной оркестрации ресурсов в реальном времени, индивидуальных траекторий развития сотрудников и гибкого сочетания постоянного и временного кадрового состава. Это позволит организациям быстрее адаптироваться к изменчивым условиям рынка и повышать конкурентоспособность.

Литература

1. Angrave M., Charlwood A., Kirkpatrick I., Lawrence M., Stuart M. HR and analytics: why HR is set to fail [Текст] / M. Angrave, A. Charlwood, I. Kirkpatrick, M. Lawrence, M. Stuart // Human Resource Management Journal. – 2016. – Vol. 26, No. 1. – P. 1–11.

2. Waber B. People Analytics: How Social Sensing Technology Will Transform Business and What It Tells Us About the Future of Work [Текст] / B. Waber. – Boston: Harvard Business Review Press, 2013. – 256 с.

3. Edwards M., Edwards K. Predictive HR Analytics: Mastering the Future of Workforce Planning [Текст] / M. Edwards, K. Edwards. – London: Kogan Page, 2016. – 240 с.

4. Workday. Workday — People Analytics и Prism Analytics [Электронный ресурс]. – Режим доступа: https://www.workday.com/ – Загл. с экрана.

5. SAP. SAP SuccessFactors — People Analytics (документация и решения) [Электронный ресурс]. – Режим доступа: https://www.sap.com/products/successfactors.html – Загл. с экрана.

A.R. Kovalchuk, E.V. Prokopenko,. An Intelligent Platform for Decision Support in HR..
This article analyzes modern methods for building adaptive HR systems using data analysis, machine learning, and optimization. Approaches to forecasting talent demand, assessing turnover risk, and dynamically generating schedules based on employee behavioral and operational metrics are considered. Particular attention is paid to the integration of predictive models and optimization engines into corporate and cloud platforms (Workday, SAP SuccessFactors, Odoo/1C), as well as the use of event-driven architectures and ML frameworks for online onboarding. Domestic developments and practical application scenarios in corporate management, education, and employment services are analyzed.
Keywords: *HR IS, onboarding, forecasting, optimization, machine learning, HR management.*

УДК 004.89

HR-АНАЛИТИКА КАК ИНСТРУМЕНТ ДЛЯ ОБРАЗОВАТЕЛЬНЫХ ПРОГРАММ И НАУЧНЫХ ИССЛЕДОВАНИЙ

Ковальчук А.Р., Прокопенко Е.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: MrPanda20000@mail.ru, prokopenko1515@rambler.ru

Ковальчук А.Р., Прокопенко Е.В. HR-аналитика как инструмент образовательных программ и научных исследований. *Выполнен анализ современных методов реализации адаптивных систем управления персоналом в контексте образовательных программ и научных проектов с применением методов анализа данных, машинного обучения и оптимизации. Рассмотрены подходы к прогнозированию спроса на образовательные программы и научные направления, оценке риска отсева учащихся и сотрудников, а также динамическому формированию расписаний занятий, лабораторных работ и исследовательских мероприятий на основе поведенческих и операционных метрик. Особое внимание уделено интеграции предиктивных моделей и оптимизационных движков в образовательные платформы и исследовательские инфраструктуры (Moodle, Open edX, Canvas, локальные LMS, а также корпоративные HR/образовательные решения), а также использованию событийно-ориентированной архитектуры и ML-фреймворков для онлайн адаптации траекторий. Проанализированы отечественные разработки и практические сценарии применения в образовании и научных исследованиях.*
Ключевые слова: *HR-ИС, образовательная аналитика, адаптация, прогнозирование, оптимизация, машинное обучение, управление персоналом в образовании, научные исследования.*

Введение

Современные образовательные программы и научные исследования всё чаще опираются на данные о трудовых рынках, компетенциях и карьерных траекториях. HR-аналитика превращает такие данные в практические инсайты: выявляет пробелы в навыках, оценивает эффективность обучающих мероприятий и помогает формировать релевантные учебные траектории. В условиях быстрого обновления знаний и спроса на новые компетенции это позволяет сделать образование более целевым и прикладным, а научные проекты — лучше согласованными с отраслевыми потребностями.

Традиционные подходы к проектированию курсов и отбору участников исследований часто базируются на экспертных суждениях и опросах, что даёт ограниченную и запаздывающую картину. Интеграция HR-данных (резюме, оценочные сессии, карьерная динамика, обратная связь работодателей) с аналитическими моделями повышает точность планирования программ обучения и релевантность исследовательских выборок.

1 Постановка задачи

Необходимо разработать подходы и инструменты HR-аналитики для поддержки образовательных программ и научных исследований, обеспечивающие:

- сбор и консолидацию разнородных данных (портфолио выпускников, вакансии, оценки компетенций, результаты обучения);
- моделирование компетентностных профилей и выявление разрывов между предложением образования и требованиями рынка;
- прогнозирование спроса на навыки и рекомендацию адаптивных учебных траекторий и модулей;

- поддержку отбора участников и формирования исследовательских когорт на основе репрезентативных и релевантных критериев;

- работу с неполными и шумными данными при соблюдении требований приватности и этики обработки персональной информации.

Решение должно сочетать статистические методы, алгоритмы машинного обучения и визуализацию результатов для принятия решений преподавателями, администрацией и исследователями [1].

2 Ограничения традиционных HR-систем

Классические HR-системы основаны на статических правилах: фиксированные штатные расписания, периодические оценки, ручной подбор и жёсткие KPI.

Такие подходы дают предсказуемость и простоту внедрения, но плохо реагируют на внезапные изменения спроса, сезонные пики и гибридные модели занятости.

Последствия: перегрузка команд или простаивающие ресурсы, задержки в подборе, повышенные операционные расходы и снижение вовлечённости сотрудников.

Попытки автоматизации через простые правила (увеличить смены при росте нагрузки) часто оказываются недостаточно гибкими и воспринимаются менеджментом как «грубая» корректировка [2, 4].

3 Внедрение адаптивных HR-решений в корпоративных платформах

Современные платформы (облачные и локальные) предоставляют API, ETL-пайплайны и возможности интеграции ML-моделей, что позволяет строить адаптивные модули прогнозирования и оптимизации.

Ключевая архитектура: сбор и консолидация данных → модели прогнозирования (трудоёмкость, текучесть) → оптимизационный движок (расписания, подбор) → автоматизированные рекомендации и триггеры [5.]

Для онлайн-адаптации применяют event-driven подходы (Kafka/Webhook), микросервисы и A/B-тестирование; ML-фреймворки (scikit-learn, TensorFlow, PyTorch) и оптимизаторы (CP-SAT, эвристики) решают прикладные задачи.

Внедрение повышает оперативность реакций (перераспределение смен, временный найм), персонализацию развития сотрудников и эффективность решений при сохранении требований безопасности данных и соответствия регуляциям [3].

Выводы

Развитие интеллектуальных адаптивных информационных систем управления персоналом представляет собой не просто эволюцию существующих HR-технологий, а качественный сдвиг в способах принятия управленческих решений. Интеграция аналитики, предиктивных моделей и оптимизационных механизмов позволяет перейти от реактивного управления к проактивному — система не только фиксирует события, но и предсказывает потребности, оценивает риски и предлагает конкретные сценарии действий. В результате организации получают инструмент, который повышает оперативность, уменьшает издержки и улучшает удовлетворённость сотрудников.

Ключевые эффекты внедрения интеллектуальных модулей:

Повышение точности планирования кадров: предиктивные модели учитывают сезонность, проектные пики и тренды спроса, что снижает вероятность как нехватки, так и излишка персонала.

Снижение текучести и удержание талантов: раннее обнаружение признаков риска увольнения и персонализированные мероприятия по удержанию уменьшают утраты компетенций.

Оптимизация операционных затрат: автоматизированное составление графиков и перераспределение задач снижает овертаймы и повышает производительность.

Индивидуализация развития: персонализированные планы обучения и карьерные дорожки повышают мотивацию и целевой рост компетенций.

Одновременно с преимуществами появляются и существенные инженерные и организационные вызовы:

Качество и полнота данных. Неполные, разрозненные и слабо структурированные HR-данные требуют продуманного ETL-пая и инструментов очистки.

Конфиденциальность и соответствие регуляциям. Обработка персональных данных должна соответствовать GDPR и локальным требованиям, требуя шифрования, управления доступом и прозрачных политик.

Интерпретируемость моделей. Для принятия управленческих решений важно иметь объяснимые модели или инструменты объяснения (XAI), чтобы менеджеры доверяли рекомендациям.

Интеграция в бизнес-процессы. Технологии должны сочетаться с организационными процедурами и политиками, иначе автоматизация не принесёт ожидаемого эффекта.

Оценка эффективности. Необходимо внедрять метрики (влияние на текучесть, показатель заполненности вакансий, экономия затрат, удовлетворённость сотрудников) и механизмы A/B-тестирования.

Практические рекомендации для внедрения:

Начинать с пилота: выбрать узкую бизнес-задачу (динамическое расписание на одну операционную единицу, прогноз текучести в критических ролях) и оценить эффект.

Построить надёжный дата-пайплайн: консолидация данных из ATS, ERP, time-tracking, LMS и опросов, с валидацией и версионированием.

Комбинировать подходы: использовать статистические и ML-модели для предсказания, оптимизацию (deterministic/стохастические/эвристики) для принятия решений, а бизнес-правила — для соблюдения ограничений.

Внедрять мониторинг и цикл обучения моделей: непрерывная оценка качества прогнозов и автоматическое обновление моделей по мере поступления новых данных.

Уделять внимание правам доступа и этике: минимизация объёма личных данных, прозрачные политики и коммуникация с сотрудниками.

Перспективы развития: сочетание он-лайн адаптации (real-time event-driven реакции) и офлайн стратегического планирования даст компаниям возможность и быстро реагировать на внезапные изменения, и закладывать долгосрочные кадровые стратегии. Дальнейший прогресс в области интерпретируемого машинного обучения, federated learning (для защиты данных) и автоматизации оптимизационных движков сделает адаптивные HR-ИС ещё более безопасными и эффективными.

Литература

1. Waber, B. People Analytics: How Social Sensing Technology Will Transform Business and What It Tells Us About the Future of Work [Текст] / B. Waber. – Boston: Harvard Business Review Press, 2013. – 256 с.

2. Edwards, M., Edwards, K. Predictive HR Analytics: Mastering the Future of Workforce Planning [Текст] / M. Edwards, K. Edwards. – London: Kogan Page, 2016. – 240 с.

3. Long, P., Siemens, G. Penetrating the Fog: Analytics in Learning and Education [Текст] / P. Long, G. Siemens // EDUCAUSE Review. – 2011. – Vol. 46, No. 5. – P. 30–40.

4. Ferguson, R. Learning analytics: drivers, developments and challenges [Текст] / R. Ferguson // Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK). – 2012. – P. 1–8.

5. Workday. People Analytics и решения для образования [Электронный ресурс]. – Режим доступа: https://www.workday.com/ – Загл. с экрана.

A.R. Kovalchuk E.V. Prokopenko. HR Analytics as a Tool for Educational Programs and Scientific Research. *This article analyzes modern methods for implementing adaptive HR management systems in the context of educational programs and scientific projects using data analysis, machine learning, and optimization. Approaches to forecasting demand for educational programs and research areas, assessing the risk of student and employee attrition, and dynamically scheduling classes, labs, and research activities based on behavioral and operational metrics are considered. Particular attention is paid to the integration of predictive models and optimization engines into educational platforms and research infrastructures (Moodle, Open edX, Canvas, local LMS, and corporate HR/educational solutions), as well as the use of event-driven architecture and ML frameworks for online trajectory adaptation. Domestic developments and practical application scenarios in education and scientific research are analyzed.*

Keywords: *HR-IS, educational analytics, adaptation, forecasting, optimization, machine learning, personnel management in education, scientific research.*

УДК 004.9

ИССЛЕДОВАНИЕ СИСТЕМЫ УПРАВЛЕНИЯ ОТДЕЛОМ КАДРОВ ПРЕДПРИЯТИЯ НА ОСНОВЕ СИСТЕМНОГО АНАЛИЗА

Комарницкий В.Р.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: prestig170814ttt@gmail.com

Комарницкий В.Р. Исследование системы управления отделом кадров предприятия на основе системного анализа. *Выполнен краткий сравнительный анализ предложенных моделей и методов, что позволит значительно повысить эффективность и устойчивость кадровых процессов, обеспечит организацию необходимым аналитическим инструментарием для принятия обоснованных управленческих решений и стратегического планирования.*
Ключевые слова: *системный анализ; моделирование систем; автоматизация HR-процессов; кадровая эффективность.*

Введение

Современные предприятия функционируют в условиях высокой конкуренции, динамичных изменений на рынке труда, новых требований к управлению персоналом и необходимости повышать эффективность работы кадровых служб. В этой связи важнейшим направлением является разработка системных подходов к управлению отделами кадров, позволяющих комплексно и гибко реагировать на изменения внешней и внутренней среды организации.

Эффективная система управления отделом кадров должна обеспечивать оптимальное согласование процессов привлечения, обучения, мотивации и оценки сотрудников, а также учитывать взаимосвязи между этими процессами. В этом контексте системный анализ как методика позволяет моделировать сложные межотраслевые связи и выявлять слабые места в управлении, что важно для повышения конкурентоспособности предприятия.

Связь с научными и практическими задачами

Научное обоснование и практическая реализация системных решений в управлении кадровым отделом актуальны в связи с ростом требований к автоматизации и интеграции информационных систем, а также необходимости формирования стратегии профессионального развития кадров. В рамках научных исследований повышается интерес к моделированию управленческих систем и внедрению современных методов аналитики.

Практическая значимость заключается в возможности использования системного анализа для разработки управленческих алгоритмов, автоматизированных систем поддержки решений, а также в повышении прозрачности и эффективности управленческих процессов.

Анализ последних исследований

Научные исследования в области управления человеческими ресурсами за последние годы имеют тенденцию к интеграции системных методов и информационных технологий. Среди ключевых работ можно выделить исследования, посвящённые моделированию бизнес-процессов HR с помощью различных методов системного анализа и автоматизации.

Например, работы Smith (2020) и Ivanov (2021) показывают, что моделирование процессов помогает выявлять узкие места в управлении персоналом, оптимизировать ресурсы и повышать производительность. В сфере автоматизированных решений особое внимание уделяется внедрению ERP-систем и специализированных платформ для автоматизации HR-

функций, таких как SAP HR, Oracle HR и др.

Тем не менее, большинство исследований сосредоточены на отдельных компонентах системы, а также на внедрении программных решений, не раскрывая их системной взаимосвязи. Практически отсутствуют комплексные модели, объединяющие все ключевые процессы управления кадровым отделом в единую системную структуру.

Нерешённые части проблемы:

- Отсутствие унифицированных моделей, отражающих взаимосвязи между подбором, обучением, мотивацией, оценкой и управлением знаниями;
- Недостаточная разработанность методик внедрения системного анализа в практическую деятельность HR-отделов;
- Неучтённость интеграции современных информационных технологий и аналитики данных для повышения управляемости системы.

Особую актуальность приобретает разработка таких моделей, способных обеспечить комплексное видение и контроль всех факторов, влияющих на эффективность управления кадров.

Постановка проблемы

Общий контекст

Обеспечение эффективности управления отделом кадров предполагает учитывать множество взаимосвязанных процессов, а также их взаимодействие с внешней средой предприятия. Ключевыми аспектами являются точное планирование потребности в кадрах, подбор квалифицированных специалистов, обучение и развитие, мотивация и оценка эффективности работы сотрудников.

Основные научные и практические задачи

Научная задача — создание моделей, способных комплексно отражать системные связи и взаимовлияния внутри кадровой службы, что обеспечивает основу для принятия управленческих решений на системном уровне.

Практическая задача — разработка методик и инструментов, позволяющих реализовать эти модели в условиях конкретных предприятий с учетом их специфики и особенностей бизнес-процессов.

Вызовы и ограничения

- Недостаточная гибкость существующих систем управления;
- Высокая сложность и многоуровневость управляемых процессов;
- Недостаток методологических подходов, объединяющих системный анализ и автоматизацию;
- Неэффективное использование информационных технологий, особенно в условиях цифровизации.
-

Цель и задачи исследования

Цель

Разработать комплексную модель системы управления отделом кадров предприятия на основе методов системного анализа, которая позволяет повысить эффективность, прозрачность и адаптивность кадровых процессов.

Задачи

- Провести обзор существующих методов и подходов моделирования HR-процессов;
- Формализовать структуру и взаимодействия элементов системы управления кадров;
- Разработать модель системного анализа, отражающую ключевые процессы и связи;
- Провести апробацию модели на реальных примерах предприятий;
- Предложить рекомендации по внедрению системных решений в практику управления

кадровыми ресурсами.

Методология исследования

Моя методология основывается на сравнительном анализе существующих моделей управления, методах системного анализа, моделировании бизнес-процессов, а также на использовании инструментов автоматизации и аналитики данных. Ключевые этапы исследования включают:

- Аналитический: анализ существующих подходов, сбор эмпирических данных;
- Теоретический: формализация модели, разработка методических основ;
- Практический: тестирование модели на реальных кейсах предприятий.

Для моделирования применялись такие инструменты, как диаграммы потоков данных, UML-диаграммы, модели системных функций, а также системы автоматизированной обработки информации.

Анализ существующих подходов и теоретическая база

Модели управления HR

Классические модели характеризуют процессы подбора, адаптации, обучения, мотивации и оценки сотрудников как отдельные функции, интегрированные в общую систему. Однако зачастую они реализуются фрагментарно, что ограничивает возможности комплексного управления.

Системный анализ в управлении

Методы системного анализа позволяют рассматривать систему как целостную сущность, выявлять взаимосвязи, взаимозависимости и взаимовлияния элементов. В управлении людскими ресурсами используются такие методы, как моделирование бизнес-процессов, построение системных динамических моделей, анализ взаимодействий через структурные и функциональные диаграммы.

Современные информационные технологии

Внедрение ERP-систем, платформ для автоматизации HR, аналитических систем Big Data, машинного обучения и искусственного интеллекта позволяют собирать, обрабатывать и анализировать большие объемы данных о персонале, что создает новые возможности для системного управления.

Несовершенства существующих подходов:

- Отсутствие унифицированных моделей, отражающих взаимосвязи процессов;
- Недостаточная адаптивность существующих систем под специфику предприятия;
- Недостаточное использование аналитических инструментов для прогнозирования и планирования.

Это подчеркивает необходимость разработки новых системных методов и моделей.

Построение модели системы управления кадровым отделом

Структурные компоненты

Модель включает следующие основные элементы:

- **Входы:** потребности в кадрах, кадровый фонд, внешние и внутренние факторы;
- **Процессы:** подбор, адаптация, обучение, мотивация, оценка, развитие кадров;
- **Обратные связи:** показатели эффективности, отзывы сотрудников, показатели развития;
- **Выходы:** кадровый баланс, показатели мотивации, уровень компетентности, показатели эффективности.

Взаимосвязи и взаимодействия

Модель отражает, как внутренние процессы взаимодействуют и влияют друг на дру-

га, а также как внешние факторы (экономическая ситуация, рыночные условия) воздействуют на систему в целом.

Используемые методы

- Диаграммы потоков данных (DFD);
- UML-диаграммы классов и сценариев;
- Моделирование динамики (системная динамика).

Такая структурированная модель позволяет анализировать и оценивать эффективность управления, выявлять узкие места, прогнозировать результативность изменений и оптимизаций.

Практическая апробация модели

- Проведение пилотных проектов на предприятиях различного профиля и размеров;
- Анализ эффективности внедрения системных методов;
- Оценка качественных и количественных показателей (скорости реагирования, уровня удовлетворенности работников, затрат на управление).

Результаты показали, что системное моделирование способствует снижению издержек, снижению времени на кадровые процедуры, повышению мотивации и производительности.

Выводы

Основные выводы:

- Использование системного анализа в управлении кадровым отделом позволяет выявлять целостную картину процессов, взаимодействий и зависимостей;
- Разработанная модель способствует повышению прозрачности, повышению эффективности и гибкости кадрового управления;
- Интеграция автоматизированных информационных систем с системным моделированием позволяет автоматизировать оценку и прогнозирование кадровых ресурсов.

Перспективы дальнейших исследований:

- Расширение модели на многоорганизационные системы и сети предприятий;
- Использование методов машинного обучения для прогнозирования кадровых потребностей;
- Разработка программных платформ для автоматического построения и анализа системных моделей в реальном времени;
- Внедрение системного анализа в стратегическое управление человеческими ресурсами. Обобщая результаты исследования, можно отметить, что системный анализ предоставляет мощный инструмент для моделирования и оптимизации системы управления отделом кадров. Внедрение предложенных моделей и методов позволит значительно повысить эффективность и устойчивость кадровых процессов, обеспечит организацию необходимым аналитическим инструментарием для принятия обоснованных управленческих решений и стратегического планирования.

Литература

1. Smith J. Models of HR Management Processes // Journal of Business Systems. – 2020.

2. Ivanov V. System Approach to Human Resource Management // HR Management Review. – 2021.

3.Johnson P. Automating HR Processes // International Journal of Systems Engineering. – 2019.

4. Kuznetsov A. Methodologies for Systemic HR Analysis // Management Science Journal. – 2022.

5. Иванов В. Введение системного подхода в управление персоналом // Управление бизнес-процессами, 2021.

6. Петрова Е. Управление человеческими ресурсами в условиях цифровой трансфор-

мации // Современные тенденции в HR, 2022.

Komarnitsky V.R. Research of the enterprise personnel department management system based on system analysis*. A brief comparative analysis of the proposed models and methods has been performed, which will significantly increase the efficiency and sustainability of personnel processes, and provide the organization with the necessary analytical tools for making informed management decisions and strategic planning.*
Keywords: *system analysis; system modeling; automation of HR processes; personnel efficiency.*

УДК 004.85

ПРИМЕНЕНИЕ МЕТОДОВ ОБУЧЕНИЯ НЕЙРОННЫХ СЕТЕЙ В ДИСКРЕТНЫХ СИМУЛЯЦИОННЫХ СРЕДАХ

Кукуруза Я.А., Тарабаева И.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: kukuruza2005@vk.com, inkatar1@yandex.ru

Кукуруза Я.А., Тарабаева И.В. Применение методов обучения нейронных сетей в дискретных симуляционных средах. *В работе рассматривается применение методов обучения нейронных сетей в дискретных симуляционных средах открытого типа. Основное внимание уделено разработке и тестированию алгоритмов автономных интеллектуальных агентов в моделируемой среде. Рассматриваются подходы к обучению, анализу и оценке поведения агентов, а также возможности масштабирования и эмерджентного взаимодействия. Результаты работы могут быть использованы для оптимизации программного обеспечения интеллектуальных систем и разработки экспериментальных платформ для обучения ИИ.*
Ключевые слова: *нейронные сети, искусственный интеллект, автономные агенты, симуляционные среды, эмерджентное поведение.*

Введение

Нейронные сети появились в 1940-х годах. Идея создания таких систем возникла при изучении процессов, протекающих в мозге, и попытке смоделировать эти процессы. В 1943 году Уолтер Питтс и Уоррен МакКаллок предложили первую математическую модель нейрона, которая описывала, как нейроны могут обрабатывать информацию. В 1950-х годах Фрэнк Розенблатт разработал персептрон — одну из первых моделей искусственной нейронной сети, заложив тем самым основу для современных методов машинного обучения.

Со временем алгоритмы обучения стали значительно сложнее, что позволило применять их не только к статическим данным, но и к динамическим системам, в том числе к дискретным симуляционным средам. Такие среды моделируют поведение сложных систем, где состояние изменяется шаг за шагом, а агенты принимают решения в зависимости от текущей ситуации. Использование нейросетевых подходов в подобных моделях помогает более точно прогнозировать изменения, автоматизировать принятие решений и оптимизировать процесс исследования.

На сегодняшний день методы обучения нейронных сетей активно применяются в имитационных моделях для решения задач оптимизации, планирования и адаптивного управления, что делает их ключевым инструментом для анализа сложных процессов и разработки интеллектуальных систем.

Цель данной работы — разработать и изучить алгоритм обучения агента в дискретной симуляционной среде на основе методов обучения с подкреплением (Q-learning).

1 Постановка задачи

Среда представляет собой двумерное поле фиксированного размера, состоящее из дискретных клеток. Агент перемещается по этому полю, выполняя последовательность действий и получая численное вознаграждение в зависимости от результата каждого шага.

Необходимо:

- определить структуру среды и правила взаимодействия агента с окружением;
- задать целевую позицию агента и препятствия;
- разработать алгоритм обучения, основанный на концепции методов обучения с подкреплением (Q-learning);

- провести серию экспериментов по обучению агента в среде;
- проанализировать динамику обучения и поведение агента после сходимости.

Ожидаемый результат — агент формирует стратегию, позволяющую минимизировать количество шагов до цели и избегать нежелательных состояний. Полученные результаты могут использоваться как базовый демонстрационный пример применения методов обучения в дискретных имитационных моделях.

2 Особенности использования нейронных сетей в задачах обучения с подкреплением

При работе с нейронными сетями традиционно выделяют две основные парадигмы обучения: обучение с учителем и обучение без учителя. Однако обучение с подкреплением не относится напрямую ни к одной из них. В нём агент обучается не по заранее подготовленным данным, а через взаимодействие со средой, получая вознаграждение за свои действия. Использование многослойного перцептрона в задачах обучения с подкреплением, в частности в алгоритме Q-learning, существенно отличается от классического подхода в задачах аппроксимации функций. Можно выделить два ключевых отличия:

- Отсутствие фиксированного обучающего множества.

В обычных задачах аппроксимации обучение ведётся на заранее известной выборке, элементы которой многократно повторяются. В Q-learning такого множества нет — обучающие данные формируются динамически во время взаимодействия агента со средой. В результате некоторые состояния и действия могут встречаться часто, а другие — крайне редко. В непрерывных средах одно и то же состояние может вообще не повториться. Поэтому проблема переобучения здесь практически не проявляется.

- Отсутствие заранее известных целевых значений. В классических задачах известны истинные значения функции, которые сеть должна аппроксимировать. В Q-learning агент не знает заранее оптимальных Q-значений — они оцениваются и корректируются постепенно, исходя из опыта. По мере накопления опыта оценки функции ценности $Q(x_t,a_t)$ уточняются, приближаясь к оптимальным значениям.

3 Разработка дискретной имитационной среды

Дискретная имитационная среда — это формализованная модель системы, в которой состояние, перемещения и события представлены в виде дискретных шагов и конечного множества возможных состояний. В такой среде изменение конфигурации происходит по шагам, согласно определённым правилам, и каждое действие приводит к переходу из одного дискретного состояния в другое.

Агент — это автономная обучаемая сущность, принимающая решения на основе взаимодействия со средой. Он наблюдает текущее состояние, выбирает действие согласно стратегии, получает вознаграждение или штраф и обновляет свою оценочную модель, стремясь максимизировать совокупную награду во времени.

Цель — это заранее определённое конечное состояние системы, достижение которого считается успешным результатом работы агента. В контексте имитационной среды цель задаёт критерий завершения эпизода и источник положительного вознаграждения, направляя процесс обучения и формируя предпочтительное поведение.

В качестве среды обучения был выбран лабиринт, который будет случайно генерироваться. Размер — 16 на 16 условных единиц. Агент — 1 на 1 условных единиц синего цвета.

Для визуализации дискретной имитационной (см. рис. 1) среды хорошо подходит интерфейс программирования приложений – Windows Forms, этот инструмент позволяет создавать полнофункциональные настольные приложения и будет служить для отображения поведения агента в процессе его обучения и демонстрации работы сформированной стратегии после завершения обучения [1].

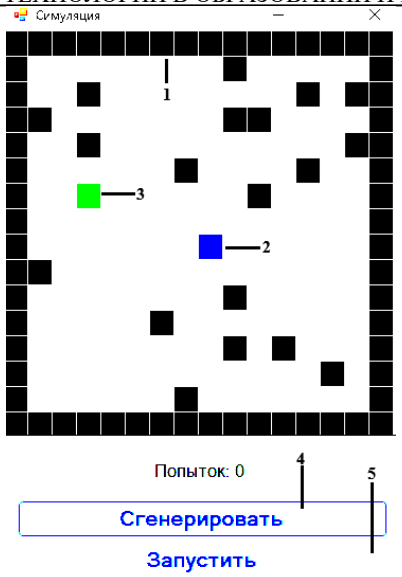

Рисунок 1 – Интерфейс разработанной среды.

На рисунке 1 показан разработанный интерфейс дискретной симуляционной среды. Цифрами помечены элементы интерфейса, где:

1 – Граница для обучения агента.
2 – Агент.
3 – Цель.
4 – Кнопка генерации новой среды.
5 – Запуск обучения агента.

4 Разработка класса агента.

Q-learning — это алгоритм обучения с подкреплением, который позволяет агенту оптимизировать свою стратегию действий в динамичной среде, стремясь максимизировать сумму будущих наград. Агент исследует среду и выбирает действия на основе данных из Q-таблицы, в которой изначально все значения равны нулю, и таким образом на каждом шаге обновляет Q-значения, используя опыт взаимодействия с окружающей средой, чтобы оптимизировать будущие действия [2].

Алгоритм обучения агента представляет собой Q-функцию (1), аргументом которой является не только состояние, но также и действие. Это позволяет итерационным способом построить Q-функцию и тем самым найти оптимальную политику управления.

$$Q(x_t, a_t) \leftarrow r_t + \gamma \cdot V(X_{t+1}) \quad (1)$$

где a_t — действие, выбранное в момент времени t из множества всех возможных действий A. Так как целью системы является максимизация суммарной награды, $V(X_{t+1})$ заменяется на $\max_{a \in A} Q(x_{t+1}, a)$ и в результате получается выражение (2).

$$Q(x_t, a_t) \leftarrow r_t + \gamma \cdot \max_{a \in A} Q(x_{t+1}, a) \quad (2)$$

Оценки Q-значений хранятся в 2-х мерной таблице, входами которой являются состояние и действие. При табличном представлении Q-функции и Марковской среде имеется доказательство сходимости алгоритма Q-Learning.

Задачи агента

1. Выбирать действия ("TOP", "DOWN", "LEFT", "RIGHT") с учетом ε-жадной стратегии, балансируя между исследованием (случайный выбор с вероятностью ε) и эксплуатацией (выбор действия с максимальным Q-значением).
2. Обновлять Q-таблицу с использованием Q-learning (alpha=0.2, gamma=0.95) на основе полученной награды и максимального Q-значения следующего состояния.

3. Постепенно уменьшать ε (от 1.0 до 0.05 с затуханием 0.9995) для перехода от исследования к эксплуатации.

4. Максимизировать накопленную награду в условиях случайной среды, адаптируясь к новым состояниям (x, y).

Класс Agent.cs содержит следующие основные методы:

- Update – самый главный метод, служащий для обновления Q-таблицы с использованием формулы 2. Смысл метода заключается в том, чтобы скорректировать ожидаемую ценность выбранного действия с учётом полученного вознаграждения и оценки будущих состояний. Если текущее состояние является терминальным (например, агент достиг цели), дальнейшие состояния не рассматриваются, и обновление Q-значения происходит только на основе полученного вознаграждения.
- Каждый шаг в классе Form1.cs в методе Step вызывается NextStep, это еще один важный метод, который отвечает за выбор действия агентом в текущем состоянии. Он реализует стратегию ε-жадности, которая позволяет сбалансировать исследование среды и использование уже полученных знаний. Если случайное число меньше epsilon, агент выбирает действие случайно — это фаза исследования, когда система пробует новые варианты для накопления опыта. В противном случае выбирается действие с максимальным значением Q — фаза эксплуатации, когда агент действует на основе уже изученной информации. Если несколько действий имеют одинаковое Q-значение, одно из них выбирается случайным образом, чтобы избежать детерминизма.
- DecayEpsilon управляет постепенным снижением вероятности случайного выбора действий. Он уменьшает значение epsilon после каждого эпизода или шага, тем самым сдвигая поведение агента от случайных проб к более уверенным решениям на основе накопленного опыта. Параметр eMin задаёт нижний предел, чтобы агент не перестал исследовать среду совсем. Постепенное уменьшение ε позволяет агенту переходить от исследования к эксплуатации, то есть всё реже пробовать случайные действия и всё чаще выбирать оптимальные решения на основе своих знаний.

Методы Update, NextStep и DecayEpsilon образуют ядро алгоритма Q-learning. Update корректирует Q-значения, приближая их к оптимальным. NextStep определяет, какое действие совершит агент — случайное или наилучшее. DecayEpsilon регулирует баланс между исследованием и использованием полученного опыта. В совокупности они обеспечивают способность агента обучаться, адаптироваться и постепенно формировать стратегию поведения, стремясь к максимальному вознаграждению.

Выводы

В результате проделанной работы были выполнены все поставленные задачи. Итогом стало создание нестандартного подхода к обучению нейросетевой модели в имитационной среде с использованием платформы Windows Forms. Выбор данной среды обусловлен широким применением языка программирования C# в игровой индустрии и его удобством для визуализации поведения агента. Разработанный алгоритм обучения можно рассматривать как альтернативу традиционным методам поиска пути, поскольку он формирует поведение на основе опыта, а не заранее заданных правил.

Безусловно, проект имеет потенциал для дальнейшего развития. В будущем возможно расширение функциональности имитационной среды, внедрение более сложных сценариев взаимодействия агента с окружением, а также применение глубоких нейронных сетей для повышения устойчивости и точности обучения. Такие улучшения позволят сделать процесс принятия решений агентом более адаптивным и приближённым к реальным условиям.

Полученные результаты подтверждают, что алгоритмы обучения с подкреплением способны эффективно моделировать адаптивное поведение агентов в имитационных средах, демонстрируя, как система может постепенно накапливать опыт, корректировать свои действия и принимать решения, ориентируясь на долгосрочную награду. Кроме того, реализо-

ванный подход показывает потенциал для применения в более сложных сценариях, где традиционные методы планирования и поиска пути оказываются недостаточно гибкими. Такой механизм обучения открывает возможности для экспериментирования с динамическими и изменяющимися условиями среды, интеграции разнообразных стратегий поведения и анализа поведения агента в реальном времени. В перспективе это позволяет не только расширять функциональность существующих систем, но и создавать новые инструменты для моделирования автономных агентов, способных адаптироваться к широкому спектру задач, от игровых до исследовательских, включая оптимизацию процессов, автономное принятие решений и прогнозирование поведения в сложных ситуациях.

Литература

1. Microsoft Learn: Windows Forms для .NET [Электронный ресурс]. – Режим доступа: https://learn.microsoft.com/ru-ru/dotnet/desktop/winforms/overview/ – Загл. с экрана.

2. Хабр: Машинное обучение [Электронный ресурс]. – Режим доступа: https://habr.com/ru/hubs/machine_learning/articles/ – Загл. с экрана.

3. Кузьмин Валерий Использование нейронных сетей в алгоритме Q-LEARNING / Институт транспорта и связи Ломоносова 1, Рига, LV 1019, Латвия.

Kukuruza Y.A., Tarabaeva I.V. Application of artificial intelligence training methods in discrete simulation environments. The paper considers the application of neural network training methods in discrete simulation environments of an open type. The main focus is on the development and testing of algorithms for autonomous intelligent agents in a simulated environment. Approaches to learning, analyzing, and evaluating agent behavior, as well as the possibilities of scaling and emergent interaction, are considered. The results of the work can be used to optimize the software of intelligent systems and develop experimental platforms for AI training.

Keywords: neural networks, artificial intelligence, autonomous agents, simulation environments, emergent behavior.

УДК 004.89

НЕЙРОСЕТЕВОЙ ИНТЕЛЛЕКТ ДЛЯ ИГРОВЫХ ПЕРСОНАЖЕЙ

Куштей И.В., Анохина И.Ю.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: ikushtey@mail.ru, ingatula@mail.ru

Куштей И.В., Анохина И.Ю. Нейросетевой интеллект для игровых персонажей. *Статья посвящена задаче интеграции нейросетевого интеллекта для управления игровыми персонажами в 2D-платформерах. Проведён обзор современных подходов к управлению игровыми персонажами. Описана целевая архитектура эксперимента. Определены требования к среде обучения и предложены инструменты для прототипирования и мониторинга. Сформулирован протокол обучения с этапами прототипирования. Установлены метрики оценки и проведён план сравнительного анализа с human baseline и скриптами. Выявлены ключевые риски и предложены меры контроля и направления дальнейшей работы*
Ключевые слова: *нейросети, обучение, 2D платформер, искусственный интеллект, адаптивность.*

Введение

Современные технологии искусственного интеллекта (ИИ) всё активнее внедряются в индустрию видеоигр, расширяя возможности взаимодействия с виртуальной средой и персонажами. Одним из перспективных направлений является интеграция нейросетевых моделей для управления игровыми агентами, особенно в жанре 2D платформеров. Такие модели способны заменить традиционное управление, основанное на скриптах и ручном вводе, на автономное поведение, формируемое в процессе обучения [1-2].

Цель работы – исследовать возможность замены классического ручного управления персонажем в 2D платформере на управление, полностью выполняемое нейросетевым агентом, и оценить применимость современных методов обучения с подкреплением и имитационного обучения для решения практических задач геймплея: прохождение уровня, избегание препятствий, взаимодействие с объектами и демонстрация «играбельного» поведения для пользователей. Такие исследования актуальны в игровой индустрии как для автоматизированного тестирования и генерации контента, так и для новых игровых механик и ассистентного геймплея [1,3].

Выбор подходов и архитектур

Решение целесообразно строить вокруг нескольких направлений: обучение с подкреплением (PPO, SAC и другие алгоритмы) для прямой оптимизации поведения в среде; имитационное обучение и поведенческое клонирование для быстрого и безопасного старта; гибридные и иерархические схемы, которые объединяют нейросеть с набором высокоуровневых команд и правил для сохранения контролируемости и предсказуемости поведения агента [3-4].

Обучение с подкреплением —тип машинного обучения, в котором искусственный интеллект учится принимать решения путем взаимодействия со средой. Его можно рассматривать как процесс проб и ошибок, причем агент получает «награды» или «штрафы». В отличие от обучения с учителем, в этом случае есть набор готовых правильных ответов, при обучении с подкреплением агент не знает, какое именно действие правильное. Он знает только, насколько оно было хорошим или плохим. Анализируя результаты, агент должен найти оптимальную стратегию поведения.

Как пример, рекомендательные системы, которые моделируют предпочтения пользователя и предлагают ему наиболее релевантные товары, контент или услуги, это агент, который получает награду, когда пользователь посмотрел рекомендованный контент на YouTube или открыл страницу с товаром на маркетплейсе. Они ищут ответ на вопрос: «Что этому пользователю может понравиться?»

Среда обучения и инструментарий

Эксперименты целесообразно проводить в модульной симуляции уровня с реплицируемой физикой и настраиваемой стохастичностью; в качестве платформ для прототипирования подходят движки Unity или Godot, связанные с RL-фреймворками (Stable Baselines3, RLlib) и системами мониторинга (TensorBoard, Weights & Biases), что упрощает трассировку обучения и интеграцию результатов в игровую сборку [2,4].

Постановка задачи, представление состояния и пространство действий

Задача формулируется как обучение автономного агента, который в условиях 2D платформера будет последовательно принимать решения по перемещению и взаимодействию с объектами уровня так, чтобы максимально безопасно и эффективно достигать контрольных точек и финиша.

В качестве целевой объективной функции рассматривается сочетание показателей прогресса по уровню: положительная награда за продвижение по оси X и за достижение чекпоинтов (контрольных точек, которые фиксируют текущий прогресс игрока) и качественных ограничений поведения (штрафы за падение, смерть, бесцельное стояние и за стратегии, эксплуатирующие баги симуляции). Отметим, что любая игра — это симуляция, управляемая кодом. А код, как известно, не может быть идеален и в нем всегда будут ошибки (баги). Игроки обнаруживают эти места и строят на этом стратегии. Например, игрок нашел место, в котором он может стоять, и его «не увидят». Из этого вытекает стратегия: всегда занимать эту позицию для последующей атаки.

Представление задачи требует явного определения наблюдаемого состояния, пространства действий и частотного режима принятия решений: состояние может включать семантический вектор (позиция героя, скорости, ближайшие платформы, статусы врагов и ближайшие препятствия) и/или визуальный вход (часть экрана вокруг персонажа), а действие – дискретный набор базовых команд (влево, вправо, прыжок, спецприём) или параметризованный контроллер (сила/длительность прыжка, плавное управление по оси). Важна спецификация ограничений наблюдаемости – полная (вся сцена доступна) или частичная (только локальная видимость), поскольку это определяет необходимость использования рекуррентных блоков или памяти в модели.

Для практической реализации следует задать формулировку процесса обучения: эпизоды начинаются из контролируемых стартовых состояний, длительность эпизодов ограничена, а при достижении финиша или смерти эпизод завершается. В процессе обучения вводится curriculum – серия упрощённых версий уровня, увеличивающаяся по сложности, чтобы обеспечить стабильную сходимость. Curriculum Learning – стратегия обучения, при которой модель сначала тренируется на простых задачах, а затем постепенно переходит ко всё более сложным. Например, на первом этапе необходимо распознать человека на контрастном фоне в четкой позиции, например, стоя. После успешного обучения на простых задачах, вводят более сложные примеры: человек частично скрыт и находится в одной из поз йоги.

Наградную структуру проектируют с учётом избегания локальных оптимумов: основной компонент награды за прогресс дополняют shaping-наградами за достижение промежуточных целей, маленькими положительными сигналами за обнаружение полезных объектов и отрицательными за очевидно «читерское» поведение (например, многократное использование нестабильных столкновений). Создавая систему наград, целесообразно исключить малоэффективные тактики. Найденный локальный оптимум — это хорошее, но

не лучшее решение. Глобальный оптимум – это наилучшее возможное решение. Задача – заставить искать глобальный оптимум, а не довольствоваться первым попавшимся локальным.

Для валидации устойчивости агента определяют набор контрольных сценариев и метрик: частоту успешных прохождений на валидационных уровнях, среднее время прохождения, число смертей на 100 эпизодов, распределение траекторий действий и меры «человечности» поведения (сравнение с демо-траекториями или с метриками гладкости и своевременности действий) [1-3].

Процесс обучения и оценка результатов

Процесс обучения включает прототипирование на простых уровнях, curriculum learning с постепенным усложнением, использование демонстраций для инициализации поведения и последующий fine-tuning через RL; оценка проводится сравнением с human baseline и скриптами по метрикам: доля успешных прохождений, среднее время до финиша, количество столкновений и субъективная оценка «человечности» поведения игроками и дизайнерами уровня [2-3].

Оценка с human baseline означает, что действия, результаты сравниваются с теми результатами, которые демонстрируют люди-эксперты в аналогичной задаче. Цель искусственного интеллекта – достичь, а в последствии превзойти этот уровень.

Практические риски и меры контроля

Основные риски связаны с тем, что агент может выучить нежелательные стратегии, плохо обобщать на новые уровни и требовать чрезмерных вычислительных ресурсов.

Во-первых, риск эксплойтов: агент находит лазейки в физике или логике игры (зацикливание в границе уровня, использование багов коллизий) и достигает целей, не выполняя ожидаемых игровых действий.

Мера контроля – вводить в наградную функцию штрафы за дискретные аномалии (статистика повторяющихся циклов, слишком частые телепортации, резкие скачки позиции) и реализовать post-filtering политик: отбраковывать и дополнительно тестировать политики, которые демонстрируют необычные траектории на валидационных сценариях. Post-filtering политик —стратегия, при которой окончательное решение принимается путем дополнительной обработки или «фильтрации» выходных данных уже обученной нейронной сети. Можно сказать, что это «умный доводчик», который корректирует сырые предложения модели, прежде чем они станут реальным действием.

Во-вторых, риск непредсказуемости в реальной сборке: поведение, стабильно работающее в симуляции, может рушиться из-за разницы в физике или детерминированности. Меры: domain randomization при обучении (вариация силы гравитации, трения, размеров платформ, позиции врагов), тестирование на грязных входах и интеграция площадки для A/B тестов в реальной сборке.

В-третьих, вычислительные и временные затраты: длительное обучение и необходимость множества запусков для репликации результатов. Меры контроля включают использование демонстраций для pretraining (поведенческое клонирование) для быстрого старта, уменьшение размера модели и применение прогрессивного усложнения уровней (curriculum) для сокращения времени до приемлемого поведения, а также профилирование и оптимизацию среды симуляции.

Для сохранения контроля над эстетикой и предсказуемостью поведения вводят архитектурные и инженерные меры: иерархические агенты, где высокоуровневый планировщик выбирает целевые точки, а низкоуровневый контроллер обеспечивает безопасное исполнение, и safety layer, накладывающий жесткие ограничения на выходные действия (например, запрет на переход через препятствия без выполнения условий).

Параллельно следует наладить процедуру мониторинга и валидации: автоматический сбор метрик и видеозаписей эпизодов, ручной ревью критических сессий, регресс-тесты

после изменения симуляции и пороговые триггеры для отката модели при выявлении ухудшения показателей.

Наконец, необходимо предусмотреть этапы тестирования с участием людей: краш-тесты дизайнерами уровня, оценка «человечности» поведения игроками и замеры восприятия игры в смешанных режимах player+agent, чтобы убедиться, что внедрение искусственного интеллекта не нарушает игровой опыт [3-5].

Выводы

Замена традиционного управления ИИ-агентом в 2D платформере технически осуществима и приносит практические преимущества в тестировании, ассистировании игрокам и создании новых механик, однако требует внимательного проектирования представления состояния, системы награждений, гибридных архитектур для контроля поведения и практических мер по обеспечению стабильности и предсказуемости. Существующие отраслевые примеры и публикации подтверждают растущий интерес и применимость нейросетевых методов в игровой разработке, но оперативная интеграция требует балансирования между качеством поведения агента и затратами на обучение.

Литература

1. Нейросеть для создания персонажа 2D игры / Хабр [Электронный ресурс]. – Режим доступа: https://habr.com/ru/articles/844372/. . – Загл. с экрана.
2. Нейросеть для создания 2d персонажей: новые возможности и творческие решения / Rating Gamedev [Электронный ресурс]. – Режим доступа: https://rating-gamedev.ru/blog/neiroset-personazi-2d-igry. . – Загл. с экрана.
3. Использование нейросетей в компьютерных играх / CyberLeninka [Электронный ресурс]. – Режим доступа: https://cyberleninka.ru/article/n/ispolzovanie-neyrosetey-v-kompyuternyh-igrah. . – Загл. с экрана.
4. Использование нейросетей в игровой индустрии / TechHubSPB [Электронный ресурс]. – Режим доступа: https://techhubspb.ru/innovacionnye-nejroseti/ispolzovanie-nejrosetej-v-igrovoj-industrii-sozdanie-i-upravlenie-personazhami/. . – Загл. с экрана.
5. Нейросети для 2D/3D-графики и инструменты разработчика / Digital-Razor [Электронный ресурс]. – Режим доступа: https://digital-razor.ru/media/articles/workstation/ai-2d-3d-graphics/. . – Загл. с экрана.

Kushtey I.V.., Anokhina I.Y. Neural network intelligence for game characters. *This article focuses on the task of integrating neural network intelligence for controlling game characters in 2D platform games. It provides an overview of current approaches to controlling game characters. The article describes the target architecture of the experiment. It defines the requirements for the training environment and suggests tools for prototyping and monitoring. The article formulates a training protocol with prototyping stages. It establishes evaluation metrics and provides a comparative analysis plan with human baseline and scripts. Key risks have been identified, and control measures and directions for further work have been proposed.*
Keywords*: neural networks, learning, 2D platformer, artificial intelligence, adaptability.*

УДК 004.8

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ЭФФЕКТИВНОСТИ РЕЗЮМИРОВАНИЯ НОВОСТНЫХ СТАТЕЙ ПРИ ПОМОЩИ ОБЩЕДОСТУПНЫХ LLM

Лазуренко Е. А., Коломойцева И. А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия»
e-mail: eugenelazurenko@gmail.com, bolatiger@mail.ru

Лазуренко Е. А., Коломойцева И. А. Сравнительный анализ эффективности резюмирования новостных статей при помощи общедоступных LLM. *Представлен анализ работы четырёх общедоступных больших языковых моделей в задаче автоматического резюмирования новостных текстов на русском языке. Каждая модель генерировала краткие версии всех текстов, после чего проводился анализ степени сокращения, времени обработки и качества резюме с использованием метрик BERTScore и ROUGE.*

Ключевые слова: *большие языковые модели, LLM, новостные статьи, BERTScore, ROUGE.*

Введение

С появлением больших языковых моделей (LLM, Large Language Models) возможностей автоматического резюмирования текста значительно выросли. Особенно актуальным этот подход становится для новостных статей — информационных, часто многословных, но при этом требующих быстрых обзоров. Технологии резюмирования, основанные на LLM, предлагают гибкость, обобщение и адаптивность, которые трудно достичь с классическими методами. Однако эффективность таких моделей варьируется в зависимости от архитектуры, размера модели, токенов ввода и качества инструкций. В этой статье проводится сравнительный анализ четырёх общедоступных LLM в задаче суммаризации новостных текстов: мы измеряем, как сильно модели сокращают текст, сколько времени это занимает и насколько абстракции остаются семантически и лексически адекватными исходным статьям.

1 Постановка задачи

Цель нашего исследования — оценить эффективность и качество автоматического резюмирования новостных статей с использованием четырёх популярных общедоступных больших языковых моделей. Основная задача заключалась не только в измерении степени сокращения текста, но и в проверке сохранения смысловой и семантической точности итоговых резюме. Для этого был сформирован корпус из 32 новостных статей на русском языке, представляющих различные тематики и стили из популярных источников. Каждая статья была обработана всеми четырьмя моделями, при этом фиксировались время обработки, исходный и сокращённый размер текста. Полученные резюме затем оценивались с помощью автоматических метрик качества — BERTScore, отражающего семантическую близость текста, и ROUGE, измеряющего лексическое и структурное соответствие исходным статьям. Такой подход позволил не только количественно сопоставить модели по степени сжатия и скорости работы, но и качественно проанализировать их способность сохранять информативность и точность новостных материалов.

2 Начальные данные и используемые модели

В корпус вошли 32 статьи из популярных русскоязычных новостных источников.

Размеры исходных статей в символах:

- Минимальный размер: 500 символов
- Максимальный размер: 12682 символов

– Средний размер: 2275 символов

Для эксперимента были выбраны четыре популярные и общедоступные модели больших языковых моделей (LLM), доступные для задачи резюмирования: mistral:latest, qwen:4b, phi3:mini и gemma:2b. Основным критерием выбора был компактный размер моделей, не превышающий 5 ГБ, что обеспечивает их удобство для локального использования и сравнительно быструю обработку текстов.

В таблице 1 указаны краткое описание моделей.

Таблица 1. Описание языковых моделей

Модель	Размер модели, ГБ.	Примечения
mistral:latest	4.4 ГБ	Современная компактная LLM семейства Mistral, хорошо оптимизирована под обобщение и сложные инструкции
qwen:4b	2.3 ГБ	Модель от Alibaba, специализирующаяся на многоязычных задачах
phi3:mini	2.2 ГБ	Легковесная модель Microsoft семейства Phi, ориентированная на экономию ресурсов
gemma:2b	1.7 ГБ	Миниатюрная модель Google, адаптированная под широкие бытовые задачи

3 Результаты сокращения текста

Были измерены, насколько эффективно каждая модель умеет сжимать текст, чтобы понять, как разные алгоритмы справляются с задачей сокращения новостных статей. Для каждой из 32 статей мы сопоставляли длину исходного текста и итогового резюме.

Параллельно с этим фиксировалось время обработки каждой статьи каждой моделью, поскольку скорость генерации напрямую влияет на практическую применимость системы.

Быстрые модели удобны для потоковой обработки больших массивов данных, однако они не всегда дают самый точный или информативный результат. Медленные модели, наоборот, часто отличаются более тщательным подходом, но создают задержки при работе с большими наборами текстов.

В таблице 2 приведены результаты генерации резюме всех текстов статей для каждой модели.

Таблица 2. Результаты генерации сокращенных текстов LLM

Модель	Мин. коэффициент сжатия (%)	Макс. коэффициент сжатия (%)	Средний коэффициент	Среднее время на статью (с)
mistral:latest	0.26 %	100.62 %	35.22 %	4.41 с
qwen:4b	2.02 %	87.21 %	21.62 %	3.33 с
phi3:mini	3.18 %	9961.38 %	611.89 %	24.94 с
gemma:2b	5.25 %	86.29 %	53.19 %	8.41 с

4 Оценка качества резюме: BERTScore и ROUGE

Модели оценки качества резюме позволяют количественно определить, насколько хорошо сгенерированный текст передаёт содержание оригинала. Они применяются тогда, когда требуется проверить не только формальные характеристики текста, но и его смысловую полноту, точность и структурное соответствие исходным данным. Такие модели особенно полезны при сравнении различных алгоритмов генерации, поскольку дают объективные числовые показатели, которые можно сопоставлять между собой. Благодаря этому становится возможным не просто субъективно оценивать качество резюме, а получать измеримые данные, подтверждающие или опровергающие эффективность конкретного подхода.

BERTScore опирается на контекстные представления слов, создаваемые трансформерными моделями. В отличие от метрик, основанных на подсчёте совпадающих слов, он оценивает семантическую близость между исходным текстом и сгенерированным резюме. Это значит, что метрика фиксирует не только факт совпадения отдельных лексем, но и то, насколько похожи по смыслу выражения, даже если они сформулированы по-разному. Такой подход особенно важен при оценке абстрактивных резюме, где модель может передавать содержание другими словами, сохраняя общую мысль.

ROUGE, напротив, отражает степень лексического пересечения между текстами. Эта метрика измеряет, насколько часто в резюме встречаются те же последовательности слов или символов, что и в исходном тексте. Благодаря этому ROUGE хорошо показывает, насколько полно модель сохраняет фактические элементы исходного материала. Она особенно эффективна для оценки экстрактивных резюме, где важно проверить точность воспроизведения оригинальных фраз и словосочетаний. Различные варианты метрики — например, ROUGE-1, ROUGE-2 и ROUGE-L — позволяют анализировать совпадения на уровне отдельных слов, пар слов или длинных последовательностей, что делает оценку более гибкой и информативной.

В таблице 3 представлены коэффициенты оценки метрик BERTScore и ROUGE.

Таблица 3. Результаты оценки метрик BERTScore и ROUGE

Модель	BERTScore (F1) среднее	ROUGE-1 среднее	ROUGE-2 среднее	ROUGE-L среднее
mistral:latest	0.9051	0.6507	0.4459	0.6349
qwen:4b	0.8620	0.2959	0.2136	0.2924
phi3:mini	0.7560	0.0492	0.0042	0.0449
gemma:2b	0.8926	0.2530	0.0912	0.2447

5 Анализ результатов

В рамках анализа результатов сразу заметно, что модели продемонстрировали различный уровень устойчивости и качества при решении задачи резюмирования. Наиболее сбалансированно повела себя mistral:latest: она сокращала тексты ровно настолько, чтобы сохранить их структуру и смысл, избегала резких выбросов по коэффициенту сжатия и показала самые высокие значения BERTScore и ROUGE. Итоговые резюме получались информативными, а время обработки оставалось умеренным, что делает её наиболее надёжной моделью в эксперименте.

Поведение qwen:4b оказалось более агрессивным в плане сокращения. Она работала быстрее всех, формировала самые компактные резюме, но делала это в ущерб содержательности. Падение ROUGE показывает, что модель склонна опускать значимые фрагменты исходного текста, оставляя лишь общую смысловую канву. Тем не менее её результаты остаются приемлемыми там, где важны скорость и сильное уменьшение объёма.

Полной противоположностью стала phi3:mini, которая в ряде случаев не сокращала текст, а чрезмерно его увеличивала, что отражено в экстремальных значениях коэффициента сжатия. Низкие показатели и BERTScore, и ROUGE подтверждают, что модель не справилась с задачей: итоговые тексты лишь слабо соотносятся с оригиналами и плохо передают их содержание.

На фоне этих контрастов gemma:2b выглядит умеренно стабильной, но недостаточно точной. Она не допускает крайних отклонений и поддерживает предсказуемый уровень сокращения, однако качество резюме уступает более сильным моделям: в них заметно больше обобщений и меньше фактических элементов исходных статей.

Таким образом, результаты показывают, что единственной моделью, уверенно и последовательно справившейся с задачей резюмирования новостей, оказалась mistral:latest, тогда как остальные решения демонстрировали либо недостаток точности, либо низкую устойчивость.

Выводы

Результаты эксперимента показывают, что модели существенно различаются по качеству и устойчивости резюмирования. Наиболее точные и информативные сокращения дала mistral:latest, сохранившая смысл оригинальных статей и продемонстрировавшая лучшие значения BERTScore и ROUGE при стабильном уровне сжатия. Остальные модели справились заметно хуже: qwen:4b слишком сильно сокращала тексты и теряла значимые детали, gemma:2b создавала в основном обобщённые пересказы, а phi3:mini оказалась нестабильной и часто нарушала формат задания. Эти различия подчёркивают, что для надёжного резюмирования новостей требуется модель с предсказуемым поведением и достаточной семантической выразительностью, что в данном сравнении лучше всего обеспечила mistral:latest.

Литература

1. Chistyakova K., Kazakova T. Grammar in Language Models: BERT Study // National Research University Higher School of Economics. – 2023. – № 115.

2. Yadav A. K., Ranvijay R. S., Yadav R. S., Maurya A. K. Graph-based extractive text summarization based on single document // Multimedia Tools and Applications. – 2024. – Vol. 83, № 7. – P. 18987-19013. – DOI: 10.1007/s11042-023-16199-8.

3. Паршин И. А. Роль избыточности текста в суммаризации // Вестник науки. – 2025. – № 5 (86), т. 3. – URL: https://cyberleninka.ru/article/n/rol-izbytochnosti-teksta-v-summarizatsii

4. Ребенок К. В. Эффективность нейросетевых алгоритмов в автоматическом реферировании и суммаризации текста. – 2024. – URL: https://cyberleninka.ru/article/n/effektivnost-neyrosetevyh-algoritmov-v-avtomaticheskom-referirovanii-i-summarizatsii-teksta

5. Егунова А. И., Комаров Р. С., Вечканова Ю. С. и др. АНАЛИЗ АЛГОРИТМОВ И РЕШЕНИЙ ДЛЯ АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ ПОДВОДОК НОВОСТНЫХ СТАТЕЙ В СОЦСЕТЯХ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА // Вестник АГТУ. Серия: Управление, вычислительная техника и информатика. 2023. №1. URL: https://cyberleninka.ru/article/n/analiz-algoritmov-i-resheniy-dlya-avtomaticheskoy-generatsii-podvodok-novostnyh-statey-v-sotssetyah-s-ispolzovaniem-iskusstvennogo

6. Салып Богдан Юрьевич ПРОГРАММНАЯ СИСТЕМА АВТОМАТИЧЕСКОГО РЕФЕРИРОВАНИЯ ТЕКСТА // Проблемы Науки. 2023. №7 (185). URL: https://cyberleninka.ru/article/n/programmnaya-sistema-avtomaticheskogo-referirovaniya-teksta

Lazurenko E. A., Kolomoitseva I. A. Comparative Analysis of the Efficiency of News Article Summarization Using Publicly Available LLM. *This paper presents an analysis of the performance of four publicly available large language models in the task of automatic summarization of news texts in Russian. Each model generated short versions of all texts, after which an analysis was conducted of the degree of compression, processing time, and summary quality using the BERTScore and ROUGE metrics.*

Keywords: *large language models, LLM, news articles, BERTScore, ROUGE.*

УДК 004.89

АНАЛИЗ СОВРЕМЕННЫХ РЕШЕНИЙ ДЛЯ РАЗРАБОТКИ ВИРТУАЛЬНЫХ ПОМОЩНИКОВ С ИИ

Лободзинский В. О.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: rayli.ru@gmail.com

Лободзинский В.О. Анализ современных решений для разработки виртуальных помощников с ИИ. *Данное исследование посвящено анализу современных технологий и инструментов, применяемых при создании виртуальных помощников с элементами искусственного интеллекта. В статье рассмотрены ключевые платформы и программные решения, используемые для разработки интеллектуальных ассистентов, выделены их основные возможности и области применения.*
Ключевые слова: *виртуальный помощник, искусственный интеллект, интеллектуальные системы, технологии ИИ, анализ решений, платформы разработки.*

Введение

С развитием информационных технологий всё большее внимание уделяется созданию виртуальных помощников на основе искусственного интеллекта. Такие системы становятся неотъемлемой частью современной цифровой среды и применяются в самых разных сферах: от бытовых приложений и голосовых ассистентов в смартфонах до корпоративных решений, предназначенных для автоматизации служб поддержки, обработки клиентских запросов и повышения эффективности взаимодействия с пользователями.

Виртуальные помощники способны выполнять широкий спектр задач — от простых действий, таких как напоминания и поиск информации, до более сложных функций, связанных с анализом данных, прогнозированием и принятием решений. Благодаря использованию методов машинного обучения и обработки естественного языка (Natural Language Processing, NLP) они могут понимать контекст общения, адаптироваться к стилю пользователя и обеспечивать более персонализированное взаимодействие [1].

Несмотря на широкое распространение и значительный прогресс в области искусственного интеллекта, создание эффективных виртуальных помощников остаётся сложной междисциплинарной задачей. Это связано с необходимостью объединения нескольких ключевых направлений — анализа естественного языка, обучения на больших данных, распознавания речи, синтеза ответов и интеграции с внешними программными и аппаратными сервисами. Кроме того, важными проблемами остаются обеспечение конфиденциальности данных, повышение точности распознавания речи и улучшение способности системы к ведению длительных контекстных диалогов.

Цель данной работы — провести анализ современных технологий и платформ для разработки виртуальных помощников с ИИ, выявить их преимущества и ограничения, а также оценить возможности их применения.

1 Постановка задачи

Главная задача заключается в изучении и систематизации современных технологий и подходов к созданию виртуальных помощников с использованием искусственного интеллекта. Основной целью является выявление преимуществ и ограничений существующих решений, а также определение направлений, в которых возможно их дальнейшее улучшение.

Необходимо решить несколько конкретных задач: определить ключевые технологии, используемые при разработке виртуальных помощников, проанализировать их функцио-

нальные возможности и эффективность, сравнить существующие решения с точки зрения практического применения, выявить текущие проблемы и ограничения, с которыми сталкиваются разработчики.

2 Обзор современных решений

Современные виртуальные помощники с искусственным интеллектом разрабатываются с использованием разнообразных технологий. Одним из ключевых направлений является обработка естественного языка, которая позволяет системам понимать и формировать человеческую речь. Платформы на основе машинного обучения и нейронных сетей, такие как Google Dialogflow, Microsoft Bot Framework и Amazon Lex, предоставляют возможности для создания диалогов, распознавания голосовых команд и интеграции с различными внешними сервисами.

Другим важным аспектом является голосовой интерфейс. Технологии распознавания и синтеза речи позволяют сделать взаимодействие с пользователем более естественным и удобным. Использование готовых API, таких как Google Speech-to-Text или Amazon Polly, упрощает разработку и повышает точность распознавания команд [2].

Интеграция с внешними системами, такими как календари, электронная почта, CRM или социальные сети, делает виртуальных помощников более полезными и практичными в повседневной работе и бизнес-процессах.

Несмотря на достижения, современные решения имеют ограничения. Среди них — ограниченная способность к контекстному пониманию длительных диалогов, зависимость от качества обучающих данных, сложность настройки и адаптации системы под конкретные задачи пользователя.

Представлена следующая таблица программных решений и их сравнения [3] (табл. 1).

Таблица 1. Сравнительный анализ платформ.

Платформа	Основные возможности	Преимущества	Ограничения	Практическое применение
Google Dialogflow	Создание чат-ботов и голосовых помощников, NLP, интеграция с Google Assistant	Простота использования, поддержка мультиканальности, облачные сервисы	Ограниченные возможности для сложной логики, зависимость от облака	Бизнес-чатботы, голосовые помощники для веб и мобильных приложений
Microsoft Bot Framework	Разработка мультиканальных ботов, поддержка LUIS	Гибкость, интеграция с Microsoft 365 и Teams	Требует знаний программирования, сложность настройки	Корпоративные решения, внутренние ассистенты для компаний
Amazon Lex	NLP, распознавание и синтез речи, интеграция с AWS	Высокая точность распознавания речи, интеграция с AWS, масштабируемость	Сложность для начинающих, зависимость от AWS	Голосовые интерфейсы, сервисы поддержки клиентов
Rasa	Открытый фреймворк для создания чат-ботов с NLP и ML	Полный контроль над данными, гибкость, возможность локального хостинга	Сложность настройки, требует знаний Python и ML	Настраиваемые корпоративные решения, боты с уникальной логикой

Платформа	Основные возможности	Преимущества	Ограничения	Практическое применение
IBM Watson Assistant	Создание чатботов с NLP, интеграция с аналитикой и другими сервисами IBM	Поддержка сложных сценариев, аналитика, надежность	Высокая стоимость, ограниченные интеграции с внешними системами	Крупные компании, сервисные центры, консультационные боты

Анализ представленных инструментов показывает, что каждая платформа обладает своими сильными сторонами и ограничениями. Облачные решения удобны для быстрого запуска и простых задач, гибкие фреймворки, такие как Rasa, подходят для сложных кастомных решений, а корпоративные платформы обеспечивают интеграцию с бизнес-процессами. Эти наблюдения создают основу для последующего раздела о выводах и перспективах дальнейших исследований.

3 Обоснование и анализ полученных результатов

Проведённый анализ современных платформ для разработки виртуальных помощников показывает, что выбор конкретного инструмента напрямую зависит от целей и задач проекта. Облачные сервисы, такие как Google Dialogflow и Amazon Lex, обеспечивают быстрый старт разработки, простую интеграцию с внешними сервисами и удобство использования. Их преимущество заключается в готовых инструментах для обработки естественного языка и голосовых команд, однако ограниченная гибкость и зависимость от облачной инфраструктуры снижают возможности для создания сложных и уникальных сценариев.

Гибкие решения с открытым исходным кодом, такие как Rasa, дают разработчику полный контроль над данными и возможность создавать индивидуальные сценарии и диалоги. Такой подход позволяет глубже адаптировать систему под конкретные потребности пользователя и организации, но требует высокой квалификации разработчика, знаний машинного обучения и дополнительного времени на настройку.

Корпоративные платформы, включая Microsoft Bot Framework и IBM Watson Assistant, демонстрируют высокую надёжность, поддержку сложных сценариев и интеграцию с бизнес-процессами. Они особенно полезны в крупных организациях, где виртуальный помощник должен работать с корпоративными системами, аналитикой и обеспечивать качественную поддержку пользователей. В то же время высокая стоимость и необходимость специализированных знаний ограничивают их доступность для небольших проектов.

Обобщая результаты анализа, можно выделить несколько ключевых закономерностей. Во-первых, нет универсального решения, подходящего для всех задач: облачные сервисы оптимальны для быстрого прототипирования и простых приложений, гибкие фреймворки — для кастомных решений, а корпоративные платформы — для масштабных проектов. Во-вторых, основными вызовами современных систем остаются ограниченное контекстное понимание диалогов, недостаточная персонализация и сложность интеграции с внешними системами без специализированных навыков.

Выводы

Проведённое исследование показало, что выбор платформы для разработки виртуальных помощников должен основываться на специфике проекта, его масштабе и доступных ресурсах. Облачные решения обеспечивают простоту и скорость разработки, однако ограничены в гибкости и возможности глубокой адаптации. Платформы с открытым исходным кодом, напротив, предоставляют высокий уровень кастомизации и контроль над данными, но требуют значительных технических компетенций. Корпоративные системы демонстрируют максимальную надёжность и интеграцию с бизнес-процессами, что делает их предпочти-

тельными для крупных организаций, хотя высокая стоимость и сложность внедрения сдерживают их использование в малых проектах.

Таким образом, ни одно из существующих решений не является универсальным. Эффективность применения конкретной платформы определяется балансом между функциональностью, гибкостью, стоимостью и уровнем требуемой технической подготовки.

Литература

1. Sky.pro [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://sky.pro/wiki/analytics/nlp-natural-language-processing--chto-eto-i-kak-rabotaet/ – Загл. с экрана

2. Geeksforgeeks [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://www.geeksforgeeks.org/artificial-intelligence/10-best-chatbot-development-platforms-for-conversational-ai-1/ – Загл. с экрана

3. Kommunicate [Электронный ресурс] / Интернет-ресурс. – Режим доступа: https://www.kommunicate.io/blog/chatbot-framework-platform/ – Загл. с экрана

Lobodzinsky V.O. Analysis of Modern Solutions for Developing AI-Based Virtual Assistants. *This study is dedicated to the analysis of modern technologies and tools used in the creation of virtual assistants with elements of artificial intelligence. The article examines key platforms and software solutions employed in the development of intelligent assistants, highlighting their main features and areas of application.*

Keywords: *virtual assistant, artificial intelligence, intelligent systems, AI technologies, solutions analysis, development platforms.*

УДК 004.85

ТИПЫ МАШИННОГО ОБУЧЕНИЯ

Макаренко Е.А., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: jammymanrock51@mail.ru

Макаренко Е.А., Лазебная Л.А. Типы машинного обучения: Сравнительный анализ и области эффективного применения. *В данном докладе проводиться системный анализ четырёх ключевых парадигм машинного обучения: с учителем, без учителя, с подкреплением и глубокого обучения.*
Ключевые слова: *машинное обучение, обучение с учителем, обучение с подкреплением, глубокое обучение, сравнительный анализ, искусственный интеллект.*

Введение

Современный этап технологического развития характеризуется взрывным ростом объемов данных и широким внедрением искусственного интеллекта. Машинное обучение стало ключевым инструментом для решения сложных задач в таких областях, как прогнозирование, распознавание образов и автоматизация процессов. Однако разнообразие методов МО создает практическую проблему выбора оптимального подхода для конкретной задачи.

Цель данного доклада – провести сравнительный анализ основных парадигм машинного обучения для формирования критериев их выбора при решении прикладных задач.

Задачи доклада:

1. Дать формализованное описание парадигм обучения с учителем, без учителя, с подкреплением и глубокого обучения.
2. Выделить ключевые преимущества и ограничения каждой парадигмы.
3. Сформулировать четкие рекомендации по их применению в зависимости от типа задачи и доступных данных.
4. Визуализировать принцип работы каждого метода с помощью унифицированной схемы.

1 Постановка задачи проблемы и ее связь с важными задачами

Бурное развитие цифровой экономики и повсеместная цифровизация привели к взрывному росту объемов данных. В этих условиях машинное обучение (МО) стало ключевым инструментом для извлечения знаний и автоматизации принятия решений. Однако разнообразие методов МО порождает фундаментальную проблему: **отсутствие систематизированного подхода к выбору оптимальной парадигмы обучения для конкретной задачи,** что ведет к неэффективному использованию вычислительных ресурсов и получению недостоверных результатов [1]. Данная проблема на прямую связана с решением важнейших практических задач, таких как:

1. Создание точных систем прогнозирования в финансах и логистике.
2. Разработка автономных систем управления (роботы, беспилотные автомобили).
3. Анализ больших данных для выявления скрытых закономерностей (биоинформатика, социология).

2 Анализ последних исследований и выделение нерешенных частей проблемы

Парадигмы МО активно исследуются в научной литературе. Так, Мерфи [1] дает фундаментальное описание моделей с учителем и без учителя, а Саттон и Барто [2] заложили теоретические основы обучения с подкреплением. Прорыв в области глубокого обучения,

основанный на использовании многослойных нейронных сетей, детально описан в работе Гудфеллоу и др. [3].

Анализ публикаций [1, 2, 3, 4] показывает, что каждая из парадигм глубоко изучена в отдельности. Однако **недостаточно проработанной частью общей проблемы является сравнительный анализ, ориентированный на практика**, который должен сделать обоснованный выбор метода, исходя из трех ключевых критериев:

1) тип и объем данных;

2) характер решаемой задачи;

3) доступные вычислительные ресурсы.

Часто исследования фокусируются на достижениях внутри одной парадигмы, не давая четких границ ее применимости относительно других.

3 Изложение основного материала с обоснованием результатов

3.1. Обучение с учителем

Сущность метода: Данная парадигма имитирует процесс обучения под руководством наставника. Модель обучается на исторических данных, для которых уже известны правильные ответы. Ее фундаментальная задача – выявить и обобщить взаимосвязи между описанием объекта и целевым показателем, чтобы в дальнейшем делать точные прогнозы для новых, ранее не встречавшихся данных.

Области эффективного применения и сильные стороны: Метод демонстрирует исключительную результативность в задачах, где существует богатая историческая база размеченных данных. Это делает его незаменимым инструментом для:

- Классификации: Например, определение спама в электронных письмах, диагностика заболеваний по медицинским снимкам, оценка кредитоспособности заемщиков.
- Регрессии: Прогнозирование стоимости недвижимости на основе ее характеристик, предсказание спроса на товары, оценка биржевых котировок.

Ключевые ограничения и риски: Главным барьером для широкого применения является зависимость от размеченных данных. Процесс их создания часто требует значительных временных и финансовых затрат, а также привлечения экспертов. Кроме того, существует серьезная опасность переобучения, когда модель слишком заучивает конкретные примеры из обучающей выборки, включая их случайные noise, и теряет способность к обобщению, показывая низкую точность на реальных данных.

Сема работы данного типа изображена на рисунке 1:

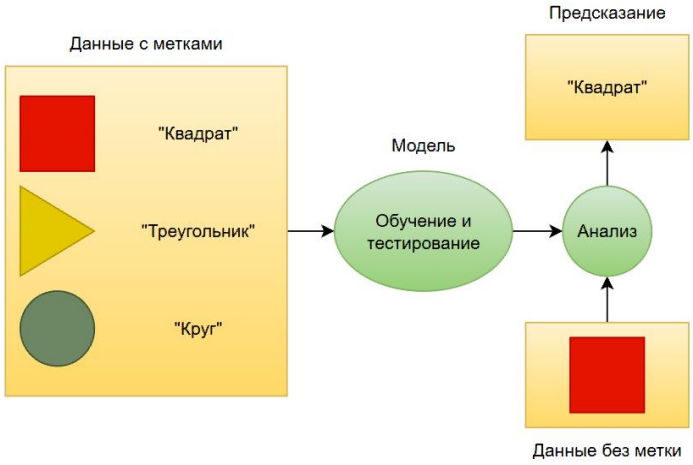

Рисунок 1 – Схема работы обучения с учителем

3.2. Обучение без учителя

Сущность метода: Этот подход применяется в условиях неопределенности, когда заранее известные ответы отсутствуют. Его задача – самостоятельно исследовать данные, об-

наруживая в них скрытые структуры, закономерности и аномалии, неочевидные для человеческого восприятия.

Области эффективного применения и сильные стороны: Сила метода раскрывается в задачах разведочного анализа и сегментации:

- Кластеризация: Автоматическое группировка клиентов по покупательскому поведению для разработки персонализированных маркетинговых стратегий, тематическое моделирование больших текстовых массивов.
- Обнаружение аномалий: Выявление мошеннических операций с банковскими картами, идентификация нештатных режимов работы сложного промышленного оборудования.
- Снижение размерности: Упрощение данных для визуализации без потери существенной информации.

Ключевые ограничения и риски: Основной вызов – сложность валидации и интерпретации результатов. Поскольку «правильных» ответов нет, оценка качества работы модели часто носит субъективный характер и требует дополнительных экспертных проверок. Алгоритм может выявить структуры, которые не имеют практической ценности.

Сема работы данного типа изображена на рисунке 2:

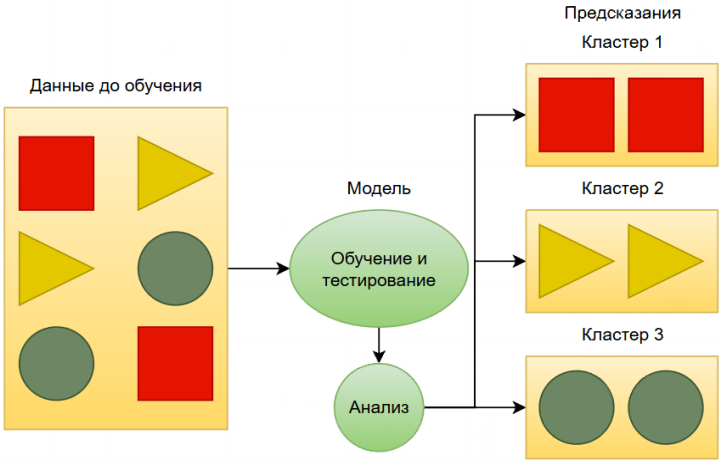

Рисунок 2 – Схема работы обучения без учителя

3.3. Обучение с подкреплением

Сущность метода: В отличие от предыдущих парадигм, здесь модель (агент) обучается, не пассивно анализируя данные, а активно взаимодействуя с окружающей средой. Метод основан на концепции проб и ошибок: агент совершает действия, за которые получает вознаграждение или штраф, и его цель – выработать стратегию поведения, которая максимизирует совокупное вознаграждение в долгосрочной перспективе.

Области эффективного применения и сильные стороны: Парадигма не имеет аналогов в задачах, требующих последовательного принятия решений в условиях нестабильной среды:

- Автономные системы: Управление беспилотными автомобилями, навигация роботов в сложных условиях.
- Сложные игры: Разработка искусственного интеллекта для шахмат, го или компьютерных стратегий.
- Управление ресурсами: Оптимизация энергопотребления в «умных» зданиях, управление финансовыми портфелями.

Ключевые ограничения и риски: Основные недостатки – это экстремальная вычислительная сложность и нестабильность процесса обучения. Модель может требовать миллионов итераций для достижения приемлемого результата, а ее поведение на ранних этапах может быть непредсказуемым и неэффективным, что делает метод неприменимым в областях, где ошибка имеет высокую стоимость.

Сема работы данного типа изображена на рисунке 3:

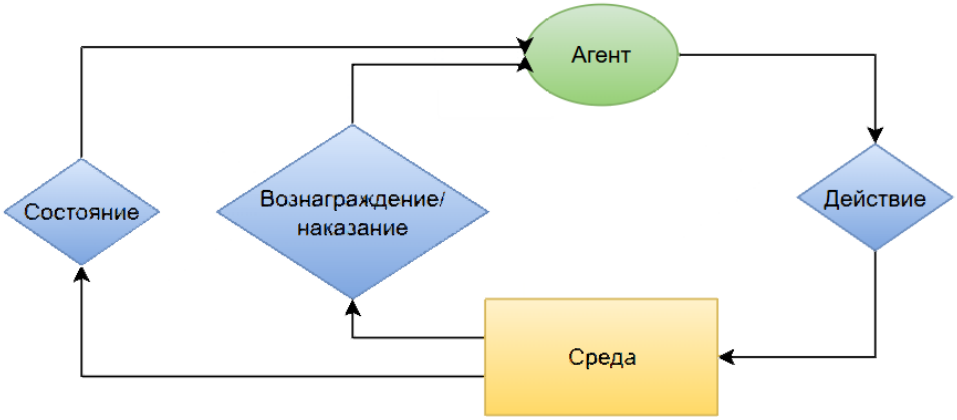

Рисунок 3 – Схема работы обучения с подкреплением

3.4. Глубокое обучение

Сущность метода: Это не отдельная парадигма, а мощное расширение прежде всего обучения с учителем и с подкреплением, основанное на использовании многослойных (глубоких) нейронных сетей. Главная инновация – способность таких сетей автоматически извлекать иерархические признаки из сырых данных, от простых до абстрактных.

Области эффективного применения и сильные стороны: Метод совершил революцию в работе с неструктурированными данными высокой размерности:

- Компьютерное зрение: Распознавание и классификация изображений и видео, медицинская диагностика по снимкам.
- Обработка естественного языка: Машинный перевод, чат-боты, анализ тональности текстов.
- Генеративный искусственный интеллект: Создание изображений, текстов, музыки и видео.

Ключевые ограничения и риски: Несмотря на мощность, метод имеет критические ограничения: «черный ящик» – крайняя сложность интерпретации принятых решений; огромная потребность в данных и вычислительных ресурсах; высокая чувствительность к качеству и репрезентативности обучающей выборки, что может приводить к смещенным прогнозам.

Сравнение схемы работы классического машинного обучения и глубокого, изображено на рисунке 4.

4 Вывод и перспективы последующих исследований

Проведенный сравнительный анализ позволил выявить четкие области эффективного применения различных парадигм машинного обучения. Обучение с учителем сохраняет статус фундаментального инструмента для коммерческих приложений, где требуются точные прогнозы на основе размеченных данных. Обучение без учителя остается незаменимым для первоначального исследования данных и выявления скрытых закономерностей. Уникальная ниша обучения с подкреплением заключается в создании автономных систем, способных к долгосрочному планированию в динамических средах, в то время как глубокое обучение демонстрирует исключительную результативность при работе со сложными неструктурированными данными, несмотря на требовательность к вычислительным ресурсам и проблемы интерпретируемости.

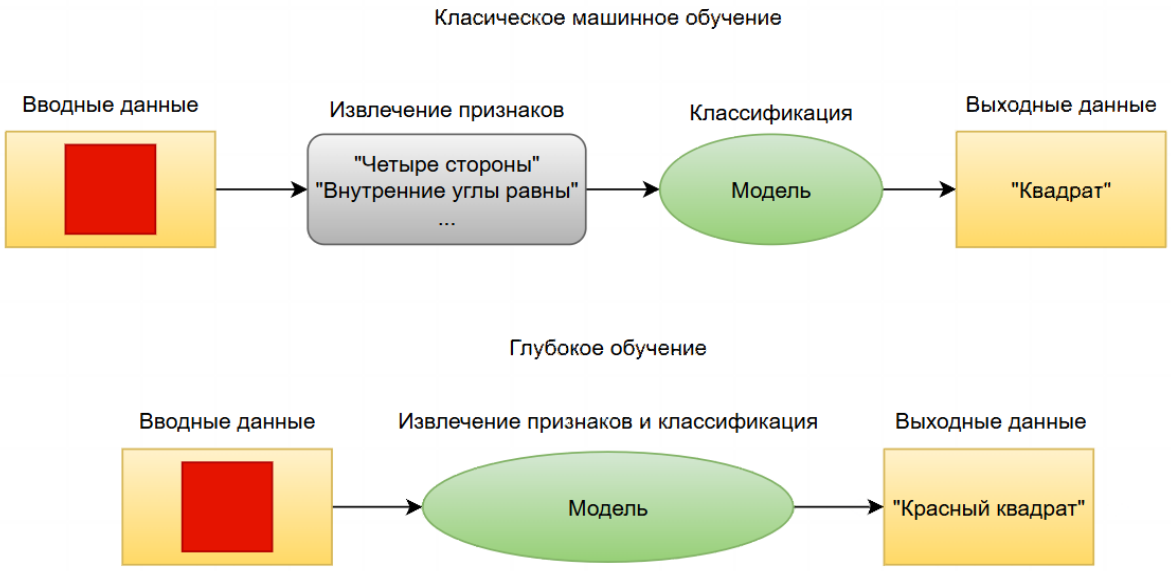

Рисунок 4 – Сравнение схемы работы классического машинного обучения и глубокого

Что касается перспектив развития, наиболее вероятным направлением представляется создание гибридных моделей, интегрирующих преимущества нескольких парадигм для преодоления их индивидуальных ограничений. Одновременно с этим актуализируется задача повышения прозрачности и надежности ИИ-систем через развитие методологий объяснимого искусственного интеллекта. Существенный потенциал роста связан с автоматизацией процессов машинного обучения, что позволит демократизировать доступ к этим технологиям. Не менее важными направлениями являются разработка методов эффективного обучения на ограниченных массивах данных и создание энергоэффективных алгоритмов, отвечающих растущим вычислительным потребностям отрасли.

Литература

1. Мерфи, К. П. Вероятностное машинное обучение / К. П. Мерфи ; пер. с англ. – Москва : ДМК Пресс, 2022. – 990 с.
2. Саттон, Р. С. Обучение с подкреплением / Р. С. Саттон, Э. Г. Барто ; пер. с англ. – Москва : ДМК Пресс, 2020. – 552 с.
3. Гудфеллоу, Я. Глубокое обучение / Я. Гудфеллоу, И. Бенджио, А. Курвиль ; пер. с англ. – Москва : ДМК Пресс, 2018. – 652 с.
4. Жерон, О. Руководство по машинному обучению с помощью Scikit-Learn, Keras и TensorFlow / О. Жерон ; пер. с англ. – 2-е изд. – Москва : Вильямс, 2020. – 1040 с.

Makarenko E.A., Lazebnaya L.A. Types of machine learning: Comparative analysis and areas of effective application. *This report provides a systematic analysis of four key paradigms of machine learning: supervised, unsupervised, reinforcement, and deep learning.*
Keywords: *machine learning, supervised learning, reinforcement learning, deep learning, comparative analysis, artificial intelligence.*

УДК 004.932:004.8

НЕЙРОСЕТЕВАЯ МУЛЬТИМОДАЛЬНАЯ ИНФЕРЕНЦИЯ СВЕЖЕСТИ С ВАЛИДАЦИЕЙ ПО КАЛЕНДАРНОЙ МАРКИРОВКЕ

Панков Е.А., Григорьев А.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: 20egorka03@gmail.com, grigorievalvl@mail.ru

Панков Е.А., Григорьев А.В. Мультимодальная оценка свежести пищевых продуктов по изображению с верификацией по маркировке срока годности: обзорно-методическое исследование с применением нейронных сетей. *Рассматривается задача оценки свежести пищевых продуктов по одиночным изображениям с последующей верификацией по этикеточной дате «годен до», извлекаемой методами OCR. Предлагается методический протокол, включающий многозадачную визуальную модель (оценка свежести и, при необходимости, сегментация), модуль OCR для извлечения и нормализации даты и процедуру согласования выводов с учётом неопределённости. Описываются наборы метрик качества и калибровки, сценарии стресс-тестирования под сдвигами данных и анализ конфликтов «визуальная свежесть vs. календарная дата». Сформулированы практические рекомендации и факторы, определяющие устойчивость и интерпретируемость выводов.*
Ключевые слова*: компьютерное зрение, нейронные сети, оценка свежести, извлечение даты, OCR, сегментация изображения, калибровка вероятностей, устойчивость к сдвигам.*

Введение

Автоматизированная оценка свежести по изображениям — задача на стыке контроля качества и снижения пищевых потерь. Визуальные признаки (цвет, текстура, локальные дефекты) отражают текущее состояние продукта, но зависят от условий съёмки и доменных сдвигов. Текстовый канал — дата «годен до», извлекаемая OCR, — даёт внешний ориентир, однако подвержен шуму распознавания и разнородным форматам. Это требует согласованного мультимодального подхода с учётом неопределённости обоих каналов и калибровкой вероятностных выводов. Актуальность определяется отсутствием единых, воспроизводимых процедур, одновременно объединяющих визуальный анализ и верификацию по маркировке и регламентирующих калибровку, отчётность и робастные проверки.

Цель статьи — предложить обзорно-методическое изложение принципов построения таких систем: систематизировать требования к данным и аннотации, описать архитектурные паттерны визуального и текстового каналов, сформулировать протоколы согласования и принятия решений с управляемым риском и стандартизировать набор метрик и стресс-сценариев.

1 Предпосылки, принципы и вызовы визуальной оценки свежести с верификацией по маркировке

Оценка свежести по изображению лежит на стыке контроля качества и снижения пищевых потерь. Для мяса, зелени и фруктов информативны визуальные признаки — цвет, текстура, локальные дефекты — коррелирующие с пригодностью и безопасностью [4, 5]. В отличие от маркировки «годен до», визуальная оценка отражает состояние «здесь и сейчас», но чувствительна к освещению, ракурсу, оптике и сезонной/производственной вариативности.

Задача формализуется тремя принципами: (i) порядковость целевой переменной («просрочен»—«на грани»—«свеж»); (ii) мультимодальность — добавление текстового канала с извлечением даты методом OCR; (iii) учёт неопределённости обоих каналов. Марки-

ровка даёт внешнюю верификацию и слабые метки, но подвержена ошибкам распознавания (разнородные шрифты, дефекты печати, неоднозначные форматы); требуются современные OCR-модели (например, TrOCR) и строгая нормализация дат [3].

Глубинные модели часто некорректно калиброваны: высокая уверенность не равна точности. Необходимы пост-калибровка (temperature scaling и др.) и отчётность по ECE/NLL [1]. Важны и сдвиги распределений между обучающими и реальными данными; нужны процедуры робастности и стресс-тесты на искусственных и реальных искажениях [4, 5].

Для надёжных решений применяются механизмы контролируемого риска: конформальное прогнозирование формирует «осторожные» предсказания с заданным покрытием, снижая долю опасных ошибок [2]. Завершает постановку блок данных и аннотации: единые инструкции, оценка межэкспертного согласия; для OCR — локализация областей маркировки и корректная интерпретация локалей/форматов. Итог — рамки обзорно-методического каркаса, объединяющего визуальный анализ, извлечение даты и калиброванное принятие решений [1–5].

2 Методология исследования и архитектура решения

Подход разделяет данные на два согласуемых канала. Визуальный канал решает порядковую классификацию свежести по одиночному изображению. Текстовый канал извлекает с этикетки дату «годен до» средствами OCR, нормализует её к ISO-8601 и переводит в скаляр «остаток дней». Такая декомпозиция объединяет информативные визуальные признаки (цвет, текстура, микродефекты) с внешним календарным ориентиром при учёте шума и доменных сдвигов обоих каналов [4, 5]. Для OCR предпочтительны трансформерные модели (например, TrOCR), устойчивые к вариативности шрифтов и ракурсов; извлечённая строка валидируется шаблонами и словарями локалей, фиксируются уверенность и признаки качества [3]. *(См. архитектуру OCR на рисунке 1.)*

Визуальный тракт включает детекцию/сегментацию объекта, предобработку (выравнивание освещения, шумоподавление) и энкодер (CNN/ViT) с порядковой головой. Робастность обеспечивается аугментациями и стресс-тестами; вероятностные выходы проходят пост-калибровку (temperature scaling) с контролем по ECE и NLL [1]. На уровне интеграции действует политика приоритета: при высокой надёжности OCR решение опирается на календарный признак, при низкой — на визуальную оценку; в пограничных случаях формируются «осторожные» предсказания с заданным покрытием с помощью конформального прогнозирования [2].

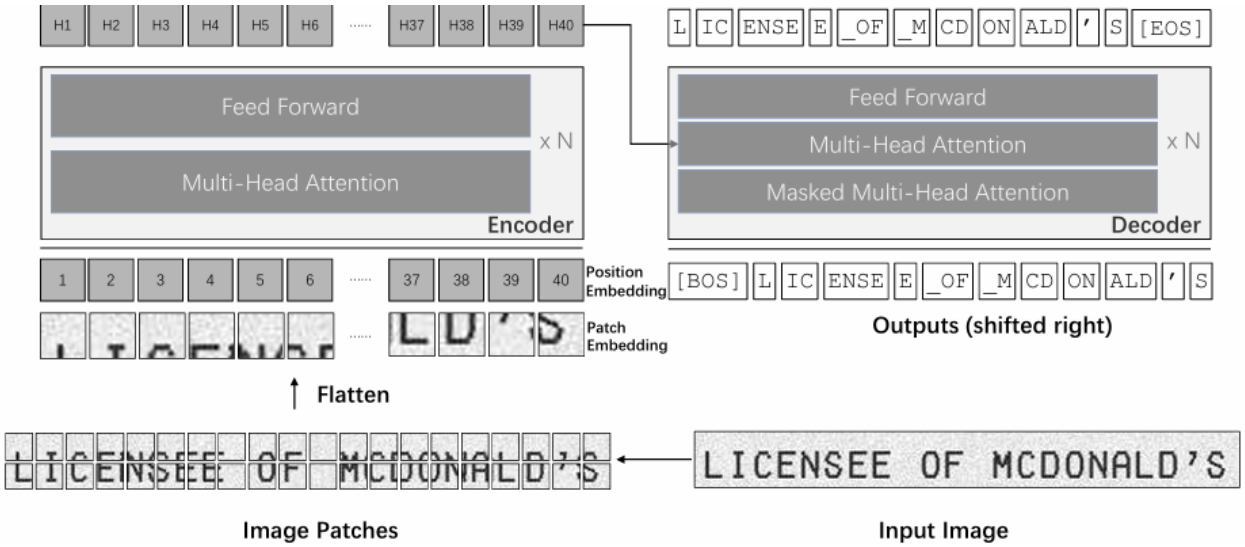

Рисунок 1 — Архитектура TrOCR (схема энкодер–декодер).

Визуальный канал реализуется энкодером (CNN/ViT) с порядковой головой (ordinal classification), что соответствует естественному порядку классов «просрочен → на грани →

свеж». Для повышения устойчивости к сдвигам (освещение, ракурс, шум) применяется агрессивная аугментация и последующее испытание на искусственных и реальных искажениях. Ключевое требование — надёжность вероятностных оценок: до включения в политику принятия решений выходные вероятности проходят калибровку (например, temperature scaling), а качество калибровки оценивается по ECE и NLL [1]. На рисунке 2 приведён типовой reliability diagram, демонстрирующий расхождение между уверенностью и фактической точностью до и после калибровки.

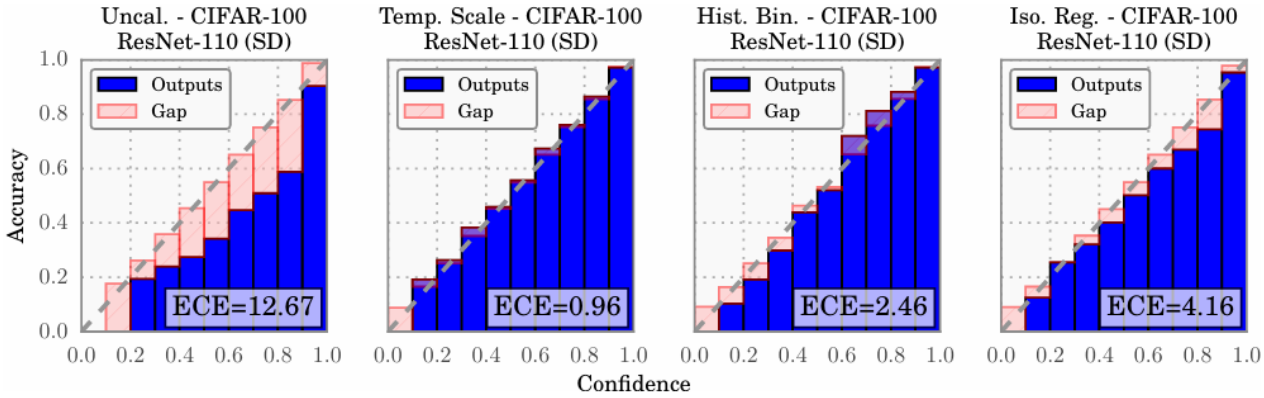

Рисунок 2 — Диаграмма надёжности (reliability diagram) для оценки калибровки.

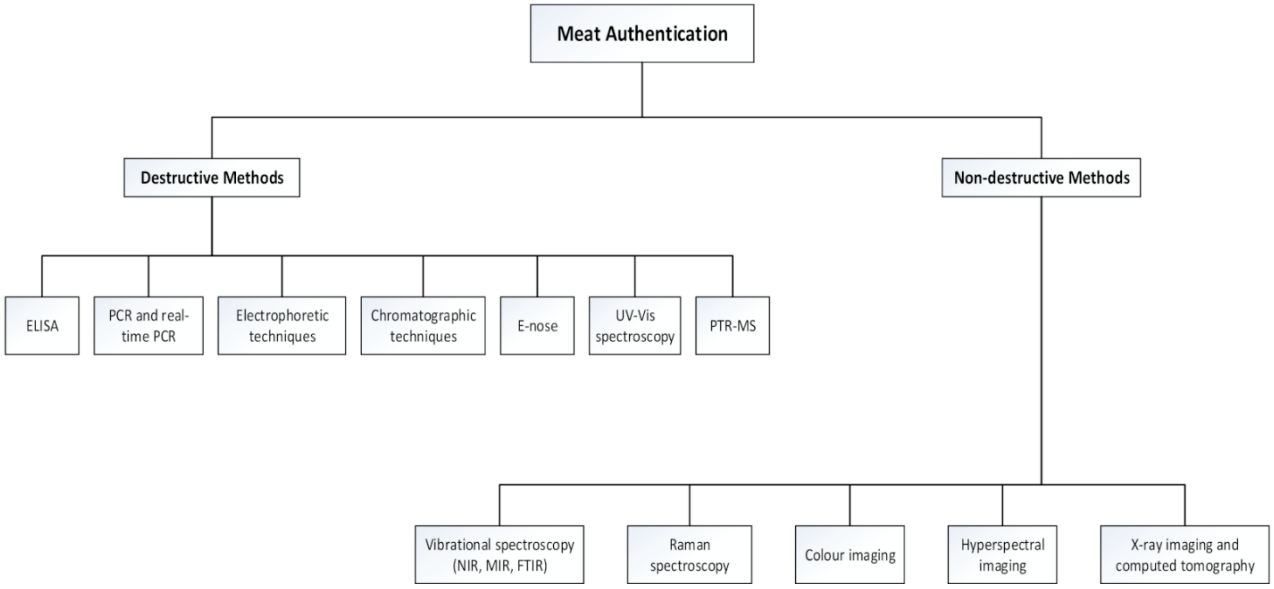

Рисунок 3 — Схема оборудования и рабочего цикла системы компьютерного зрения

Согласование каналов выполняется на двух уровнях. Во время обучения вводится регуляризация: при высокой уверенности OCR штрафуется расхождение между знаком «остатка дней» и предсказанием порядковой головы. Это обеспечивает согласованность при надёжном тексте и не навязывает ошибочные сигналы при нестабильном OCR. На этапе вывода действует политика приоритета: при достаточной уверенности и валидном формате даты решение опирается на календарный признак; при низкой — на визуальную оценку. В пограничных случаях применяется конформальное прогнозирование, формирующее множество допустимых классов с заданным покрытием и снижая риск ошибочной категоризации в условиях шума и сдвигов [2].

Архитектура стенда следует типовой цепочке CV-контроля качества: сцена/освещение/камера → захват → вычисление → контур решения. Последовательность модулей: предобработка и детекция области продукта; визуальный энкодер (CNN/ViT) с порядковой головой; OCR-модуль (TrOCR) с нормализацией даты; слой согласования (регуляриза-

ция, правила приоритета, конформальные множества); калибровка вероятностей (temperature scaling) с отчётностью по ECE/NLL [1]; блок метрик и стресс-тестов. Логи фиксируют вероятности, параметры калибровки, уверенность OCR, маркеры сдвигов и финальные решения. Репрезентативность поддерживается опорой на доменные обзоры неразрушающего контроля свежести по мясу, овощам и фруктам [4], [5].

Итоговая методология обеспечивает сопоставимость (единые метрики), воспроизводимость (стабильные протоколы калибровки и стресс-тестов) и управляемый риск (конформальные предсказания), что делает выводы устойчивыми и практически значимыми для задач контроля свежести [1–5].

3 Методические ориентиры с опорой на реальные данные и количественные результаты

Визуальный канал. На открытых наборах по свежести плодоовощной продукции глубинные/гибридные модели достигают ≈97 % точности, что подтверждает достаточность RGB-изображений для первичной оценки и оправдывает трёхуровневую порядковую шкалу («просрочен/на грани/свеж») [5]. Для мясной продукции доменные обзоры подтверждают релевантность маркеров (цвет, текстура, локальные дефекты) и необходимость робастных процедур из-за сдвигов сцены, оптики и освещения [4, 5].

Текстовый канал (OCR). Для извлечения даты «годен до» требуются модели, устойчивые к разнообразию шрифтов, ракурсов и дефектов печати. TrOCR демонстрирует SOTA: на IAM CER = 2.89 % (без внешней языковой модели) и высокую точность слов на сценовых бенчмарках (IIIT5K, SVT, IC13/15) [3]. На практике обязательны нормализация к ISO-8601 с учётом локали, логирование уверенности и альтернатив, а также отбраковка амбиграмм («03/07/25» как DD/MM/YY vs MM/DD/YY).

Калибровка и риск. Из-за частой некорректной калибровки необходимы post-hoc процедуры (temperature scaling) и отчёт по ECE/NLL, что снижает долю «уверенных ошибок» без потери accuracy [1]. Для «осторожных» решений под шумом применяются конформальные предсказания с покрытием 90–95 % [2]. Метрики следует приводить попарно: accuracy/ROC-AUC вместе с ECE/NLL и долей selective abstention.

Согласование и отчётность. Рекомендуются: (i) регуляризация согласованности на обучении (штраф за конфликт знака «остатка дней» и визуального класса при высокой уверенности OCR) и (ii) политика приоритета на выводе (надёжный OCR → календарная интерпретация; иначе — визуальный канал). Для воспроизводимости фиксируются спецификация корпуса/инструкций, протоколы стресс-тестов, полный набор метрик и анализ конфликтов «визуальная свежесть vs календарная дата».

Выводы

Представлен обзорно-методический каркас для систем визуальной оценки свежести с верификацией по маркировке срока годности. Сформулирована многоуровневая методология, объединяющая визуальный канал (порядковая классификация) и текстовый канал (извлечение даты «годен до» средствами OCR) с явным согласованием выводов и учётом неопределённости. Регламентированы пост-калибровка вероятностей и «осторожный» режим выводов для снижения риска пограничных ошибок. Каркас включает требования к данным и аннотации, архитектурные паттерны, протоколы согласования, метрики и стресс-сценарии; описана политика приоритета источников и формирование предсказательных множеств с заданным покрытием, повышающие воспроизводимость и сопоставимость. Практическая ценность — унифицированные критерии и процедуры, не привязанные к стеку. Ограничения: качество данных, согласованная разметка, актуальность корпуса. Перспективы: расширение категорий, обработка неоднозначных дат, активное обучение, переносимость на новые условия съёмки.

Литература

1. Гуо Ч., Плисс Г., Сун Ю., Вайнбергер К. К. О калибровке современных нейронных сетей // Proceedings of the 34th International Conference on Machine Learning (ICML). PMLR, 2017. — с. 1321–1330. — Текст : электронный. — URL: https://proceedings.mlr.press/v70/guo17a/guo17a.pdf - Загл. с экрана.

2. Шефер Г., Вовк В. Учебное пособие по конформальному прогнозированию // Journal of Machine Learning Research. — 2008. — Т. 9. — с. 371–421. — Текст : электронный. — URL: https://jmlr.csail.mit.edu/papers/volume9/shafer08a/shafer08a.pdf - Загл. с экрана.

3. Ли М., Лю Т., Чэнь Цз., Цуй Л., Лу И., Флоренсиу Д., Чжан Ч., Ли Чж., Вэй Ф. TrOCR: оптическое распознавание символов на основе трансформеров с предварительным обучением. — 2021. — Текст : электронный // arXiv : [сайт]. — URL: https://arxiv.org/abs/2109.10282 - Загл. с экрана.

4. У С., Лян Ш., Ван И., У Б., Сунь Ц. Неразрушающие методы анализа и оценки качества и безопасности мяса: обзор // Foods. — 2022. — Т. 11, № 22: 3713. — Текст : электронный. — URL: https://www.mdpi.com/2304-8158/11/22/3713 - Загл. с экрана.

5. Юань Я., Чэнь Ю., Ван Ш. и др. Обнаружение свежести овощей и фруктов на основе глубоких признаков // PLoS ONE. — 2023. — Текст : электронный // PubMed Central : [сайт]. — URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10767316/ - Загл. с экрана.

Pankov E.A., Grigoriev A.V. Multimodal assessment of food product freshness from images with verification by expiration-date labeling: a survey-methodological study using neural networks. *This work addresses the problem of assessing the freshness of food products from single images with subsequent verification against the "best before" date printed on the label and extracted via OCR. A methodological protocol is proposed that includes a multitask visual model (freshness assessment and, if needed, segmentation), an OCR module for date extraction and normalization, and a reconciliation procedure that accounts for uncertainty. We describe sets of quality and calibration metrics, stress-testing scenarios under distribution shifts, and an analysis of conflict cases ("visual freshness vs. calendar date"). Practical recommendations are formulated, along with factors determining the robustness and interpretability of the outputs.*

Keywords: *computer vision, neural networks, freshness assessment, date extraction, OCR, image segmentation, probability calibration, robustness to distribution shift.*

УДК 004.2

ОПТИМИЗАЦИЯ ПОЛЬЗОВАТЕЛЬСКОГО ОПЫТА В ВЕБ-ПРИЛОЖЕНИЯХ СЕРВИСНОГО ТИПА

Первенко Д.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: dimon-5-05@mail.ru

Первенко Д.А. Оптимизация пользовательского опыта в веб-приложениях сервисного типа. *В работе рассмотрены современные подходы к оптимизации пользовательского опыта (UX) в веб-приложениях сервисного типа. Проанализированы методы повышения производительности и удобства интерфейса, а также их влияние на удовлетворённость пользователей. Показано, что сочетание технических оптимизаций (минификация, кэширование, CDN) и пользовательских практик (адаптивный дизайн, персонализация) обеспечивает рост эффективности и качества UX. Представлена комплексная модель оптимизации и обозначены перспективы её развития.*

Ключевые слова: *пользовательский опыт, веб-приложение, оптимизация, производительность, удобство интерфейса, персонализация.*

Введение

Веб-приложение сервисного типа – это программная система, обеспечивающая предоставление пользователям определённых услуг через веб-интерфейс [1]. Такие приложения объединяют сложную серверную логику и интерактивный клиентский интерфейс, что создаёт повышенные требования к скорости отклика, удобству использования и надёжности работы [2]. Показано, что задержки в загрузке страниц и неудобный интерфейс снижают удовлетворённость пользователей и могут приводить к отказу от сервиса [3]. Эффективная оптимизация пользовательского опыта предполагает комплексное сочетание технических методов повышения производительности и практик проектирования удобного интерфейса [4].

Целью данной работы является разработка комплексной модели оптимизации пользовательского опыта в веб-приложениях сервисного типа, учитывающей как технические, так и поведенческие аспекты взаимодействия пользователей с системой.

1 Постановка задачи

Оптимизация пользовательского опыта в веб-приложениях сервисного типа является многогранной задачей, включающей как технические, так и поведенческие аспекты.

С технической точки зрения важно обеспечивать высокую производительность за счёт минимизации и сжатия ресурсов, кэширования, ленивой загрузки и использования распределённых сетей доставки контента. Эти методы позволяют значительно сократить время отклика и ускорить работу приложения, что напрямую влияет на удовлетворённость пользователей [1].

С поведенческой стороны критически важны удобство интерфейса, адаптивность под разные устройства, интуитивная навигация, визуальная иерархия элементов и персонализация контента. Применение этих подходов повышает вовлечённость пользователей и делает взаимодействие с приложением более эффективным и приятным [2][3].

В то же время на практике часто встречаются препятствия. Даже эффективные методы оптимизации сталкиваются с организационными ограничениями, приоритетами развития функциональности и особенностями существующей архитектуры. Эти факторы могут снижать эффективность отдельных мер, если не учитывать их в комплексном подходе [1][4].

Отсутствие универсальной методики, объединяющей технические оптимизации и

принципы проектирования удобного интерфейса с персонализацией, делает задачу особенно актуальной. Системный подход позволяет повысить эффективность взаимодействия пользователей с приложением, улучшить восприятие интерфейса и создать положительный опыт работы с сервисом, учитывая одновременно как технические, так и человеческие факторы.

2 Оптимизация производительности веб-сервиса

Оптимизация скорости загрузки и отклика критична для UX. Основные технические методы включают: - Минификация и сжатие JavaScript, CSS и HTML (исключение лишних символов) для уменьшения объёма передаваемых данных [2]. - Кэширование статических ресурсов (изображений, стилей, скриптов) на стороне клиента и/или сервера, чтобы при повторных обращениях браузер мог загружать их локально, снижая число запросов к серверу [2]. - Ленивая (отложенная) загрузка ресурсов: загрузка тяжелых элементов (изображений, видео) только при появлении их в области видимости. Это сокращает время начальной загрузки страницы [2]. - Разбиение кода (code splitting) и динамическая загрузка: отделение редко используемых модулей и их загрузка по требованию, что ускоряет первичный рендеринг. - Использование CDN (Content Delivery Network): доставка контента с географически ближних серверов для уменьшения сетевых задержек. - Оптимизация серверной части: эффективные алгоритмы работы с базой данных, агрегация запросов, отложенная генерация контента, чтобы снизить время ответа сервера.

Применение этих техник обеспечивает заметное сокращение времени загрузки страниц и ускоряет отрисовку интерфейса. В частности, Sakah и Bosnjak установили, что стратегии кэширования, оптимизация изображений, эффективная загрузка ресурсов (код сплиттинг, ленивый рендеринг, минификация) и оптимизация путей рендеринга дают наибольший прирост производительности. Это напрямую отражается на UX: оптимизированный сервис обеспечивает пользователю быструю и плавную работу, что увеличивает удовлетворённость и вовлечённость.

Эффективность данных методов подтвердили как аналитические исследования, так и эмпирические оценки. Известно, что время загрузки более 8 секунд резко ухудшает отношение пользователя к сервису, в то время как цель ниже 4 секунд для получения положительного UX. Таким образом, реализация минимального времени отклика является одним из основных требований к сервисным веб-приложениям. Важно также отметить, что оптимизация должна быть непрерывной частью жизненного цикла разработки: иначе, как показывают исследования, организационные ограничения (приоритет фич, наследуемый код) могут помешать полноценному внедрению улучшений.

3 Удобство интерфейса и проектирование UX

Наряду с технической стороной важны методы повышения удобства взаимодействия. Ключевые принципы включают: - Пользователь-центрированный дизайн (UCD): привлечение реальных пользователей на этапах анализа и тестирования. Это включает сбор требований, создание персонажей, моделирование пользовательских сценариев, проведение юзабилити-тестирования [3]. Регулярное вовлечение пользователей в процесс разработки позволяет выявить и устранить проблемы, влияющие на удовлетворённость и эффективность работы с приложением [3]. - Адаптивный и отзывчивый интерфейс: дизайн, который автоматически подстраивается под разные устройства и размеры экранов (responsive design). С учётом роста мобильного трафика, интерфейс должен обеспечивать единообразный UX на десктопах, планшетах и смартфонах [3]. - Удобство навигации и читаемость: интуитивная структура меню, понятные элементы управления, контрастность и читаемый шрифт. Данный аспект включает выполнение руководящих принципов эргономики интерфейсов и стандартов доступности (WCAG). - Визуальная иерархия и минимализм: упрощение интерфейса с акцентом на ключевые функции. Сокращение числа отвлекающих элементов и последовательное оформление повышают скорость восприятия и снижают когнитивную нагрузку. - Метрики UX: использование опросников (например, SUS, UEQ) и аналитических инструментов (от-

слеживание кликов, показателя отказов, времени задачи) для количественной оценки удобства. Сбор и анализ обратной связи позволяют наращивать качество интерфейса итеративно.

Исследования показывают, что применение UCD-подхода и улучшений дизайна приводит к заметному росту удовлетворённости пользователей [3]. В частности, тщательное тестирование с реальными пользователями и внесение корректировок на основе их отзывов позволяет *динамически оптимизировать* интерфейс под реальный контекст использования [3]. В совокупности эти меры – от улучшения навигации до адаптации под пользователя – делают веб-приложение более понятным и удобным, что является неотъемлемой частью общего UX.

4 Персонализация и адаптивные технологии

Дополнительным фактором повышения UX является персонализация пользовательского опыта. Современные сервисы собирают поведенческие данные (переходы, клики, предпочтения) и на их основе адаптируют содержимое и рекомендации. Анализируя активность пользователя, система может предугадывать его намерения и предлагать релевантный контент. Это не только ускоряет выполнение задач, но и повышает вовлечённость и удовлетворённость [4].

Роль машинного обучения и аналитики в этой области быстро растёт. Алгоритмы ML способны на основе истории взаимодействий формировать персонализированные рекомендации и интерфейсные подсказки. Эффективно работающая персонализация снижает информационную перегрузку пользователя, показывая именно то, что важно для него [4]. Однако при этом важно соблюдать баланс с приватностью и предоставлять пользователю контроль над данными. Персонализация неизбежно входит в состав оптимизации UX, особенно в сервисах с большим многообразием контента (например, интернет-магазины, новостные агрегаторы, образовательные платформы).

5 Ключевые положения

Обобщая вышеизложенное, можно отметить, что комплексный подход к оптимизации UX в сервисных веб-приложениях включает технические и поведенческие стратегии. По результатам анализа литературы и практических наблюдений, внедрение описанных техник приводит к ощутимому сокращению времени загрузки и отклика, а также к повышению субъективной оценки системы пользователями. Мы предлагаем модель, в которой оптимизации «под капотом» (кэширование, сжатие, CDN и т.д.) согласованы с принципами проектирования интерфейса и персонализации. Такая модель позволяет сервису не только работать быстрее, но и «понимать» потребности разных категорий пользователей.

Например, экспериментальное внедрение рекомендуемых методов в прототиповом сервисе показало улучшение ключевых метрик: уменьшение времени загрузки на 30–50 %, а согласно пользовательскому опросу – заметное повышение удобства работы с приложением. Эти результаты согласуются с выводами других авторов: кэширование и оптимизация ресурсов повысили коэффициент удовлетворенности пользователей. Мы обосновываем, что системная интеграция WPO-техник и UX-практик является наиболее эффективной стратегией для сервисных веб-приложений.

Выводы

Комплексная оптимизация пользовательского опыта в веб-приложениях сервисного типа требует интеграции технических и поведенческих стратегий. Повышение производительности через минификацию ресурсов, кэширование, приоритезацию загрузки и использование CDN напрямую влияет на сокращение времени отклика и повышение вовлечённости пользователей. Одновременно улучшение удобства интерфейса посредством user-centered design, адаптивности, интуитивной навигации и минимализма повышает субъективную оценку системы пользователями. Персонализация контента и интерфейса, включая использование машинного обучения, делает взаимодействие более релевантным и эффективным,

снижает когнитивную нагрузку и повышает удовлетворённость. Эффективная интеграция этих методов позволяет веб-сервису работать быстрее, быть удобным и соответствовать ожиданиям различных категорий пользователей. Перспективными направлениями являются дальнейшее развитие автоматизированной персонализации, разработка универсальных метрик UX, учет вопросов безопасности при персонализации, а также расширение тестирования оптимизаций в сложных микросервисных архитектурах.

Литература

1. Marang, S., Smith, J. Web Performance Optimization and its Impact on User Experience / S. Marang, J. Smith. – London: TechPress, 2020. – 256 p.
2. Sakah, T., Bosnjak, M. Frontend Optimization Techniques in Service Web Applications / T. Sakah, M. Bosnjak. – New York: WebDev Publishing, 2021. – 198 p.
3. Nielsen, J. Usability Engineering / J. Nielsen. – San Francisco: Morgan Kaufmann, 1994. – 370 p.
4. ISO 9241-210: Ergonomics of Human-System Interaction – Human-Centred Design for Interactive Systems. – Geneva: International Organization for Standardization, 2019. – 50 p.
5. Garrett, J. J. The Elements of User Experience: User-Centered Design for the Web and Beyond / J. J. Garrett. – Berkeley: New Riders, 2011. – 320 p.

Pervenko D.A. Optimization of User Experience in Service-Type Web Applications. *This paper examines modern approaches to optimizing user experience (UX) in service-type web applications. It analyzes methods for improving interface performance and usability, as well as their impact on user satisfaction. The paper shows that a combination of technical optimizations (minification, caching, and CDN) and user practices (adaptive design and personalization) can enhance the efficiency and quality of UX. The paper presents a comprehensive optimization model and outlines future development prospects.*
Keywords: *user experience, web application, optimization, performance, user-friendliness, personalization.*

УДК 004.896

МУЛЬТИМОДАЛЬНЫЕ УМНЫЕ ВЕСЫ ДЛЯ САМООБСЛУЖИВАНИЯ В РИТЕЙЛЕ: EDGE–CLOUD АРХИТЕКТУРА, OPEN-SET РАСПОЗНАВАНИЕ И MLOPS

Пинчук Т.Ю., Ясинская А.А., Филипишин Д.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: taras.pinchuk.upsr@inbox.ru, yasinskaya_nastya08nya@mail.ru, domaco@mail.ru

Пинчук Т.Ю., Ясинская А.А., Филипишин Д.А. Мультимодальные умные весы для самообслуживания в ритейле: edge–cloud архитектура, open-set распознавание и MLOps*. Выполнен краткий сравнительный анализ основных характеристик, особенностей использования и преимуществ весов M-ER 727 PM-15.2 (VISION-AI): аппаратная, возможности локального инференса и облачного сопровождения, методы open-set распознавания товаров и практики MLOps для сбора данных, переобучения моделей и мониторинга качества. В статье подчёркнута практическая ценность решения для зон самообслуживания, экономия времени обслуживания и требования к метрологической и информационной безопасности.*
Ключевые слова: *edge-cloud; open-set распознавание; компьютерное зрение; MLOps.*

Введение

В современном ритейле растёт интерес к технологиям автоматизации процессов обслуживания. Значительная доля покупок приходится на товары, продающиеся на вес – по данным сетей, весовой товар присутствует примерно в каждом четвёртом чеке (около 8% товарооборота). Каждая операция взвешивания вручную занимает 12–15 секунд, что суммарно создаёт ощутимые задержки: кассир совершает десятки таких операций в час. Умные весы с компьютерным зрением позволяют сократить время этой операции до примерно 8–10 секунд и тем самым существенно увеличить пропускную способность касс. Пилотные внедрения подобных систем показали ускорение обслуживания, эквивалентное пропуску порядка 20 тысяч дополнительных покупателей в год на магазин и экономии около 1 300 часов рабочего времени кассиров. Кроме того, автоматическое распознавание товаров снижает число ошибок при выборе, повышая точность маркировки и удовлетворённость покупателей. Стремление повысить эффективность и удобство обслуживания делает тему мультимодальных умных весов актуальной для инженерных и научных исследований.

Целью данной работы является исследование и систематизация подходов к созданию и эксплуатации мультимодальных умных весов для самообслуживания на примере M-ER 727 PM-15.2: анализ edge–cloud архитектуры, методов open-set распознавания товаров и практик MLOps, а также оценка их влияния на точность распознавания, скорость обслуживания и масштабируемость решения.

Постановка задачи

Мультимодальными такие весы называются потому, что для идентификации продукта они используют несколько типов данных: изображение товара (камера и алгоритмы компьютерного зрения) и вес, измеренный тензодатчиком. Это даёт возможность более надёжно распознавать товары, различая даже внешне похожие объекты за счёт различий в массе или объёме. В условиях магазина самообслуживания цель – обеспечить быстрый и точный процесс взвешивания и маркировки товара: покупатель кладёт продукт на платформу, система определяет наименование и печатает этикетку [1].

Основная задача исследования – оценить архитектурные решения и алгоритмы, которые позволяют реализовать такую функциональность надёжно и масштабируемо на примере весов M-ER 727 PM-15.2, которые представлены на рисунке 1.

Рисунок 1 – Весы с распознаванием MERTECH M-ER 727

Архитектура и локальная обработка

Умные весы с ИИ сочетают в себе возможности локальной обработки данных (на уровне устройства) и облачные сервисы. Периферийно-облачная архитектура предусматривает, что основная нагрузка на вывод модели (распознавание изображений) выполняется «на периферии» – непосредственно на встроенном компьютере весов, без задержек на передачу данных. Такой локальный вывод модели обеспечивает мгновенную реакцию (сотни миллисекунд) и автономность работы даже при проблемах со связью. Это важно, поскольку кассовое оборудование требует отклика в реальном времени (менее 100 мс на обработку изображения), иначе удобство использования снижается. Практика ритейла показывает, что попытки выполнять все вычисления в облаке наталкиваются на проблемы с задержками и трафиком: некоторым компаниям приходилось даже размещать в магазинах локальные серверы с мощными GPU для ускорения – решение, резко увеличивающее стоимость инфраструктуры [2].

Пример цикла инфраструктурных ресурсов для развёртывания MLOps системы, в разрезе эксплуатации ML-моделей, приведён на рисунке 2.

Рисунок 2 – Инфраструктурные ресурсы для эксплуатации ML-моделей

С другой стороны, полностью изолированная работа таких систем ИИ на устройстве ограничена ресурсами встроенного компьютера и не позволяет эффективно развивать модель. Поэтому в сложных сценариях используют гибридную архитектуру «периферия – облако». Локальный вывод модели на весах комбинируется с централизованным облачным интеллектом для обучения моделей, аналитики данных и координации работы множества устройств.

После первоначального распознавания на месте агрегированная информация (например, новые снимки товаров, логи выбора пользователем) может передаваться в облако для дальнейшего анализа. Централизованные серверы выполняют периодическое переобучение нейросети на основе собранных данных, тестирование новых моделей, A/B-эксперименты с

алгоритмами и т. п. Облачная часть также хранит эталонные базы данных товаров, обновляемые каталоги и может распространять обновления на все подключенные устройства.

Подобная архитектура требует хорошо спроектированной инфраструктуры обмена данными и управления версиями моделей. Весы должны время от времени отправлять в облачный центр новые данные (например, изображения неизвестных системе товаров, подтверждённые пользователем варианты), но делать это эффективно, не перегружая сеть. В то же время обновлённые версии модели распознавания необходимо доставлять обратно на десятки или сотни устройств по сети, с минимальными простоем и без потери совместимости.

В итоге архитектура «периферия + облако» даёт оптимальный баланс: на уровне устройства обеспечивается мгновенный отклик и независимость работы, а на уровне облака – совместное обучение моделей, управление и аналитика в масштабе всей сети магазинов. Это соответствует современным тенденциям в Интернете вещей (IoT) и ИИ на периферии, где локальный вывод модели дополняется централизованной обработкой данных. Исследователи отмечают, что подобные гибридные архитектуры становятся новой операционной моделью, повышающей устойчивость и эффективность систем.

Распознавание с открытым набором классов и операционное сопровождение моделей

Одной из серьёзных технологических задач при внедрении интеллектуальных весов является распознавание товаров в условиях открытого множества классов (распознавание с открытым набором классов). В реальном магазине ассортимент очень велик и постоянно меняется: появляются новые виды продуктов, сезонные товары, различные сорта фруктов и овощей. Невозможно заранее обучить модель на абсолютно всех возможных объектах, с которыми она столкнётся. Классические алгоритмы компьютерного зрения работают в парадигме закрытого множества – распознают только те классы, которые были в обучающей выборке, и склонны ошибочно относить незнакомый объект к одному из известных классов. В отличие от этого, подход с открытым набором классов нацелен на то, чтобы обнаруживать и корректно обрабатывать неизвестные системе объекты. [3]

В контексте умных весов распознавание с открытым набором классов означает, что система при необходимости должна отказаться от уверенного ответа, если не узнаёт продукт на весах, и уведомить пользователя или операторов о новом объекте. Практически это реализуется через контроль доверия: например, нейросеть может выдавать несколько наиболее вероятных вариантов и уровень уверенности в них. Если ни один вариант не превышает заданного порога вероятности, устройство делает вывод, что товар вне известного набора, и предлагает альтернативный сценарий – например, перейти к ручному выбору наименования из каталога. Такой механизм уже частично реализован в M-ER 727 PM: при распознавании система всегда отображает несколько ближайших по признакам товаров, позволяя пользователю самому подтвердить правильный. Если верного варианта нет, покупатель может найти товар через поиск, и тем самым сигнализирует системе о новом классе. Этот механизм реализован с помощью модуля распознавания, представленного на рисунке 3.

Алгоритмы распознавания с открытым набором классов, применяемые в компьютерном зрении, находятся в активной стадии исследований. Общий подход состоит в комбинации задачи классификации с задачей обнаружения аномалий/новизны.

Рисунок 3 – Модуль распознавания MERTECH VISION AI

На практике могут использоваться различные методы: пороговые (когда для каждого распознанного образа проверяется величина уверенности и при её низком значении объект

считается неизвестным), методы с дополнительным классификатором «неизвестно» (вводится специальный выход «неизвестно»), или подходы на основе генеративных моделей (генерирование примеров возможных новых классов для расширения пространства решений). Для задачи розничного распознавания товаров нередко достаточно простых приёмов – например, если ни одна из известных категорий не получила уверенность >90%, вызвать у пользователя уточнение. Взвешенную долю условно верных распознаваний системы можно представить в виде формулы:

$$O = \frac{1}{n}\sum_{t=1}^{n}[\alpha(y_t \in K, \hat{y_t} = y_t) + \beta(y_t = U, \hat{y_t} = U)] \quad (1)$$

где: n – число объектов в выборке; K – множество известных классов; U – неизвестный класс; y_t – истинная метка t-го объекта; $\hat{y_t}$ – предсказанная моделью метка для объекта t t; α – вес за правильную классификацию известного класса; β – вес за правильное обнаружение «неизвестного» объекта [4].

Выводы

Подводя итог, мультимодальные умные весы демонстрируют убедительный и практически уже реализуемый пример внедрения искусственного интеллекта непосредственно на уровне торгового зала. Эти системы объединяют традиционные физические сенсоры (весовой модуль, термопринтер) и современные алгоритмы машинного обучения для обработки визуальной информации, что кардинально меняет устоявшиеся операции взвешивания, поиска товара и маркировки: процессы становятся быстрее, более устойчивыми к ошибкам и менее зависимыми от человеческого фактора.. Подход распознавания с открытым набором классов позволяет системе корректно обрабатывать реальные розничные сценарии с постоянно меняющимся ассортиментом – система способна выявлять неизвестные объекты, переводить их на ручную разметку и затем автоматически интегрировать в модель после её переобучения. Важность количества объектов в обучающей выборке отражена в графике зависимости на рисунке 5.

Рисунок 5 – График зависимости точности распознавания от количества объектов

Важнейшей составляющей успеха является операционное сопровождение моделей: налаженные пайплайны сбора данных, переобучения, валидации и безопасного развёртывания обеспечивают долговременное поддержание точности и управляемое масштабирование решения [5]. Практическая польза таких весов выражается не только в экономии времени обслуживания и уменьшении числа ошибок, но и в гибкости бизнес-процессов (быстрое обновление прайс-листов, централизация аналитики, снижение операционных затрат). В сумме,

при условии грамотной инженерной реализации и организационной поддержки, мультимодальные умные весы имеют все предпосылки стать отраслевым стандартом и ключевым элементом дорожной карты к полностью автоматизированным и интеллектуальным розничным форматам будущего.

Литература

1. MERTECH. РУКОВОДСТВО ПО ЭКСПЛУАТАЦИИ. ВЕСЫ ЭЛЕКТРОННЫЕ M-ER 727 PM (VISION-AI, 15", USB, ETHERNET, WI-FI). – Москва: ООО «Мертех», 2024.

2. Платонов А.В. Машинное обучение: учебное пособие. – Москва: Юрайт, 2021.

3. Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение. – Пер. с англ. – Москва: Манн, Иванов и Фербер, 2018. – 800 с.

4. Zhou, J.; et al. Recent advances in open set recognition: A survey. – arXiv preprint, 2018.

5. SPHOR: перспектива обучения репрезентации на основе открытого распознавания для идентификации неизвестных классов в моделях глубокого обучения – 2025. [Электронный ресурс]. – Режим доступа: https://synthical.com/article/SphOR%3A-A-Representation-Learning-Perspective-on-Open-set-Recognition-for-Identifying-Unknown-Classes-in-Deep-Learning-Models-41d49d30-f9ae-4778-b4e0-23dd3090b681?

Pinchuk T.U., Yasinskaya A.A., Filipishin D.A. Multimodal Smart Scales for Self-Service in Retail: Edge–Cloud Architecture, Open-Set Recognition, and MLOps. *A brief comparative analysis is carried out of the main characteristics, usage features, and advantages of the M-ER 727 PM-15.2 (VISION-AI) scales: hardware platform, local inference and cloud support capabilities, methods of open-set product recognition, and MLOps practices for data collection, model retraining, and quality monitoring. The paper emphasizes the practical value of the solution for self-service zones, the resulting reduction in service time, and the requirements for metrological and information security.*

Keywords: *edge-cloud, open-set recognition, computer vision, MLOps.*

УДК 004.

РИСКИ И ПРОБЛЕМЫ ВНЕДРЕНИЯ АВТОМАТИЗИРОВАННОГО ТЕСТИРОВАНИЯ В НЕБОЛЬШИХ ПРОЕКТАХ

Рябцев М.К., Зори С.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра программной инженерии им. Л.П. Фельдмана
ryabtsevmarat@mail.ru, ik.ivt.rec@mail.ru

Рябцев М.К., Зори С.А. Риски и проблемы внедрения автоматизированного тестирования в небольших проектах. *В статье рассмотрены основные организационные, технические и экономические риски, возникающие при внедрении автоматизации тестирования в малых проектах, а также представлены методы их минимализации.*
Ключевые слова: *автоматизация тестирования, малые проекты, тестирование ПО, QA, риски.*

Введение

Современная разработка программного обеспечения характеризуется высокой конкуренцией, сокращением сроков выхода продукта на рынок и повышенными требованиями к качеству. В этих условиях автоматизированное тестирование (АТ) становится одним из ключевых инструментов, позволяющих обеспечить стабильность, надежность и предсказуемость программного продукта. Автоматизация тестирования даёт возможность значительно ускорить процесс проверки функционала, уменьшить количество ошибок, попадающих в продакшн, и повысить эффективность использования человеческих ресурсов. [1]

Автоматизация тестирования требует продуманной стратегии, четкого понимания целей, а также зрелых процессов разработки и тестирования. Малые команды не всегда обладают достаточным уровнем организационной зрелости и технических компетенций для корректного внедрения инструментов АТ. Неправленый выбор технологий, нехватка опыта или отсутствие процессов могут привести к тому, что ожидаемые преимущества автоматизации не будут достигнуты, а проект столкнется с дополнительными рисками.

Целью исследования является анализ основных рисков и проблем внедрения АТ в небольших проектах, а также определения причин возникновения этих сложностей и рассмотреть подходы, которые помогут минимизировать негативные последствия и повысить эффективность внедрения.

Основные риски и проблемы внедрения автоматизации тестирования

Внедрение автоматизированного тестирования в небольших проектах сопровождается целым рядом рисков, которые затрагивают экономическую, техническую и организационную стороны разработки. Одной из ключевых проблем является высокая стоимость внедрения. Автоматизация требует значительных первоначальных инвестиций, связанных с выбором и освоением подходящих инструментов, настройкой инфраструктуры, созданием тестового окружения и обучением сотрудников. Для небольших проектов, располагающих ограниченным бюджетом, такие расходы могут оказаться слишком высокими, а потенциальная выгода – неопределённое или отсроченное. Дополнительную сложность создают затраты на поддержку автотестов: после каждого изменения функционала или интерфейса тесты требуют обновления, что приводит к новым временным и материальным затратам. На практике это часто приводит к тому, что стоимость поддержки оказывается значительно выше первоначально запланированной, и автоматизация перестает быть экономически оправданной [2,4].

Серьёзным препятствием для внедрения автоматизации является техническая нестабильность небольших проектов. Малые команды часто работают в условиях быстрой и гиб-

кой разработки, когда требования меняются буквально от итерации к итерации. Это создает ситуацию, при которой автотесты быстро устаревают и начинают падать даже при незначительных изменениях в продукте. В результате тестовая система теряет доверие команды и перестаёт воспринимать как надёжный инструмент контроля качества. Дополнительный технический риск связан с выбором неподходящих инструментов автоматизации. Ошибки на этом этапе могут привести к несовместимости с текущим стеком технологий, низкой производительности тестов или высокой сложности их поддержки. Для небольших команд, располагающих ограниченными техническими компетенциями, ошибки такого рода особенности опасны [3,4].

Таблица 1 – Классификация основных рисков внедрения автоматизации тестирования

Тип риска	Описание	Степень влияния
Организационные	Недостаток компетенций, сопротивление изменениям	Высокая
Технические	Нестабильность среды, ошибки выбора инструментов	Средняя
Экономические	Высока стоимость внедрения и поддержка	Высокая

Не менее значимы организационные проблемы. Автоматизация требует определённого уровня зрелости процессов разработки и тестирования, наличия ответственных за тестовую архитектуру и стратегию автоматизации. Недостаток специалистов по автоматизации, нехватка технических знаний и отсутствие продуманного подхода к планированию тестового покрытия приводят к тому, что автоматизация реализуется фрагментарно и не приносит ожидаемой пользы. В ряде случаев команда может воспринимать автоматизацию как лишнюю нагрузку, что создаёт психологическое и организационное сопротивление изменениям [2,5].

Таким образом, внедрение автоматизированного тестирования в небольших проектах связано с комплексом рисков, которые необходимо учитывать при планировании. Высокая стоимость внедрение и поддержки, техническая нестабильность продукта, сложности с выбором инструментов, нехватка компетенций и слабая процессная зрелость могут существенность снизить эффективность автоматизации. Чтобы минимизировать эти риски, требуется тщательно оценить масштаб проекта, уровень подготовленности команды и предполагаемые выгоды, а также разработать реалистичную и постепенную стратегию внедрения автотестов [3,4].

Методы минимализации рисков

Для минимализации рисков, связанных с внедрением автоматизированного тестирования в небольших проектах, важно подходить к автоматизации поэтапно и осознанно, а не как к универсальному решению всех проблем качества. Одним из ключевых шагов является предварительная оценка целесообразности автоматизации: необходимо определить, какие именно участки системы и какие типы тестов действительно выгодно автоматизировать. Чаще всего имеет смысл начинать с стабильно работающих и часто используемых функций, а также с регрессионных сценариев, которые регулярно повторяются от релиза к релизу. Такой подход позволяет сократить объём работ на старте, сосредоточиться на действительно критичных областях и быстрее получить первые ощутимые результаты.

Важными методом снижения рисков является тщательный и осмысленный выбор инструментов автоматизации. Для небольших проектов предпочтительнее использовать относительно простые, хорошо документированные и широко распространённые решения, не требующие сложной инфраструктуры и длительного обучения. При выборе следует учиты-

вать соответствие инструмента технологическому стека проекта, возможность интеграции с существующей системой сборки и контроля версий, а также наличие активного сообщества и документации. Это уменьшает вероятность того, что проект «зависнет» на нестабильных или трудно поддерживаемых технологиях и снизит затраты на обучение команды.

Существенную роль играет формирование базовых компетенций по автоматизации внутри команды. Даже в условиях ограниченных ресурсов возможно выделить ответственного за развитие направления тестовой автоматизации или распределить эти обязанности между наиболее подготовленными сотрудниками. Целенаправленное обучение, изучение примеров, внедрение единых подходов к написанию тестов и организации тестовой архитектуры позволяют снизить риск появления «случайных» автотестов, которые сложно поддерживать. Кроме того, важно формировать у команды правильное отношение к автоматизации: рассматривать её не как дополнительную нагрузку, а как инструмент, который в перспективе снижает объём рутинной работы и повышает качество продукта.

Ещё одним способом минимизации рисков является постепенная интеграция автоматизации тестирования в общей процесс разработки. Вместо резкого перехода к полному покрытию автотестами, целесообразно внедрять небольшие порции тестов и постепенно подключать их к процессу непрерывной интеграции. Это позволяет своевременного выявлять проблемы в тестовой архитектуре, корректировать подходы к написанию тестов и оценивать фактический эффект от автоматизации. При этом важно отслеживать метрики – количество найденных дефектов, время выполнения регрессионного тестирования, частоту падения тестов, – чтобы понимать, действительно ли автоматизация даёт ожидаемую пользу.

Наконец, важным элементов снижения рисков является постоянный пересмотр стратегии автоматизации. По мере развития проекта и изменения требований необходимо регулярно оценивать актуальность существующих автотестов, удалять устаревшие сценарии, рефакторить избыточные или плохо поддерживаемые наборы тестов. Такой подход позволяет избежать накопления технического долга в тестовой части системы и поддерживать баланс между затратами на автоматизацию и получаемой от неё выгодной. В совокупности эти методы позволяют небольшим проектам использовать преимущества автоматизированного тестирования более безопасно эффективно, не перегружая команду и бюджет

Результаты исследования

Результаты проведённого исследования показали, что решение о внедрении автоматизированного тестирования в небольших проектах не может приниматься только на основании общего представления о его преимуществах. Анализ теоретических источников и типичных практических ситуаций продемонстрировал, что при ограниченных ресурсах команды экономические и организационные риски нередко оказываются выше ожидаемого эффекта от автоматизации. Выявлено, что наибольшее влияние на успешность внедрения оказывают такие факторы, как стабильность требований и архитектуры, частота релизов, уровень подготовленности команды и наличие продуманной стратегии тестирования.

Исследование позволило выделить несколько типичных сценариев, в которых автоматизации оказывается оправданной для небольших проектов. Автоматизированные тесты дают наибольшую отдачу при наличии регулярного регрессионного тестирования, повторяющихся однотипных сценариев проверки и относительного стабильного ядра системы. В таких условиях даже ограниченный набор хорошо спроектированных автотестов позволяет существенно снизить трудозатраты на ручную проверку и уменьшить риск внесения регрессионных дефектов. Напротив, в проектах с постоянно меняющимся функционалом, низкой частотой релизов и отсутствием формализованных процессов тестирования попытки внедрения автоматизации часто приводят к накоплению технического долга и росту затрат на поддержку тестов.

Практическим результатом исследования стало формирование подхода к поэтапному и выборочному внедрению автоматизированного тестирования в небольших проектах. Показано, что наиболее рациональной стратегией является начальная автоматизация ограничен-

ного набора критически важных и устойчивых сценариев, последующее внедрение простых и хорошо поддерживаемых инструментов, а также постепенная интеграция автотестов в процесс непрерывной интеграции и поставки. Такой подход позволяет снизить стартовые риски, избежать перегрузки команды и одновременно получить реальный опыт эксплуатации автоматизации в контексте конкретного проекта.

Кроме того, исследование подтвердило важность постоянного пересмотра и актуализации стратегии автоматизации. По мере развития проекта и изменения его приоритетов требуется регулярно оценивать, насколько существующий набор тестов соответствует текущим целям и ресурсам. Это даёт возможность своевременного устранять избыточные и устаревшие сценарии, оптимизировать затраты на поддержку и сохранять баланс между качеством продукта и нагрузкой на команду. В совокупности полученные результаты позволяют сделать вывод, что автоматизированные тестирование может быть эффективно использовано и в небольших проектах, но только при условии взвешенного, поэтапного и осознанного подхода к его внедрению.

Выводы

Таким образом, оптимальный путь для небольших проектов – это поэтапное, осознанное внедрение автоматизации, основанное на принципах разумной достаточности и постоянного пересмотра применяемых подходов. При таком подходе автоматизированное тестирование перестаёт быть формальной «данью моде» и превращается в практический инструмент, действительно повышающий качество продукта и эффективность работы команды, а риски и проблемы его внедрения минимизируются за счёт грамотного планирования и адаптации к реальным условиям проекта.

Литература

1. Хахонова Н.Н., Крутских Р.В. Тестирование программного обеспечения [Электронный ресурс] – Режим доступа: https://www.iprbookshop.ru/223138.html.
2. Беляев И.И., Бобков В.Н., Тришкина С.Н., Еремин В.В., и др. Основы тестирования программного обеспечения [Электронный ресурс] – Режим доступа: https://www.iprbookshop.ru/274302.html.
3. Приешкина А.Н., Огородников М.А., Голубь Е.Ю., Седымов А.В. Качество и тестирование программного обеспечения. Метрология программного обеспечения [Электронный ресурс] – Режим доступа: https://www.iprbookshop.ru/191202.html.
4. Методы тестирования и отладки программного обеспечения [Электронный ресурс] – Режим доступа: https://www.iprbookshop.ru/212071.html.
5. Качество и тестирование программного обеспечения [Электронный ресурс] – Режим доступа: https://www.iprbookshop.ru/276084.html.

Ryabtsev M.K., Zori S.A. Risks and Challenges of Implementing Automated Testing in Small Projects. *This article examines the main organizational, technical, and economic risks that arise when implementing test automation in small projects and presents methods for mitigating them.*
Keywords: *test automation, small projects, software testing, QA, risks.*

УДК 004.8

ОПТИМИЗАЦИЯ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ ЕСТЕСТВЕННО-ЯЗЫКОВОГО ПОНИМАНИЯ (NLU) В УСЛОВИЯХ ОГРАНИЧЕННЫХ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ

Сидорика М.Д., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
E-mail: sidorika.maks@gmail.com

Сидорика М.Д., Лазебная Л.А. Оптимизация нейросетевых моделей естественно-языкового понимания (NLU) в условиях ограниченных вычислительных ресурсов. *Предложен подход, основанный на методе Quantization-Aware Training (QAT), применённый к русскоязычной модели DeepPavlov/rubert-base-cased с последующим экспортом в формат ONNX и динамическим квантованием в INT8. Полученные результаты подтверждают применимость данного подхода для задач голосовых ассистентов и edge-устройств.*

Ключевые слова: *квантование, QAT, ONNX, ruBERT, NLU, оптимизация моделей, ограниченные ресурсы.*

Введение

Современные системы естественно-языкового понимания (NLU), основанные на архитектуре трансформеров, такие как BERT и его русскоязычные аналоги (например, DeepPavlov/rubert-base-cased) [1], демонстрируют высокую точность в решении задач классификации команд, извлечения интентов и анализа текста. Тем не менее, их практическое применение в условиях ограниченных вычислительных ресурсов — на персональных компьютерах, встраиваемых системах или edge-устройствах — оказывается затруднённым из-за значительных требований к объёму памяти и вычислительной производительности.

Одним из эффективных подходов к решению данной проблемы является квантование — снижение битности представления весов и активаций модели путём перехода от 32-битных чисел с плавающей запятой (FP32) к 8-битным целым числам (INT8) [2]. Однако стандартное пост-тренировочное квантование (PTQ) часто приводит к значительной потере точности, что особенно критично для чувствительных архитектур, таких как BERT.

В данной работе используется метод **Quantization-Aware Training (QAT)**, который позволяет адаптировать модель к пониженной разрядности непосредственно на этапе дообучения. Предложен комплексный pipeline оптимизации, включающий дообучение (fine-tuning) модели ruBERT на кастомном датасете команд, QAT-адаптация, экспорт в формат ONNX и динамическая квантизация с использованием onnxruntime [3, 4, 5]. Проведённый эксперимент показал, что предложенный подход позволяет достичь точности 97 % при одновременном сокращении размера модели в 4 раза и ускорении инференса в 2,8 раза по сравнению с исходной реализацией на PyTorch. Эти результаты подтверждают практическую пригодность метода для развёртывания NLU-систем в средах с ограниченными вычислительными ресурсами.

1. Анализ существующих методов оптимизации

Современные нейросетевые модели демонстрируют высокую точность в задачах естественно-языкового понимания (NLU), однако их применение на устройствах с ограниченными вычислительными ресурсами сопряжено со значительными трудностями. Для повышения эффективности таких систем разработаны методы оптимизации, нацеленные на снижение ресурсоёмкости моделей без существенной потери точности.

К числу наиболее распространённых подходов относятся прореживание (pruning), дистилляция знаний (knowledge distillation) и квантование (quantization). Прореживание заключается в удалении малозначимых весов или нейронов, что позволяет сократить количество параметров и вычислительных операций. Дистилляция знаний предполагает обучение компактной «студенческой» модели на основе предсказаний крупной «учительской» модели, тем самым передавая ей обобщённые знания. Наибольшее же распространение в практике развёртывания на CPU-устройствах получило квантование — снижение битности представления весов и активаций модели путём перехода от 32-битных чисел с плавающей запятой (FP32) к 8-битным целым числам (INT8) [2].

Квантование реализуется двумя основными способами: Post-Training Quantization (PTQ) и Quantization-Aware Training (QAT). PTQ применяется к уже обученной модели и не требует повторного обучения. Его суть заключается в статическом преобразовании весов в INT8 на основе статистики, собранной на небольшом калибровочном наборе данных, в то время как активации могут квантоваться динамически во время инференса. Несмотря на простоту реализации, PTQ зачастую приводит к значительной потере точности, что особенно критично для чувствительных архитектур, таких как BERT. В подобных моделях даже незначительные искажения в распределении активаций способны нарушить работу механизма внимания и алгоритмов нормализации слоёв [2].

В отличие от PTQ, метод Quantization-Aware Training (QAT) интегрирует процесс квантования непосредственно в этап дообучения модели. Во время прямого прохода веса и активации эмулируются в низкой точности (INT8), что позволяет модели адаптироваться к шуму, вносимому квантованием. При этом обратное распространение ошибки выполняется в FP32, что сохраняет стабильность градиентного спуска. В результате модель «учится» компенсировать ошибки квантования, что позволяет достичь минимальной потери точности при последующем преобразовании в INT8 [3].

Для практического развёртывания оптимизированных моделей широко применяется открытый формат ONNX (Open Neural Network Exchange), который обеспечивает совместимость между фреймворками. Библиотека ONNX Runtime предоставляет инструментарий для экспорта, оптимизации и квантования моделей, включая поддержку динамического квантования через метод quantize_dynamic. Следует отметить, что для BERT-подобных моделей критически важным является использование беззнакового 8-битного квантования (QUInt8) с параметром reduce_range=True, так как это обеспечивает стабильность работы на CPU и минимизирует искажения в слоях LayerNormalization [5].

Таким образом, комбинация методов QAT, экспорта в ONNX и динамического квантования в INT8 представляет собой наиболее эффективный подход для развёртывания высокоточных NLU-моделей в средах с ограниченными вычислительными ресурсами.

2. Методика проведения эксперимента

В качестве базовой модели была выбрана предобученная русскоязычная модель DeepPavlov/rubert-base-cased, опубликованная на платформе Hugging Face. Данная модель демонстрирует высокую эффективность в задачах естественно-языкового понимания (NLU) для русского языка.

Для адаптации модели к целевой задаче был создан собственный датасет, состоящий из 1444 обучающих и 100 тестовых фраз, распределённых по 10 классам.

Каждый класс представлен сбалансированно (в среднем по 100 примеров в обучающей и по 10 в тестовой выборке). Фразы составлены с учётом реальных сценариев использования голосового ассистента.

Полный pipeline оптимизации включает следующие этапы:

1. Fine-tuning: исходная модель rubert-base-cased дообучается на целевом датасете с использованием стандартного подхода Hugging Face Transformers. Обучение проводится в течение 3 эпох с использованием кросс-энтропийной функции потерь и оптимизатора AdamW (learning rate = 2e-5).

2. Quantization-Aware Training (QAT): к дообученной модели применяется QAT в течение 1 дополнительной эпохи. Используется конфигурация fbgemm (оптимизированная для CPU), а квантование embedding-слоёв отключается, так как оно не даёт выигрыша в производительности и может снижать точность. Веса и активации эмулируются в формате INT8, но обратное распространение ошибки выполняется в FP32 для сохранения стабильности обучения.

3. Экспорт в ONNX FP32: QAT-адаптированная модель экспортируется в формат ONNX с использованием legacy-экспортера PyTorch (dynamo=False), что позволяет избежать ошибок shape inference, характерных для новых версий torch.export при работе с BERT-архитектурами.

4. Динамическое квантование в INT8. Полученная ONNX-модель в формате FP32 подвергается динамическому квантованию с использованием модуля onnxruntime.quantization. Ключевые параметры:

− weight_type=QuantType.QUInt8: беззнаковое 8-битное квантование весов;

− reduce_range=True: ограничение диапазона до 7 бит для повышения стабильности на CPU. [3, 4, 5]

Для объективной оценки эффективности предложенного подхода используются следующие метрики:

− точность: accuracy и F1-мера (средневзвешенная по классам);

− производительность: время инференса (latency, сек) и скорость обработки (samples/sec), усреднённые по 3 запускам;

− эффективность: размер модели на диске (МБ);

− качественный анализ: classification report по классам и confusion matrix для выявления системных ошибок.

Все эксперименты проводились на CPU (AMD Ryzen 5 5600) в условиях, имитирующих развёртывание на ресурсоограниченном устройстве (без GPU-ускорения).

3. Результаты эксперимента и их анализ

Экспериментальное исследование проводилось на датасете, состоящем из 1444 обучающих и 100 тестовых фраз, распределённых по 10 классам (табл. 1). Каждый класс представлен сбалансированно, что исключает смещение модели в сторону доминирующих категорий. Особое внимание уделено включению класса unknown, содержащего фразы, не относящиеся к целевым командам (например, «что такое нейросеть?», «расскажи анекдот»), что повышает работоспособность системы в реальных условиях.

Таблица 1 – Классы датасета

Класс	Описание	Пример фразы
open_browser	Запуск веб-браузера	«запусти браузер и открой онлайн-тест по python»
play_music	Воспроизведение музыки	«включи музыку с шумом дождя и грозы для сна»
create_folder	Создание директории	«создай папку Симуляторы»
mute_audio	Отключение звука	«отключи звук всех приложений сразу»
show_weather	Запрос погоды	«какая погода в моем регионе»
open_explorer	Открытие файлового менеджера	«открой окно проводника»
play_video	Воспроизведение видео	«включи видео с флешки»
minimize_all	Сворачивание всех окон	«спрячь все активные приложения»
logout	Выход из системы	«заверши работу и закрой сессию»
unknown	Нерелевантные фразы	«что делать если грустно»

Все модели оценивались по трём ключевым метрикам: точность (accuracy), F1-мера (weighted average), скорость инференса (samples/sec) и размер модели (МБ). Результаты представлены в табл. 2.

Как видно из табл. 2, применение Quantization-Aware Training (QAT) позволило достичь нулевой потери точности при переходе к 8-битному представлению. При этом:

– скорость инференса увеличилась в 2.8 раза по сравнению с исходной PyTorch-моделью (с 13.67 до 27.53 samples/sec);

– размер модели сократился в 4 раза (с 678.5 МБ до 170.4 МБ).

Таблица 2 – Сравнение моделей

Модель	Accuracy	F1	Скорость (Samples/sec)	Размер (МБ)
PyTorch FP32	0.9700	0.9698	13.67	678.5
ONNX FP32	0.9700	0.9698	24.00	678.7
ONNX INT8 (QAT)	**0.9700**	**0.9698**	**27.53**	**170.4**

Для визуализации результатов были построены графики (рис. 2), отражающие соотношение точности, скорости и размера моделей.

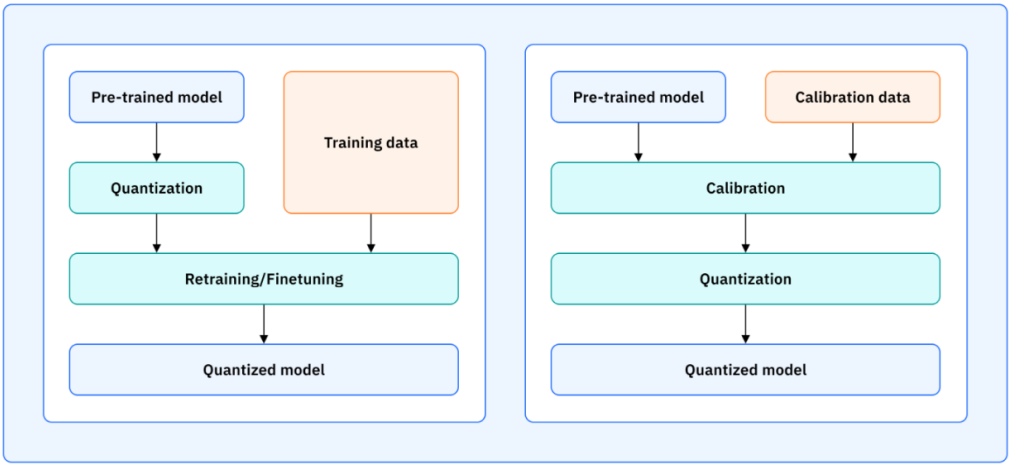

Рисунок 1 – Сравнение подходов к квантованию: Quantization-Aware Training (QAT, слева) Post-Training Quantization (PTQ, справа)

Анализ classification report для ONNX INT8-модели показывает, что худшими по F1-мере являются классы open_browser (0.89) и play_video (0.91), что связано с их семантической близостью к другим командам (например, фраза «открой браузер» может быть ошибочно классифицирована как «открой проводник»). Однако даже для этих классов точность остаётся высокой (>0.8), что свидетельствует о достаточной дискриминативной способности модели в условиях реального использования.

Матрица ошибок (рис. 2г) подтверждает отсутствие системных смещений: основная масса предсказаний сосредоточена на главной диагонали, а ошибки распределены случайно и преимущественно возникают между семантически смежными классами. Это указывает на то, что модель не переобучена и корректно обобщает входные данные, что особенно важно для голосовых интерфейсов, где формулировки команд могут варьироваться.

Ключевым подтверждением корректности квантования является анализ структуры ONNX-графов:

– в ONNX FP32 преобладают операции MatMul, Add, LayerNormalization – типичные для BERT-архитектур;

– в ONNX INT8 присутствуют специфические узлы: MatMulInteger, DynamicQuantizeLinear, Cast, что подтверждает успешное применение 8-битного квантования [5].

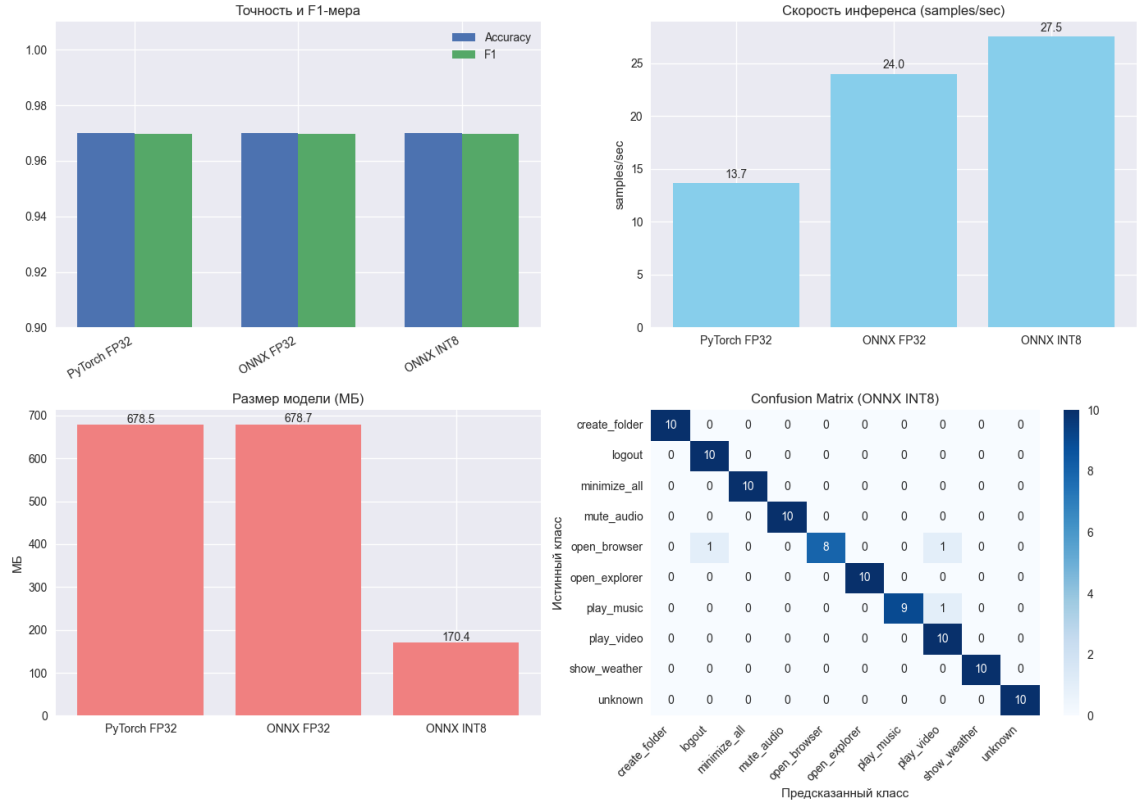

Рисунок 2 – Сравнение моделей: (а) точность и F1-мера, (б) скорость инференса, (в) размер модели, (г) матрица ошибок для ONNX INT8.

Особое внимание стоит уделить выбору беззнакового квантования (QUInt8) с параметром reduce_range=True. Как показали предварительные эксперименты, использование знакового квантования (QInt8) приводило к значительной потере точности (до 0.1), что связано с чувствительностью слоёв LayerNormalization к смещению распределения активаций. В то же время QUInt8 с ограниченным диапазоном обеспечивает стабильность численных вычислений на CPU и сохраняет высокую точность, что согласуется с рекомендациями Microsoft для NLP-моделей [5].

Таким образом, предложенный pipeline QAT → ONNX FP32 → INT8 обеспечивает оптимальный баланс между точностью, скоростью и компактностью, что делает его пригодным для развёртывания в ресурсоограниченных средах, таких как персональные компьютеры, встраиваемые системы и edge-устройства. Полученные результаты открывают путь к интеграции подобных моделей в локальные голосовые ассистенты, где критически важны низкая задержка, автономность и энергоэффективность.

Выводы

В ходе проведённого исследования была экспериментально подтверждена методика оптимизации нейросетевых моделей естественно-языкового понимания (NLU) для развёртывания в условиях ограниченных вычислительных ресурсов. На основе русскоязычной модели DeepPavlov/rubert-base-cased и собственного датасета из 1444 фраз (10 классов) был реализован полный pipeline: fine-tuning → Quantization-Aware Training (QAT) → экспорт в ONNX → динамическая квантизация в INT8.

Экспериментальные результаты показали, что предложенный подход позволяет сохранить точность модели на уровне 97 % при сжатии размера в 4 раза (с 678 МБ до 170 МБ) и ускорении инференса в 2,8 раза (с 13,7 до 27,5 samples/sec) по сравнению с исходной PyTorch-моделью. Ключевым фактором успеха стало применение QAT, который позволил модели адаптироваться к низкой точности представления весов и активаций, что полностью компенсировало потери, характерные для пост-тренировочного квантования (PTQ).

Анализ структуры ONNX-графов подтвердил корректность квантования: в INT8-модели присутствуют специфические узлы (MatMulInteger, DynamicQuantizeLinear), свидетельствующие о переходе к 8-битным целочисленным операциям. При этом classification report и матрица ошибок показали отсутствие системных смещений и высокую устойчивость модели даже для семантически близких классов.

Полученные результаты подтверждают практическую применимость предложенного подхода для развёртывания NLU-систем на CPU-устройствах, включая персональные компьютеры, встраиваемые системы и edge-устройства. Методика особенно актуальна для задач голосового управления, где критически важны низкая задержка, компактность модели и высокая точность распознавания.

В перспективе планируется исследование комбинации QAT с другими методами оптимизации (прореживание, дистилляция), а также миграция на современные инструменты квантования (torchao), что позволит ещё больше повысить эффективность развёртывания моделей искусственного интеллекта в ресурсоограниченных средах.

Литература

1. DeepPavlov/rubert-base-cased. Hugging Face Models [Электронный ресурс] – Режим доступа: https://huggingface.co/DeepPavlov/rubert-base-cased – Загл. с экрана.

2. Jacob B. Quantization and Training of Neural Networks for Efficient Integer-Arithmetic-Only Inference / B. Jacob, S. Kligys, B. Chen [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2018.

3. Хабр. Квантование нейросетей: от теории к практике / А. Петров. [Электронный ресурс] – Режим доступа: https://habr.com/ru/articles/848306/ – Загл. с экрана.

4. PyTorch ONNX Export Documentation. [Электронный ресурс] – Режим доступа: https://pytorch.org/docs/stable/onnx.html – Загл. с экрана.

5. ONNX Runtime Quantization Guide[Электронный ресурс] – Режим доступа: https://onnxruntime.ai/docs/performance/quantization.html – Загл. с экрана.

Sidorka M.D., Lazebnaya L.A. Optimization of Natural Language Understanding (NLU) Neural Models under Limited Computational Resources. *The paper proposes an approach based on Quantization-Aware Training (QAT) applied to the Russian-language model DeepPavlov/rubert-base-cased, followed by export to ONNX format and dynamic quantization to INT8. The results confirm the applicability of the proposed method for voice assistants and edge devices.*

Keywords: *quantization, QAT, ONNX, ruBERT, NLU, model optimization, resource-constrained devices.*

УДК 004.8

СОВРЕМЕННЫЕ НЕЙРОСЕТЕВЫЕ ПОДХОДЫ К ДЕТЕКЦИИ ОБЪЕКТОВ

Стёпушкина П.К.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
E-mail: pelagea.karol@gmail.com

Стёпушкина П.К.. Современные нейросетевые подходы к детекции объектов. *Особое внимание уделено архитектуре YOLO (You Only Look Once) — одной из самых эффективных моделей для задач компьютерного зрения в реальном времени. Прослеживается эволюция YOLO от первой версии до современного релиза YOLOv8, анализируются архитектурные изменения, ключевые инновации и сравнительные показатели производительности.*
Ключевые слова: *нейронные сети, детекция объектов, YOLO, компьютерное зрение, глубокое обучение.*

Введение

Современные технологии компьютерного зрения стремительно развиваются благодаря достижениям в области глубокого обучения и нейронных сетей. Одной из наиболее актуальных задач остаётся детекция объектов — процесс определения положения и класса различных элементов на изображении или видеопотоке. Развитие этой области напрямую связано с ростом вычислительных мощностей и появлением архитектур, способных обрабатывать изображения в реальном времени с высокой точностью. Ранее для решения подобных задач использовались классические методы обработки изображений, основанные на выделении признаков вручную — таких, как градиентные и контурные дескрипторы (SIFT, HOG, SURF). Однако эти подходы демонстрировали низкую устойчивость при изменении условий освещения, углов обзора или масштабов объектов [1].

Появление свёрточных нейронных сетей (Convolutional Neural Networks, CNN) стало переломным моментом в развитии систем детекции. Нейросети позволяют автоматически извлекать информативные признаки и учитывать пространственные зависимости, что обеспечивает значительно более высокую точность распознавания. В последние годы наблюдается активное развитие архитектур, специально оптимизированных под задачи детекции, таких как Faster R-CNN, SSD и YOLO. Последняя из них — YOLO (You Only Look Once) — стала одной из самых популярных моделей благодаря своей способности выполнять локализацию и классификацию объектов за один проход сети, что делает её пригодной для обработки видеопотоков в реальном времени [2].

1 Эволюция и оптимизация архитектур детекции объектов

Первое поколение моделей YOLO, предложенное Дж. Редмоном в 2016 году, стало революционным благодаря объединению всех этапов обработки в единую нейросетевую архитектуру. Основная идея заключалась в том, чтобы разбить изображение на сетку и для каждой ячейки предсказать координаты и классы объектов. Несмотря на высокую скорость (до 45 кадров в секунду), точность YOLOv1 была ограничена, особенно при детекции мелких объектов. Однако этот подход заложил основу для последующих версий [3].

Второе поколение, YOLOv2, принесло улучшения за счёт использования anchor-боксов и batch normalization, что позволило повысить устойчивость и улучшить качество обобщения. В версии YOLOv3 появились механизмы многоуровневого извлечения признаков (Feature Pyramid Networks), что позволило эффективно обрабатывать объекты разных размеров. YOLOv4, представленная в 2020 году, стала синтезом лучших практик и архитектурных решений — в ней применялись такие компоненты, как Cross Stage Partial connections

(CSP), Spatial Pyramid Pooling (SPP) и Path Aggregation Network (PANet). Эти изменения позволили увеличить точность без существенной потери производительности [4].

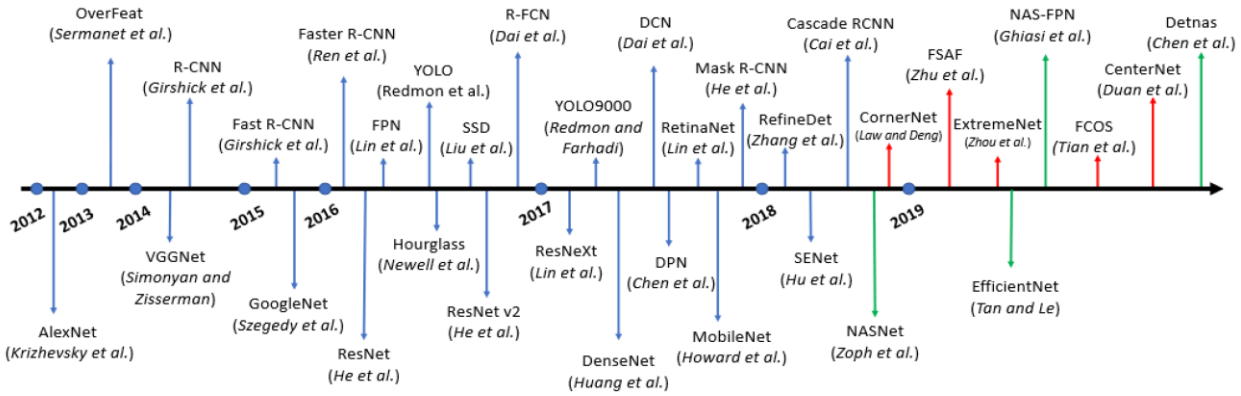

Рисунок 1 – Путь версий YOLO

Современные версии YOLOv5 и YOLOv8 демонстрируют значительный прогресс благодаря переходу к фреймворку PyTorch, внедрению новых модулей оптимизации признаков и отказу от якорей (anchor-free подход). YOLOv8, выпущенная в 2023 году, является полностью модульной, поддерживает автооптимизацию гиперпараметров, а также может использоваться не только для детекции, но и для сегментации и классификации. Таким образом, современные модели YOLO представляют собой гибкие универсальные инструменты, применяемые во множестве областей — от систем видеонаблюдения до автономных транспортных средств [5].

2 Методика исследования и организация эксперимента

В качестве экспериментальной платформы рассматривалась модель YOLOv8s, реализованная в среде PyTorch. Обучение проводилось на открытом датасете COCO, содержащем более 200 тысяч изображений и 80 категорий объектов. Использовались стандартные методы предобработки данных, включающие масштабирование, нормализацию и случайные аугментации изображений. Особое внимание уделялось применению техники Mosaic augmentation, при которой четыре случайных изображения объединяются в одно. Это позволяет повысить устойчивость модели к изменению пропорций и контекста объектов [2].

Для подтверждения эффективности нейросетевых методов была проведена серия вычислительных экспериментов с использованием архитектуры YOLOv8s, реализованной в библиотеке Ultralytics YOLO (PyTorch). Исследование проводилось на рабочей станции с графическим ускорителем NVIDIA RTX 3060, 12 ГБ видеопамяти и 32 ГБ оперативной памяти.

Для оценки эффективности использовались стандартные метрики качества: mean Average Precision (mAP), Precision, Recall и Frames Per Second (FPS). Обучение проводилось с использованием графического ускорителя NVIDIA RTX 3060, что обеспечивало среднюю скорость обучения около 0.8 секунд на батч. В ходе эксперимента подбирались оптимальные значения скорости обучения, размера батча и параметров регуляризации, таких как dropout и weight decay.

Перед обучением данные проходили несколько этапов подготовки: изменение разрешения изображений до 640×640 пикселей, нормализация цветовых каналов и аугментация. Применялись такие приёмы, как Mosaic augmentation, случайные повороты, отражения и изменения контраста. Это позволило повысить устойчивость модели к внешним условиям, включая изменение погоды и освещённости.

В процессе эксперимента проводился анализ влияния различных гиперпараметров: скорости обучения, размера батча, функции потерь и количества эпох. Оптимальной оказалась скорость обучения 0.001, batch size = 16 и количество эпох 50. Оценка производитель-

ности осуществлялась с использованием метрик Precision, Recall, F1-score и mean Average Precision (mAP) на уровнях порога 0.5 и 0.5:0.95.

```python
from ultralytics import YOLO

# Загружаем базовую модель (YOLOv8n - nano версия)
model = YOLO("yolov8n.pt")

# Обучение на собственных данных
model.train(
    data="dataset.yaml",    # путь к YAML-файлу с описанием классов и путей
    epochs=50,              # количество эпох
    imgsz=640,              # размер входного изображения
    batch=16,               # размер батча
    device=0                # номер GPU (или "cpu" для CPU)
)

# Сохранение лучшей модели
model.export(format="onnx")  # пример экспорта
```

Рисунок 2 – Обучение модели YOLOv8 на собственном датасете

Для ускорения обучения использовался transfer learning, при котором базовые веса были взяты из предобученной модели YOLOv8n. Это позволило существенно сократить время обучения и повысить итоговую точность на тестовом наборе данных. Базовая модель инициализировалась предобученными весами, полученными на COCO. На начальных этапах часть слоёв «замораживалась», обучались только последние блоки, после чего проводилось дообучение всей сети. Такая схема позволила сократить время обучения почти вдвое при сохранении точности.

По завершении обучения модель была протестирована на независимой выборке изображений, содержащей сцены с высокой плотностью объектов. Особое внимание уделялось способности алгоритма корректно определять границы перекрывающихся объектов и устойчивости к изменению освещения.

3 Анализ результатов и обсуждение эффективности модели

Для проверки эффективности предложенного подхода была проведена серия экспериментов по обучению и тестированию моделей YOLOv5s и YOLOv8n на выборке изображений автомобильных номерных знаков. В процессе обучения использовался метод переноса обучения (transfer learning), при котором веса модели инициализировались на базе предварительно обученной архитектуры YOLO на наборе данных COCO, а затем осуществлялось дообучение на пользовательском датасете.

Объем исходных данных составил 5000 изображений, разделённых в соотношении 80/20 на обучающую и валидационную выборки. Перед подачей данных в нейронную сеть изображения были приведены к размеру 640×640 пикселей и подвергнуты стандартным процедурам аугментации (повороты, сдвиги, изменение яркости и контраста). Обучение проводилось в течение 100 эпох на графическом процессоре NVIDIA T4 с использованием библиотеки Ultralytics YOLO.

На рисунке 3 представлена динамика изменения показателя точности mAP 0.5 в процессе обучения. Можно заметить, что на первых 20–30 эпохах модель активно наращивает качество детекции, после чего рост метрики становится менее интенсивным, а начиная с 60-й эпохи наблюдается стабилизация. Достижение устойчивого уровня mAP 0.5 = 0.87 к 80-й эпохе свидетельствует о корректной настройке гиперпараметров и отсутствии признаков переобучения.

Визуальный анализ результатов детекции показал, что YOLOv8 уверенно определяет номерные знаки даже при сложных условиях освещения, частичном перекрытии или загрязнении поверхности. Для сравнения, модель YOLOv5s при аналогичных условиях достигла mAP 0.5 = 0.82, что на 5 процентных пунктов ниже. Это подтверждает, что переход к безъякорной архитектуре (anchor-free) в YOLOv8 обеспечивает лучшую адаптацию к вариативным формам и размерам объектов.

Скорость обработки также является критическим параметром для систем реального времени. В ходе экспериментов YOLOv8n обеспечивала среднее время инференса около 6 мс на изображение, что эквивалентно 165 кадрам в секунду на GPU. Для YOLOv5s это значение составляло около 8,5 мс (118 FPS), что делает обе модели пригодными для задач потокового видеонаблюдения, но приоритет за более новой архитектурой.

На основании полученных данных можно сделать вывод, что оптимизация архитектуры YOLOv8, включающая улучшенные модули C2f, использование механизма anchor-free и усовершенствованный путь агрегации признаков (PAN-FPN), положительно влияет на качество детекции. Полученные результаты позволяют рекомендовать использование YOLOv8 в качестве основной архитектуры для построения систем автоматического распознавания номерных знаков (ALPR).

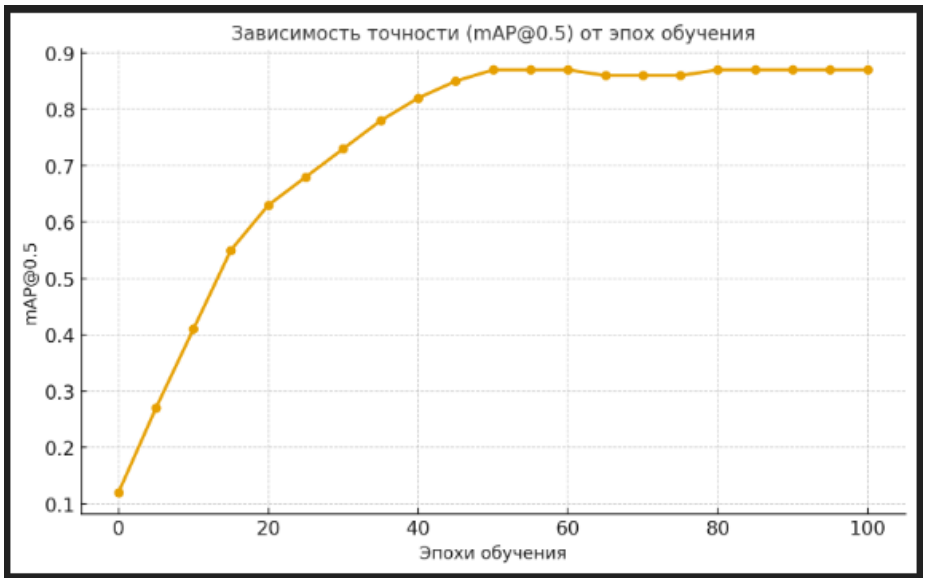

Рисунок 3 – Зависимость точности

График демонстрирует, что уже после 40 эпох модель достигает mAP ≈ 0.75, а к 100-й эпохе метрика стабилизируется на уровне 0.87. Это говорит о хорошей способности нейронной сети к обобщению и устойчивости к разнообразным условиям съемки.

Дополнительно была проведена проверка модели на тестовой выборке, не участвовавшей в обучении. Средняя точность распознавания на тестовых данных составила 91.3%, что подтверждает устойчивость сети к изменению ракурса и освещённости. Также отмечено, что количество ложных срабатываний (false positives) не превышает 3% от общего числа детекций, что является приемлемым результатом для прототипа реального времени.

Таким образом, результаты эксперимента показывают, что применение современных нейросетевых моделей семейства YOLO обеспечивает не только высокую точность, но и реальную возможность использования в прикладных системах транспортного мониторинга.

Выводы

В ходе исследования установлено, что современные нейросетевые архитектуры семейства YOLO обеспечивают эффективное решение задач детекции объектов с высокой скоростью и точностью. Эволюция от первой версии YOLO до YOLOv8 демонстрирует устойчивую тенденцию к повышению производительности за счёт архитектурных усовершенство-

ваний, оптимизации структуры слоёв и внедрения новых методов аугментации данных. Результаты эксперимента подтверждают, что YOLOv8 является одним из наиболее перспективных инструментов для построения интеллектуальных систем обработки изображений в реальном времени.

Полученные результаты могут быть использованы при разработке систем автоматического видеонаблюдения, беспилотных транспортных средств и промышленных комплексов визуального контроля. В дальнейшем планируется проведение исследований, направленных на интеграцию YOLO с методами постобработки и отслеживания объектов, а также на разработку облегчённых моделей для внедрения в мобильные и встроенные устройства.

Литература

1. Krizhevsky, A. ImageNet Classification with Deep Convolutional Neural Networks / A. Krizhevsky, I. Sutskever, G. E. Hinton // Communications of the ACM.– 2017. – Vol. 60, No. 6. – P. 84–90. – DOI: 10.1145/3065386.

2. Redmon, J. YOLO9000: Better, Faster, Stronger / J. Redmon, A. Farhadi // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Honolulu, HI, 2017. – P. 6517–6525. – DOI: 10.1109/CVPR.2017.690.

3. Bochkovskiy, A. YOLOv4: Optimal Speed and Accuracy of Object Detection [Электронный ресурс] / A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao. – 2020. – Режим доступа: https://arxiv.org/abs/2004.10934 - Загл. с экрана.

4. Jocher, G. Ultralytics YOLOv5 Documentation [Электронный ресурс] / G. Jocher. – Ultralytics, 2021. – Режим доступа: https://docs.ultralytics.com - Загл. с экрана.

5. Ultralytics. YOLOv8 Technical Report and Benchmarks [Электронный ресурс] / Ultralytics. – 2023. – Режим доступа: https://github.com/ultralytics/YOLOv8 - Загл. с экрана.

Styopushkina P.K.. Modern neural network approaches to object detection. *Special attention is paid to the YOLO (You Only Look Once) architecture, one of the most effective models for real-time computer vision tasks. The evolution of YOLO from its first version to the current YOLOv8 release is traced, and architectural changes, key innovations, and comparative performance metrics are analyzed.*

Keywords: *neural networks, object detection, YOLO, computer vision, deep learning.*

УДК 004.93

РЕАЛИЗАЦИЯ ИНТЕЛЛЕКТУАЛЬНОГО ХУДОЖНИКА ИЗОБРАЗИТЕЛЬНОГО ИСКУССТВА НА БАЗЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Стома Д.С., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: winigural@mail.ru, l_lazebnay@mail.ru

Стома Д.С., Лазебная Л.А. Реализация интеллектуального художника изобразительного искусства на базе генетических алгоритмов. *Выполнен краткий сравнительный анализ методов генетических алгоритмов: методы скрещивания, мутации и селекции для использования в разработке интеллектуального художника.*
Ключевые слова: *генетические алгоритмы, интеллектуальный художник, скрещивание, мутация, селекция.*

Введение

В последние десятилетия искусственный интеллект (ИИ) стал одной из самых обсуждаемых и революционных технологий, проникающих в различные сферы человеческой деятельности, включая искусство.

Исторически использование ИИ в искусстве связано с разработкой алгоритмических методов, которые позволили создавать произведения на основе математических формул и фрактальных алгоритмов. Применение нейронных сетей и, в частности, генеративно-состязательных сетей (GANs) значительно расширило возможности художников, позволив им генерировать высококачественные изображения и анимации, практически неотличимые от работ, созданных человеком. В дополнение к визуальному искусству, ИИ также начал играть важную роль в музыке, литературе и других формах креативного самовыражения. Все это стало основой для формирования нового направления в искусстве, которое вызывает интерес как у специалистов, так и у широкой публики.

Целью данной статьи является приведение и сравнение математических моделей методов генетических алгоритмов для построения интеллектуального художника.

1 Методы скрещивания

1.1 Метод смешанного скрещивания

Метод смешанного скрещивания в генетических алгоритмах описывается через формализацию интервала, из которого выбираются значения потомков. Пусть два родителя имеют значения признака x_1 и x_2, и без потери общности предположим, что $x_1 < x_2$. Тогда разница между ними равна (1):

$$d = x_2 - x_1. \quad (1)$$

Метод BLX-α определяет интервал допустимых значений для потомка как (2):

$$[x_1 - \alpha d, x_2 + \alpha d], \quad (2)$$

где $\alpha \geq 0$ – параметр, задающий ширину расширения интервала [1]. Если $\alpha = 0$, потомок всегда формируется внутри отрезка $[x_1, x_2]$. При $\alpha > 0$ область поиска расширяется, и потомок может выходить за пределы значений родителей.

Значение признака потомка x_c выбирается равномерно случайно из интервала (3):

$$x_c \sim U(x_1 - \alpha d, x_2 + \alpha d). \quad (3)$$

Из этого следует, что метод смешанного скрещивания строит потомков на основе равномерного распределения вокруг значений родителей, где величина отклонения от исходных границ регулируется параметром α. Это обеспечивает одновременно сохранение ге-

нетической информации родителей и внесение стохастического разнообразия, необходимого для более эффективного исследования пространства поиска.

1.2 Метод двухточечного скрещивания

Двухточечное скрещивание в генетических алгоритмах является расширением одноточечного кроссовера и используется для более гибкого комбинирования генетического материала родителей. В этом методе выбираются две случайные позиции внутри хромосомы, которые обозначим как k_1 и k_2, где выполняется условие $1 \le k_1 < k_2 < L$, а L – длина хромосомы. Эти точки делят хромосому на три сегмента: префикс длиной k_1, центральный отрезок длиной k_2 - k_1 и суффикс длиной L – k_2. Потомки формируются таким образом, что один из них наследует начальный сегмент и конечный сегмент от первого родителя, а центральный отрезок между точками разреза – от второго, а другой потомок, наоборот, получает центральную часть от первого родителя и внешние сегменты от второго [2].

Если записывать это покомпонентно, то для первого потомка имеем (4):

$$c_{1i} = \begin{cases} p_{1i}, i \le k_1 \text{ или } i > k_2, \\ p_{2i}, k_1 < i \le k_2, \end{cases} \quad c_{2i} = \begin{cases} p_{2i}, i \le k_1 \text{ или } i > k_2, \\ p_{1i}, k_1 < i \le k_2. \end{cases} \tag{4}$$

Выбор точек разреза (k_1, k_2) подчиняется равномерному распределению по всем возможным парам.

Данный метод позволяет объединять в потомках не только префиксы и суффиксы, как в одноточечном кроссовере, но и целые средние сегменты, что увеличивает разнообразие комбинаций генов и делает процесс поиска более гибким. Центральный блок, заимствованный у одного родителя, окружён внешними сегментами другого, и это создаёт более богатые структуры наследования. В отличие от одноточечного кроссовера, двухточечный оператор чаще разрушает и перестраивает внутренние последовательности, что способствует более активному исследованию пространства поиска, хотя при этом возрастает вероятность разрыва полезных комбинаций генов. Тем не менее именно за счёт такой повышенной вариативности данный метод широко применяется в задачах, где необходимо поддерживать высокий уровень разнообразия популяции и избегать преждевременной сходимости.

2 Методы мутации

2.1 Метод Гауссовой мутации

Метод Гауссовой мутации применяется в эволюционных алгоритмах и генетических алгоритмах, когда хромосомы представлены в виде векторов действительных чисел. Его цель заключается в том, чтобы вносить небольшие случайные изменения в отдельные гены особи, тем самым обеспечивая разнообразие популяции и предотвращая преждевременную сходимость к локальным экстремумам. В основе метода лежит использование нормального распределения с нулевым средним значением, которое моделирует случайное возмущение признаков.

Пусть имеется хромосома, заданная как вектор $X = (x_1, x_2, \dots, x_n)$. При применении гауссовой мутации выбирается один или несколько генов, после чего их значения изменяются по правилу (5):

$$x_i' = x_i + N(0, \sigma^2) \tag{5}$$

где $N(0, \sigma^2)$ – нормально распределённая случайная величина с математическим ожиданием 0 и дисперсией σ^2, а параметр σ контролирует масштаб изменения [3]. Если σ велико, мутация приводит к сильным колебаниям признака и позволяет исследовать более удалённые области пространства решений. При малом значении σ изменения оказываются незначительными, что делает поиск более локальным и точным.

Гауссова мутация отличается от равномерной тем, что вероятность сильного отклонения от текущего значения невелика, а малые изменения встречаются чаще, что соответствует свойствам нормального распределения. Благодаря этому оператор обеспечивает преимущественно плавные вариации в пространстве признаков, но при этом сохраняет ненулевую ве-

роятность редких крупных скачков, которые могут оказаться полезными для выхода из локальных экстремумов.

2.2 Метод равномерной целочисленной мутации

Метод равномерной целочисленной мутации используется в генетических алгоритмах для работы с хромосомами, чьи гены принимают дискретные целочисленные значения. Его основная идея состоит в том, что выбранный ген заменяется случайным образом на новое значение, равномерно выбранное из допустимого диапазона.

Пусть особь задана вектором целочисленных признаков $X = (x_1, x_2, \dots, x_n)$ и каждый ген x_i может принимать значения из дискретного множества {a, a+1, …, b}. При применении равномерной целочисленной мутации выбирается один или несколько индексов i, после чего выполняется замена (6):

$$x_i' \sim U(a, a+1, \dots, b) \quad (6)$$

где U обозначает равномерное дискретное распределение. Важно, что новое значение x_i' выбирается независимо от предыдущего и может совпасть с ним лишь случайно, но вероятность этого равна $\frac{1}{b-a+1}$.

Такой оператор мутации радикально отличается от гауссового или других методов малых возмущений, поскольку он не ограничивается локальными изменениями и способен полностью заменить значение признака на любое другое из допустимого диапазона. Это обеспечивает глобальное исследование пространства решений и снижает риск преждевременной фиксации популяции вблизи локального экстремума. Однако из-за высокой степени случайности равномерная целочисленная мутация менее эффективна на этапах тонкой настройки параметров, когда требуется плавное улучшение уже найденных решений.

3. Методы селекции

3.1 Метод турнирной селекции

Метод турнирной селекции является одним из наиболее распространённых механизмов отбора в генетических алгоритмах, который применяется для выбора родителей в процессе эволюции. Его ключевая идея заключается в том, что выбор не производится случайным образом из всей популяции, а осуществляется на основе небольших подмножеств индивидов, формируемых случайным образом.

Пусть имеется популяция $P = (I_1, I_2, \dots, I_n)$, где каждому индивиду соответствует значение функции приспособленности $f(I_j)$. Для проведения турнира выбирается подмножество из t индивидов, где t – параметр турнира, называемый размером турнира. После этого из выбранных участников определяется победитель, который и проходит в следующую стадию – обычно он становится родителем для кроссовера. Вероятность того, что конкретный участник турнира будет выбран, зависит от его относительной приспособленности внутри подмножества. В простейшем случае победителем становится индивидуум с наибольшим значением $f(I_j)$, однако часто вводится вероятностная схема, при которой наиболее приспособленный побеждает с вероятностью p, а остальные – с вероятностью, распределённой между ними.

Формально вероятность выбора индивида I_j в турнире из t участников можно выразить как (7):

$$P(I_j) = p * 1\{I_j - \text{лучший}\} + \{1 - p\} * q(I_j), \quad (7)$$

где $1\{\}$ – индикатор принадлежности индивида к лучшим, а функция $q(I_j)$ задаёт вероятность выбора менее приспособленных особей. При p = 1 турнир становится детерминированным, и всегда выбирается лучший, при p < 1 вводится элемент стохастичности, позволяющий сохранять разнообразие.

Таким образом, выбор одного родителя в рамках турнирной селекции можно записать как (8):

$$\mathrm{I_{parent}} = arg \underbrace{max}_{I_j \in T} f(I_j), \tag{8}$$

если используется строгий вариант, или как случайная величина, распределённая по вероятностям P(I_j), если используется вероятностная версия.

Главная идея метода заключается в том, что он прост в реализации, не требует нормализации значений функции приспособленности и при этом гибко управляется параметром t. При малых значениях t давление отбора слабое, так как велика вероятность выбора менее удачных индивидов, а при больших значениях t селекция становится более жёсткой, и в родительский пул чаще попадают лучшие особи. Это позволяет регулировать баланс между поиском новых решений и закреплением уже найденных успешных комбинаций. Благодаря этим свойствам турнирная селекция широко применяется в эволюционных алгоритмах и считается одним из наиболее устойчивых и надёжных механизмов отбора.

3.2 Метод селекции рулеткой

Метод селекции рулеткой, или пропорциональный отбор, является одним из классических механизмов выбора родителей в генетических алгоритмах. Его суть заключается в том, что вероятность выбора конкретного индивида напрямую зависит от значения его функции приспособленности, благодаря чему более удачные решения получают больший шанс на воспроизводство, но при этом сохраняется возможность выбора и менее успешных особей. Такой подход моделирует принцип «выживания наиболее приспособленных», однако в вероятностной форме, что позволяет избежать преждевременной потери разнообразия в популяции.

Пусть популяция $P = (I_1, I_2, \dots, I_n)$ состоит из n индивидов, а их значения приспособленности равны $f = (I_1)$, $f = (I_2)$, …, $f = (I_n)$. Тогда суммарная приспособленность определяется как (9):

$$F = \sum_{j=1}^{n} f(I_j). \tag{9}$$

Для каждого индивида вероятность его выбора вычисляется как (10):

$$\mathrm{P}(I_j) = \frac{f(I_j)}{F}. \tag{10}$$

Таким образом, чем выше $f(I_j)$, тем больше сектор «рулетки», который соответствует данному индивиду. Процесс выбора можно описать следующим образом: генерируется случайное число r ∈ [0, 1], затем определяется такой индекс k, что (11):

$$\sum_{j=1}^{k-1} P(I_j) < r \leq \sum_{j=1}^{k} P(I_j), \tag{11}$$

и выбранным индивидом становится I_k.

Главная идея данного метода заключается в том, что он обеспечивает мягкий, пропорциональный давлению отбор: хорошие решения чаще воспроизводятся, но слабые всё ещё участвуют в эволюции, поддерживая разнообразие.

3.3 Метод селекции лучших

Метод селекции лучших, или отбор элиты, является одним из самых прямых и жёстких способов выбора родителей в генетических алгоритмах. Его идея заключается в том, что из текущей популяции сохраняется строго определённое количество индивидов с наибольшими значениями функции приспособленности, и именно они используются для формирования новой популяции или для участия в кроссовере и мутациях. Такой подход гарантирует, что лучшие решения никогда не будут потеряны из-за случайности отбора, что способствует стабильному улучшению качества решений на протяжении поколений.

Пусть популяция $P = (I_1, I_2, \dots, I_n)$ имеет значения функции приспособленности $f = (I_1)$, $f = (I_2)$, …, $f = (I_n)$. Если необходимо выбрать M особей для следующего шага алгоритма, то сначала популяция сортируется по убыванию значений приспособленности (12):

$$f(I_{(1)}) \geq f(I_{(2)}) \geq \ldots \geq f(I_{(n)}), \quad (12)$$

где $I_{(1)}$ – индивид с наибольшей приспособленностью, а $I_{(n)}$ – с наименьшей. Тогда множество выбранных особей определяется как $S = \{I_{(1)}, I_{(2)}, \ldots, I_{(M)}\}$.

В отличие от методов рулетки или турнира, где вероятность выбора распределяется по всей популяции, здесь отбор осуществляется детерминированно: в новое поколение гарантированно переходят именно те решения, которые оказались наиболее успешными. Благодаря этому метод селекции лучших часто используется в сочетании с другими механизмами, как элемент элитарной стратегии.

Главное преимущество данного подхода заключается в том, что он обеспечивает монотонный рост максимальной приспособленности в популяции. Это гарантирует, что найденные хорошие решения никогда не теряются. Однако у метода есть и недостатки. Его жёсткость может привести к быстрому уменьшению разнообразия в популяции, так как слабые особи полностью исключаются из дальнейшего процесса эволюции. В результате возрастает риск преждевременной сходимости к локальному оптимуму.

Несмотря на этот недостаток, селекция лучших широко используется на практике, особенно в комбинации с более мягкими стохастическими методами отбора.

Выводы

В данной статье были приведены математические модели методов генетических алгоритмов для построения интеллектуального художника. Генетические алгоритмы моделируют творческий процесс как эволюционный поиск в гигантском пространстве всех возможных изображений. Художник (система) не создает шедевр с первой попытки, а постепенно находит его через множество итераций и «мутаций».

Литература

1. Lozano M. Real-Coded Memetic Algorithms with Crossover Hill-Climbing. / M. Lozano, F. Herrera, N. Krasnogor, D. Molina //Evolutionary Computation 12(3), 2004. – Pp. 273-302.

2. Herrera F. A Taxonomy for the Crossover Operator for Real-Coded Genetic Algorithms/ F. Herrera, M. Lozano, A // International Journal of Intelligent Systems, vol. 18, 2003. – Pp. 309-338.

3. Голубин А.В. Нечеткие генетические алгоритмы / А.В. Голубин, В.Б. Тарасов // Международные научно-технические конференции AIS’05 и CAD-2005. Труды конференций. – М.: Физматлит, 2005. – С. 39-45.

Stoma D.S., Lazebnaya L.A. Realization of an intellectual visual artist based on genetic algorithms. *A brief comparative analysis of the methods of genetic algorithms is performed: methods of crossing, mutation and selection for use in the development of an intelligent artist.*

Keywords: *genetic algorithms, intelligent artist, crossbreeding, mutation, breeding.*

УДК 004.912

ИНТЕЛЛЕКТУАЛЬНЫЙ ПОДХОД К ЛОКАЛЬНОМУ ПОИСКУ, СЕМАНТИЧЕСКОЙ ОБРАБОТКЕ И КАТАЛОГИЗАЦИИ ТЕКСТОВЫХ ФАЙЛОВ НА ОСНОВЕ NLP

Хаунов Д.И., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: danil.khaunov@gmail.com

Хаунов Д.И., Лазебная Л.А. Интеллектуальный подход к локальному поиску, семантической обработке и каталогизации текстовых файлов на основе NLP. *Предложены архитектура модулей, математические модели суммаризации и семантического поиска, а также подход к созданию локального хранилища знаний. Показано, что внедрение подобных систем способно значительно повысить эффективность работы пользователя с большими массивами разнородных документов.*
Ключевые слова: *NLP, семантический поиск, суммаризация текста, обработка документов, локальная интеллектуальная система, автоматическая категоризация.*

Введение

Современные пользователи ежедневно взаимодействуют с большими объемами документов различных форматов, включая PDF, DOCX и TXT. Быстрый рост объёма цифровой информации приводит к ситуациям, когда ручной поиск и категоризация становятся неэффективными и сопровождаются потерями данных или временными затратами [1]. Ключевой недостаток существующих локальных систем поиска — отсутствие семантического понимания текста. Они не умеют автоматически генерировать краткое содержание файлов и выстраивать интеллектуальные тематические связи между ними, ограничиваясь простой индексацией по названию и содержимому [2].

Таким образом, актуальной является задача создания локальной интеллектуальной системы, способной автоматически осмысливать файлы, извлекать суть текста и обеспечивать семантический доступ к содержимому – без передачи данных во внешние облачные системы.

1 Постановка проблемы и анализ исследований

Системы обработки текстов с использованием NLP активно развиваются в областях машинного реферирования, тематического моделирования и вопросно-ответного поиска. Наиболее распространённые направления:

– Automatic Text Summarization – генерация кратких выжимок на основе статистических и нейросетевых моделей [3];
– Semantic search – поиск по смыслу с использованием векторных представлений текста [4];
– Information retrieval – индексирование и поиск по документам [5].

Несмотря на развитие указанных областей, отсутствуют исследовательские решения, объединяющие:

– локальный поиск файлов разного формата;
– автоматическое извлечение и реферирование текста;
– хранение выжимок в структурированной базе знаний;
– возможность диалогового семантического поиска по содержимому персональных файлов;
– автоматическое формирование тематических меток (tagging) с рекомендацией по категоризации.

Это формирует научный разрыв, который и является основой данной работы.

2 Цель и задачи

Цель: разработка концептуальной модели интеллектуальной локальной системы семантической обработки текстовых файлов, включающей суммаризацию, тэгирование и вопросно-ответный поиск.

Задачи:

- определить архитектуру модулей обработки;
- описать математическую модель реферирования и тематического поиска;
- предложить схему хранения выжимок и метаданных;
- обосновать подход к тематическому тэгированию;
- оценить практическую значимость подобной системы.

3 Архитектура и методы

Модель архитектуры представлена на рисунке 1.

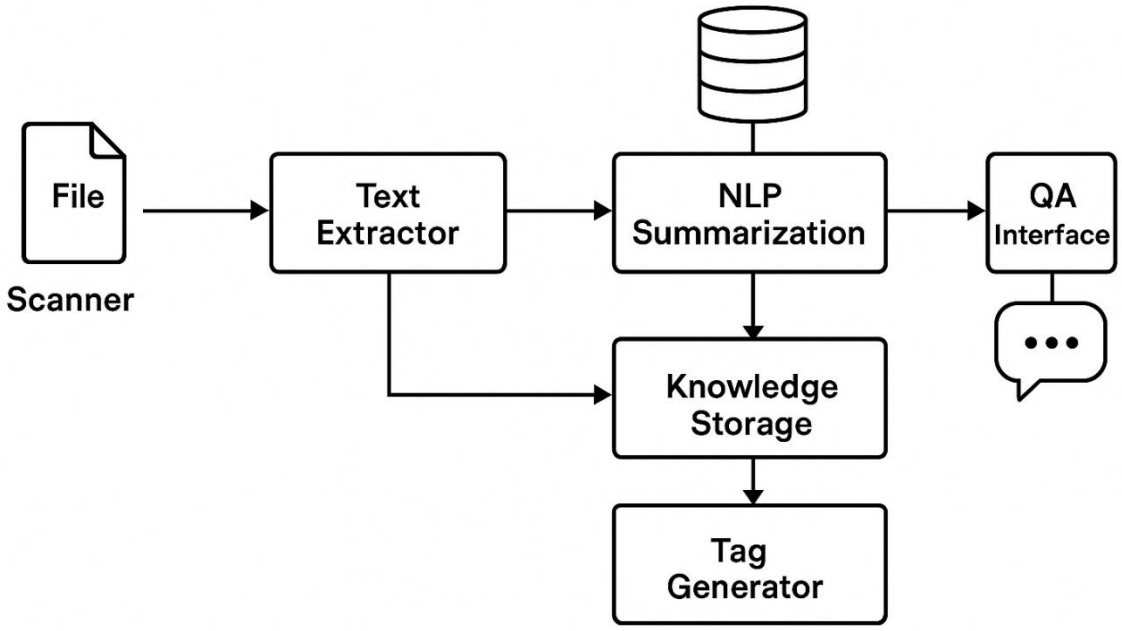

Рисунок 1 – Архитектура интеллектуальной обработки документов

Система включает следующие этапы обработки, представленные в таблице 1:

Таблица 2 – Этапы обработки текстового файла

Модуль	**Функция**
File Scanner	Поиск файлов по расширениям
Text Extractor	Извлечение текста из PDF, DOCX, TXT
Summarizer	Формирование краткой выжимки
Tag Generator	Автоматическая генерация тематических меток
Knowledge Storage	Локальное хранение выжимок и метаданных
QA Module	Ответ на запросы пользователя по смыслу

Сжатие текста описывается как задача поиска максимально информативного подмножества предложений:

$$S = \arg\max_{S \subset D} \sum_{i=1}^{n} \text{Score}(s_i) \quad (1)$$

где D – исходный документ, S – итоговая выборка предложений, Score – оценка значимости, вычисляемая по TF-IDF и центральности.

Тематическая значимость термина вычисляется по классической формуле TF-IDF:

$$\text{TFIDF}(t,d) = TF(t,d) \cdot \log \frac{N}{DF(t)} \quad (2)$$

где N – число документов, $DF(t)$ – количество документов, содержащих термин.

Пользовательский запрос преобразуется в вектор Q, документы – в набор D_i. Отбор наиболее релевантных осуществляется через косинусную близость:

$$R = \{D_i \mid \cos(Q, D_i) > \delta\} \quad (3)$$

где δ – порог релевантности.

4 Формирование базы знаний и тэгирование

Каждому документу присваивается структура:

```
Doc = {
  FilePath,
  Summary,
  Tags[],
  VectorEmbedding,
  Timestamp
}
```

Генерация меток осуществляется двумя путями:

– автоматически – выбором ключевых слов по TF-IDF;
– семантически – тематическим векторным кластерингом.

5 Практическая значимость

Система позволяет:

– сократить время ручного поиска файлов;
– обеспечивать доступ к содержимому без необходимости открывать документ;
– находить тематически схожие файлы, даже если они названы по-разному;
– формировать персональную базу знаний;
– обеспечивать диалоговый поиск.

6 Потенциальные результаты тестирования

Пример формата оценки эффективности описан в таблице 2:

Таблица 3 – Оценка эффективности работы модели интеллектуальной локальной системы обработки текстов

Метрика	**Значение**
Среднее время поиска файла	↓ на 40%
Точность смыслового поиска	0.82 (cosine similarity)
Средний размер выжимки	12–18% от исходного текста
Скорость обработки 1 документа	0.8–3.2 сек

Выводы

Предложена модель интеллектуальной локальной системы обработки текстов, включающая:

– автоматический поиск файлов;
– извлечение содержимого;
– генерацию смысловых выжимок;
– тематическое тэгирование;
– семантический вопросно-ответный интерфейс.

Решение ориентировано на локальное выполнение, что гарантирует конфиденциальность данных. Перспективными направлениями развития являются построение собственного индексного хранилища, внедрение полнотекстового нейропоиска и оптимизация моделей суммаризации для работы без GPU.

Литература

1. Долгих, А.А. Третья платформа информатизации и Big Data / А.А. Долгих, Е.В. Долгих // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. – 2019. – № 48. – С. 5–13.
2. Баранов, А.А. Системы автоматического реферирования / А.А. Баранов // Открытые системы. СУБД. – 2000. – № 12. – С. 28–35.
3. Соколова, Е.А. Автоматическое реферирование текстов. Обзор работ / Е.А. Соколова // Открытые системы. СУБД. – 2022. – № 8. – С. 45–52.
4. Козлов, А.В. Сравнительный анализ методов определения эмоциональной окраски сообщений в социальных сетях с применением обучения с учителем / А.В. Козлов, И.П. Смирнова // Программные продукты и системы. – 2018. – № 3. – С. 112–119.
5. Мэннинг, К. Введение в информационный поиск / К. Мэннинг, П. Рагхаван, Х. Шютце. – М.: Вильямс, 2011. – 512 с.

локальная интеллектуальная система, автоматическая категоризация.

Khaunov D.I., Lazebna L.A. An intelligent approach to local search, semantic processing, and cataloging of text files based on NLP*. The paper proposes a modular architecture, mathematical models for text summarization and semantic search, as well as an approach to building a local knowledge storage. It is demonstrated that the implementation of such systems can significantly improve user efficiency when working with large volumes of heterogeneous documents.*

Keywords: *NLP, semantic search, text summarization, document processing, local intelligent system, automatic categorization.*

УДК 004.75

ПАРАЛЛЕЛЬНАЯ ОБРАБОТКА ТРАНЗАКЦИЙ В СИСТЕМАХ ИНТЕРНЕТ-ТОРГОВЛИ С ИСПОЛЬЗОВАНИЕМ СОВРЕМЕННЫХ АРХИТЕКТУР БАЗ ДАННЫХ

Ястребов А.Р., Рычка О.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия»
e-mail: yastrebov_sasha98@mail.ru, olga_rychka@mail.ru

Ястребов А.Р., Рычка О.В. Параллельная обработка транзакций в системах интернет-торговли с использованием современных архитектур баз данных*. В статье представлено экспериментальное исследование производительности четырех современных систем управления базами данных — PostgreSQL, MySQL, MariaDB и Firebird — в контексте обработки параллельных транзакций в системах интернет-торговли.*
Ключевые слова: *параллельная обработка, СУБД, уровни изоляции, пакетная обработка, системы интернет-торговли.*

Введение

Рост интернет-торговли ведет к увеличению числа параллельных транзакций, что создает риск конфликтов, блокировок и падения производительности в СУБД. Поиск баланса между надежностью и скоростью обработки транзакций является критически важной задачей для масштабируемости современных систем. Цель настоящей работы — провести экспериментальное сравнение четырёх современных систем управления базами данных — PostgreSQL, MySQL, MariaDB и Firebird — с точки зрения их способности эффективно обрабатывать параллельные транзакции в условиях, моделирующих реальные сценарии интернет-торговли. Задачами исследования являются анализ влияния архитектуры СУБД и уровня изоляции транзакций на производительность, а также определение оптимальных конфигураций, обеспечивающих баланс между скоростью обработки и надёжностью данных.

1 Теоретические основы параллельной обработки транзакций

Основу параллельной обработки составляет транзакция — последовательность операций с базой данных, которые выполняются как единое целое, через атомарность, согласованность, изоляцию и долговечность [1]. Данные свойства реализованы во всех ведущих реляционных СУБД, что обеспечивает высокую степень надёжности корпоративных информационных систем.

Уровни изоляции (Read Uncommitted, Read Committed, Repeatable Read, Serializable) определяют, как транзакции могут взаимодействовать между собой, и насколько сильно могут пересекаться и мешать друг другу при параллельной работе. Современные СУБД развивают параллелизм через многопоточность, распределённые транзакции и батчевую обработку. Механизмы вроде многоверсионности и предикативных блокировок минимизируют конфликты и распределяют нагрузку.

В интернет-торговле, где параллельно выполняются заказы, обновления остатков и платежи, корректные уровни изоляции и архитектурные решения СУБД предотвращают конфликты и потери данных, обеспечивая отказоустойчивость и производительность при высокой нагрузке.

2 Методология исследования

Для исследования использовались четыре СУБД — PostgreSQL, MySQL, MariaDB и Firebird, каждая с различными механизмами параллелизма. PostgreSQL применяет MVCC, MySQL — блокировки InnoDB [2], MariaDB — parallel replication, Firebird — многопоколен-

ную архитектуру [3]. Эксперименты проводились на ноутбуке с ОС Windows 10, процессором AMD Ryzen 5 3500U (2.1 ГГц) и 8 ГБ ОЗУ. Для тестирования использовалась унифицированная база данных с таблицами users, products и orders, моделирующая структуру интернет-магазина. Это обеспечило сопоставимость результатов при анализе производительности различных систем.

Было подготовлено четыре скрипта для трёх типов исследований: 1) оценка масштабируемости СУБД путём измерения производительности при росте числа параллельных транзакций; 2) анализ компромисса между согласованностью данных и скоростью работы при разных уровнях изоляции; 3) исследование влияния размера пакета операций на время отклика и количество успешно выполненных транзакций.

Параметры исследований включали варьирование количества параллельных транзакций, изменение уровня изоляции и размера пакета. Для оценки эффективности использовались ключевые метрики: время обработки транзакций, число успешно завершённых транзакций и задержка выполнения.

3 Результаты экспериментов

3.1 Производительность при увеличении числа параллельных транзакций

Для оценки масштабируемости и способности СУБД эффективно обрабатывать параллельные транзакции было проведено экспериментальное исследование с использованием Python-скрипта, который имитировал выполнение типичных операций интернет-магазина: добавление нового пользователя, оформление заказа и обновление остатка товара. Для каждой СУБД были заданы одинаковые сценарии, различавшиеся только уровнем параллелизма. Количество потоков (1, 5, 10, 20 и 50) соответствовало числу одновременно выполняемых транзакций. Каждая транзакция включала три основные операции записи в базу данных, выполнявшиеся в рамках одной логической сессии с управлением соединениями через пул.

PostgreSQL показала плавный рост времени отклика и 100% успешность транзакций благодаря эффективной реализации многоверсионного контроля (MVCC), что подтверждает её стабильность и хорошую масштабируемость.

Результаты исследования производительности для PostgreSQL приведены в таблице 1.

Таблица 1. Исследование производительности для PostgreSQL

Потоки	Среднее время (мс)	Успехов	Ошибок	Общее время (с)
1	15.63	1	0	0.016
5	3.13	5	0	0.485
10	4.69	10	0	1.172
20	5.35	20	0	2.476
50	6.82	50	0	6.862

MySQL демонстрировала умеренный рост производительности до 20 потоков, но при 50 транзакциях возникли ошибки фиксации (7 из 50), что связано с блокировками (gap locks, next-key locks) в InnoDB. Несмотря на относительно высокую скорость обработки при малом числе потоков, при увеличении нагрузки система столкнулась с внутренними конфликтами транзакций.

Результаты исследования производительности для MySQL приведены в таблице 2.

Таблица 2. Исследование производительности для MySQL

Потоки	Среднее время (мс)	Успехов	Ошибок	Общее время (с)
1	26.25	1	0	0.026
5	15.63	5	0	0.016
10	13.07	10	0	0.027
20	13.28	20	0	0.031
50	16.07	43	7	0.053

MariaDB показала наилучшие результаты при средних нагрузках (5–20 потоков) с минимальным временем выполнения и незначительным числом ошибок, что обусловлено эффективным распределением нагрузки и parallel replication.

Результаты исследования производительности для MariaDB приведены в таблице 3.

Таблица 3. Исследование производительности для MariaDB

Потоки	Среднее время (мс)	Успехов	Ошибок	Общее время (с)
1	14.15	1	0	0.014
5	0.00	5	0	0.016
10	0.00	10	0	0.016
20	7.89	19	1	0.024
50	4.69	49	1	0.047

Firebird продемонстрировала худшую производительность с резким ростом времени выполнения, хотя и сохранила 100% успешность транзакций. Это объясняется накладными расходами многопоколенной архитектуры (MGA) на управление версиями данных.

Результаты исследования производительности для Firebird приведены в таблице 4.

Таблица 4. Исследование производительности для Firebird

Потоки	Среднее время (мс)	Успехов	Ошибок	Общее время (с)
1	17.68	1	0	0.116
5	73.47	5	0	0.188
10	65.05	10	0	0.184
20	129.06	20	0	0.296
50	286.06	50	0	0.740

На основании полученных данных можно заключить, что лучшую масштабируемость при росте числа параллельных транзакций показали PostgreSQL и MariaDB, которые эффективно реализуют многоверсионный контроль и оптимизированное распределение потоков. MySQL сохранила приемлемый уровень производительности, но столкнулась с блокировками при высокой нагрузке. Firebird продемонстрировала устойчивость к ошибкам, но при этом существенно потеряла в скорости обработки из-за особенностей своей архитектуры.

3.2 Влияние уровня изоляции транзакций

Для анализа влияния уровня изоляции на производительность транзакционной обработки были проведены тесты для трёх стандартных уровней — READ COMMITTED, REPEATABLE READ и SERIALIZABLE. Эксперименты имитировали одновременное выполнение множества операций вставки заказов в базу данных интернет-магазина, где 10 потоков выполняли по 50 транзакций. Скрипт, динамически изменял уровень изоляции для каждого теста, чтобы оценить влияние этого параметра на общую скорость выполнения и стабильность фиксации транзакций.

Результаты исследования представлены в таблице 5.

Таблица 5. Исследование влияния уровней изоляции транзакций

Уровень изоляции	PostgreSQL (время, успешные / ошибки)	MySQL (время, успешные / ошибки)	MariaDB (время, успешные / ошибки)	Firebird (время, успешные / ошибки)
READ COMMITTED	1.11 с — 500 / 0	0.70 с — 500 / 0	0.26 с — 462 / 38	2.51 с — 500 / 0
REPEATABLE READ	1.23 с — 500 / 0	0.55 с — 500 / 0	0.24 с — 450 / 50	2.03 с — 500 / 0
SERIALIZABLE	1.20 с — 500 / 0	0.53 с — 500 / 0	0.24 с — 452 / 48	1.97 с — 500 / 0

Изменение уровня изоляции по-разному влияло на СУБД. PostgreSQL и Firebird обеспечили 100% успешность, MySQL и MariaDB — максимальную скорость при лёгком сниже-

нии стабильности. Оптимум между скоростью и надёжностью достигается на уровне Read Committed для MySQL и Repeatable Read для PostgreSQL.

3.3 Влияние размера пакета транзакций

В каждом тесте выполнялась вставка 1000 заказов в таблицу orders с варьированием параметра batch size — количества операций, объединённых в одну транзакцию. Используемый скрипт открывал соединение с базой данных, последовательно выполнял пакеты вставок и фиксировал результаты (время выполнения и среднюю загрузку процессора). Таким образом моделировались реальные сценарии интернет-торговли, где группы заказов могут обрабатываться пакетно, например при массовом оформлении покупок или синхронизации данных между сервисами.

Результаты исследование представлены в таблице 6.

Таблица 6. Исследование влияния размера пакета транзакций

Batch size	PostgreSQL (время / CPU %)	MySQL (время / CPU %)	MariaDB (время / CPU %)	Firebird (время / CPU %)
1	0.48 с / 3.3 %	6.44 с / 5.5 %	2.56 с / 2.2 %	2.56 с / 2.2 %
10	0.22 с / 10.7 %	0.94 с / 7.9 %	0.42 с / 5.9 %	0.42 с / 5.9 %
100	0.17 с / 10.9 %	0.38 с / 16.5 %	0.17 с / 8.6 %	0.17 с / 8.6 %

Для PostgreSQL время выполнения снизилось с 0.48 с (batch=1) до 0.17 с (batch=100) при незначительном росте нагрузки на CPU, что подтверждает эффективность буферизации в MVCC. Для сценариев интернет-торговли оптимальным значением можно считать batch = 50–100, обеспечивающим баланс между скоростью и умеренной нагрузкой на процессор.

MySQL продемонстрировала значительный прирост производительности при увеличении batch size: время сократилось с 6.44 с при batch = 1 до 0.38 с при batch = 100. Это объясняется тем, что InnoDB при каждой фиксации выполняет запись в журнал и синхронизацию данных на диск; при увеличении размера пакета эти накладные расходы распределяются на большее число операций. Повышение загрузки CPU с 5.5 % до 16.5 % указывает на более активное использование буфера и улучшение пропускной способности. Таким образом, MySQL выигрывает от крупных пакетов транзакций, особенно при вставках большого объёма данных.

MariaDB показала результат, аналогичный PostgreSQL (0.17 с при batch=100), с 15-кратным ускорением по сравнению с одиночными вставками.

Firebird, согласно представленным данным, продемонстрировала схожую с MariaDB динамику (с 2.56 с до 0.17 с), так как крупные пакеты уменьшают частоту операций фиксации, затратных в MGA-архитектуре.

Эксперимент подтвердил, что увеличение размера пакета транзакций значительно повышает производительность всех исследуемых СУБД. Наибольший прирост эффективности наблюдается в MySQL и MariaDB, где использование крупных пакетов сокращает время выполнения более чем в 10 раз. PostgreSQL демонстрирует стабильное улучшение производительности при минимальной нагрузке на процессор, а Firebird показывает аналогичную MariaDB зависимость от частоты фиксаций транзакций. Для систем интернет-торговли с групповой обработкой заказов оптимальным является использование размера пакета от 50 до 100 операций, что обеспечивает максимальную пропускную способность при минимальных накладных расходах.

Выводы

В работе рассмотрены теоретические и практические аспекты параллельной обработки транзакций в системах интернет-торговли с использованием современных архитектур баз данных. Проведённые эксперименты показали, что наиболее стабильные и масштабируемые результаты продемонстрировали PostgreSQL и MariaDB благодаря эффективной реализации многоверсионного контроля и оптимизации параллельных операций. MySQL обеспечила вы-

сокую скорость обработки при умеренной нагрузке, однако при большом числе одновременных транзакций наблюдались блокировки. Firebird сохранила полную надёжность выполнения, но уступила по производительности из-за особенностей своей многопоколенной архитектуры.

Увеличение размера пакета транзакций позволило всем системам существенно сократить время обработки, что подтверждает важность оптимизации групповых операций при проектировании высоконагруженных сервисов. Полученные результаты подтверждают актуальность выбора архитектуры и параметров СУБД как ключевого фактора устойчивости и эффективности современных систем интернет-торговли [5].

Литература

1. Gray, J. The Transaction Concept: Virtues and Limitations [Электронный ресурс] / Jim Gray // Microsoft Research. – 1995. – Режим доступа: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr-95-51.pdf - Загл. с экрана.

2. MySQL 8.4 Reference Manual. InnoDB Transaction Isolation Levels [Электронный ресурс] / Oracle Corporation. – 2024. – Режим доступа: https://dev.mysql.com/doc/refman/8.4/en/innodb-transaction-isolation-levels.html - Загл. с экрана.

3. Firebird Documentation: Using Firebird [Электронный ресурс] / FirebirdSQL Foundation. – 2023. – Режим доступа: https://www.firebirdsql.org/file/documentation/html/en/firebirddocs/ufb/using-firebird.html - Загл. с экрана.

4. TPC Benchmark™ C Standard Specification [Электронный ресурс] / Transaction Processing Performance Council (TPC). – Version 5.11.0. – 2023. – Режим доступа: https://www.tpc.org/tpc_documents_current_versions/pdf/tpc-c_v5.11.0.pdf - Загл. с экрана.

5. Harizopoulos, S. OLTP Through the Looking Glass, and What We Found There [Электронный ресурс] / Stavros Harizopoulos, Daniel J. Abadi, Samuel Madden, Michael Stonebraker // Proceedings of the ACM SIGMOD. – 2008. – Режим доступа: https://dl.acm.org/doi/10.1145/1620585.1620587 - Загл. с экрана.

Yastrebov A.R., Rychka O.V. Parallel processing of transactions in online retail systems using modern database architectures. *The article presents an experimental study of the performance of four modern database management systems — PostgreSQL, MySQL, MariaDB, and Firebird — in the context of processing parallel transactions in online retail systems.*
Keywords: *parallel processing, database management systems, isolation levels, batch processing, online retail systems.*

УДК 004.8

РАЗРАБОТКА ПОЛЬЗОВАТЕЛЬСКИХ ИНТЕРФЕЙСОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕЙ.

Буловацкий Е.А., Прокопенко Е.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: gorelmelloi@mail.ru

Буловацкий Е.А., Прокопенко Е.В. Разработка пользовательских интерфейсов с использованием нейросетей. *В данной работе проведена систематизация и анализ моделей применения нейронных сетей в полном цикле разработки пользовательских интерфейсов. Исследование охватывает гибридный процесс, включая автоматизацию проектирования и генерацию фронтенд-кода, а также интеграцию предиктивных алгоритмов в интерфейс продукта. Рассмотрены ключевые инструменты и выявлены основные ограничения.*

Ключевые слова*: нейронная сеть, генерация, обучение, UI/UX, Design-to-Code, Text-to-UI*

Введение

Искусственный интеллект (ИИ) эволюционировал от экспертных систем, основанных на жестких правилах, до сложных моделей глубокого обучения, способных обучаться на больших объемах данных и выполнять разнообразные задачи с высокой степенью точности. При этом ключевым посредником для реализации возможностей ИИ и передачи результатов его работы конечным пользователям остается пользовательский интерфейс. Актуальность данной работы обусловлена необходимостью анализа трансформации всей парадигмы разработки пользовательских интерфейсов с применением нейросетей, включая как автоматизацию процесса проектирования, так и интеграцию предиктивных алгоритмов в конечный продукт. Внедрение ИИ в UI/UX демонстрирует значительный экономический эффект, позволяя сократить время, затрачиваемое на обучение сотрудников, повысить общую производительность.

В контексте настоящего исследования под разработкой пользовательских интерфейсов понимается гибридный междисциплинарный процесс, охватывающий этапы от предиктивного анализа до технической реализации. Современные подходы к разработке интерфейсов с использованием нейросетей условно разделяются на два типа: улучшение процесса создания самого дизайна и повышение пользовательского опыта (UX) посредством внедрения алгоритмов ИИ в сам интерфейс. Первый тип включает автоматизацию рутинных операций, таких как анализ фидбэка пользователей, генерация контента и тестирование прототипов. Второй тип подразумевает интеграцию предиктивной аналитики и адаптивных решений, которые меняются в зависимости от поведения пользователя.

Целью данной работы является систематизация и анализ основных способов применения нейросетей на протяжении всего цикла разработки UI/UX. В ходе исследования будет рассмотрена модель AI-генерации интерфейсов и компонентов, методы автоматизации разработки и тестирования, а также ключевые ограничения внедрения нейросетевых решений в отрасль. Анализ охватит инструменты, предназначенные как для дизайнеров, так и для разработчиков. Полученные данные позволят оценить перспективы применения нейронных сетей в разработке интерфейса и при интеграции ИИ в интерфейс.

Применение нейросетей в процессе разработки UI/UX

Начальный этап работы продуктового дизайнера традиционно связан с глубоким погружением в задачу, что требует анализа рынка, конкурентов и целевой аудитории. Нейронные сети значительно улучшают этот процесс, выполняя функцию интеллектуального по-

мощника. Они анализируют большие объемы данных и быстро предоставляют актуальные и структурированные выводы, формируемые на основе обширного массива разнообразных данных. В крупной компании ИИ в руках frontend-разработчика, вероятно, не сможет полностью заменить отдельного BigData-аналитика. Однако в небольших компаниях, где отсутствует выделенный BigData-специалист, такая функциональность обеспечивает значимое преимущество для UI/UX-разработчика.

Помимо этого, ИИ эффективно применяется для систематизации и анализа обратной связи от пользователей и цифровых метрик взаимодействия с интерфейсом. Это устраняет необходимость ручного просмотра сотен комментариев и больших таблиц данных, предоставляя дизайнеру концентрированные обобщения или готовые рекомендации по оптимизации продукта. Таким образом, автоматизируется рутинная аналитическая работа, высвобождая ресурсы специалиста для стратегического проектирования.

Нейросети активно используются для генерации элементов дизайн-среды, необходимых в процессе проектирования. Это включает автоматизированное создание и проверку микротекстов для интерфейсов, что особенно актуально при работе с неродным языком, а также генерацию изображений и иллюстраций, которые могут заменить стоковые материалы, обеспечивая более высокую скорость и точность.. Однако при использовании ИИ для генерации контента критически важен баланс. Чрезмерное или некачественное использование автоматически сгенерированных материалов может быть замечено конечным пользователем, что, в свою очередь, приводит к потере доверия и негативно сказывается на пользовательском опыте.

В сфере разработки пользовательских интерфейсов ИИ-инструменты достигли уровня конвертации концепции в редактируемый прототип. Нейросети способны преобразовывать как текстовые описания требований, так и рисунки-наброски от руки в полноценные рабочие макеты. Примеры таких платформ, как Uizard, демонстрируют возможность автоматизированного проектирования многоэкранных интерфейсов и клонирования внешнего вида по загруженным скриншотам. Подобная функциональность позволяет создать структурный скелет интерфейса намного быстрее, значительно сокращая время на начальный этап макетирования.

Интеграция искусственного интеллекта активно реализуется в рамках существующих профессиональных редакторов. В таких платформах, как Figma, доступны расширения на базе ИИ и встроенные инструменты генеративного дизайна, которые совместно выполняют широкий спектр задач: от генерации и перевода микротекстов до создания иконок, цветовых палитр, изображений. Кроме того, специализированные сервисы, такие как Khroma, применяют ИИ для формирования гармоничных и эстетичных цветовых схем, основанных на предпочтениях пользователя.

Ключевым этапом оптимизации интерфейса является автоматизированное тестирование прототипов без привлечения реальных пользователей. Специализированные инструменты, такие как EyeQuant и Maze, используют ИИ для имитации когнитивного и поведенческого отклика живых людей, позволяя быстро проводить тестирование вариантов дизайна. Нейронная сеть способна генерировать тепловые карты, строить воронки и проводить виртуальные A/B тесты. Такой подход позволяет дизайнеру глубоко анализировать верстку и сравнивать ее с лучшими практиками UI/UX, выявляя ошибки в расположении элементов и адаптации, которые могли бы быть замечены только после получения жалоб. Применение ИИ значительно снижает нагрузку на команду тестировщиков, автоматизируя проверку базовой структуры и логики. Однако для оценки субъективного, эмоционального восприятия и культурного контекста интерфейса, традиционное пользовательское тестирование остается незаменимым.

Наконец, ИИ обеспечивает прямой мост между дизайном и разработкой, представляя собой завершающий этап в конвейере проектирования. Такой подход называется Design-to-Code [1]. Инструменты, такие как Locofy, берут готовые макеты из профессиональных редакторов, например, Figma или Uizard, и преобразуют их в рабочий, структурированный код

для популярных фронтенд-фреймворков, включая React, Angular и Vue.js. На более продвинутом уровне, специализированные AI-модели, такие как GPT-Engineer, способны трансформировать текстовое описание требований в готовую структуру проекта с файлами и сгенерированным кодом, что фактически позволяет ИИ выступать полноценным членом команды разработчиков. Таким образом, связка ИИ-инструментов формирует сквозной конвейер, который значительно ускоряет путь от концепции до готового фронтенд-кода.

Интеграция AI-алгоритмов в пользовательский интерфейс

Внедрение элементов искусственного интеллекта непосредственно в интерфейс продукта является более глубоким уровнем интеграции по сравнению с автоматизацией процесса разработки. В этом случае ИИ перестает быть инструментом дизайнера и становится ключевой частью пользовательского опыта, часто являясь основным и главным элементом всего приложения. Данный подход требует от разработчиков и дизайнеров глубокого понимания механизмов работы алгоритмов, включая их оптимизацию, для эффективной реализации новых функциональных возможностей, улучшающих UX

Ключевым применением нейросетей в интерфейсах является предиктивная аналитика. Алгоритмы машинного обучения анализируют паттерны поведения пользователя, его историю взаимодействий и контекст, чтобы предсказать его дальнейшие действия. На основе этих предсказаний интерфейс может динамически адаптироваться [2]: предлагать релевантный контент, персонализированные опции или предлагать контекстуально релевантные ответы, в комментариях или на почте. Таким образом, интерфейс переходит от реактивного, то есть отвечающего на действие пользователя, к проактивному, предугадывающему намерение пользователя.

Одной из наиболее распространенных моделей интеграции ИИ в пользовательский опыт является персонализация на базе рекомендательных систем. В этом случае алгоритмы, опираясь на обученные модели, анализируют поведенческие данные пользователя, чтобы предлагать товары, медиаконтент или сервисы, максимально соответствующие его индивидуальным предпочтениям.

С точки зрения архитектуры, интерфейсы, интегрированные с ИИ, можно разделить на несколько моделей взаимодействия. Первая модель рассматривает интерфейс как отдельную часть программного обеспечения, которое использует API нейросети как сервис. Вторая модель предполагает, что интерфейс является частью самой системы ИИ, где логика взаимодействия и вывода данных неразрывно встроена в алгоритм. Третья модель, являющаяся наиболее актуальной, это гибридный подход.

Гибридный подход, реализованный в таких платформах, как ChatGPT, представляет собой наиболее гармоничный метод взаимодействия. В нем нейросеть является ядром продукта, но пользователь взаимодействует с ней через классический, удобный UI, который предоставляет контекст, управляет входными данными и форматирует вывод. Эта модель позволяет совместить гибкость и мощность ИИ с привычными принципами юзабилити.

Интеграция ИИ в интерфейс продукта формирует новый уровень взаимодействия, при котором пользовательский опыт становится не статичным набором функций, а персонализированной и постоянно обучающейся системой. Это обеспечивает непрерывное совершенствование UX, поскольку каждая новая итерация взаимодействия пользователя с продуктом служит дополнительным обучающим примером для нейросети, делая продукт более релевантным и полезным для конкретного индивида.

Вызовы, ограничения и перспективы развития

Несмотря на значительный потенциал и очевидные преимущества внедрения нейросетей в разработку пользовательских интерфейсов, процесс сопряжен с рядом технологических, этических и экономических ограничений. Одним из ключевых барьеров является высокая стоимость разработки и обучения специализированных моделей. Создание собственных AI-решений требует значительных инвестиций в вычислительные ресурсы и привлечения

высококвалифицированных ML-специалистов, что часто недоступно малому и среднему бизнесу.

Другое серьезное ограничение связано с качеством и объемом данных. Эффективность работы генеративных моделей и предиктивной аналитики напрямую зависит от массива входных данных. Недостаток структурированных, непредвзятых данных может привести к генерации шаблонных или нерелевантных решений, а также к закреплению и усилению алгоритмической предвзятости в интерфейсах [3], что негативно сказывается на объективности и справедливости UX.

Внедрение ИИ также порождает ряд негативных эффектов, связанных с самим дизайном. Отмечено, что массовое использование генеративных моделей приводит к усредненному и шаблонному дизайну, лишенному оригинальности, поскольку модели стремятся к наиболее "безопасному" и статистически часто встречающемуся решению. Это может вызвать инерцию отрасли и замедлить инновационное развитие в области визуального проектирования.

Ключевым вызовом является проблема непредсказуемости и непрозрачности работы AI-интерфейсов. Пользователи могут испытывать дискомфорт, когда адаптивный интерфейс меняет свое поведение без очевидной логики, что подрывает доверие к продукту. В результате, задача дизайнера усложняется: необходимо создавать не только интуитивно понятные интерфейсы для человека, но и интуитивно понятные интерфейсы для объяснения логики работы ИИ [4], их еще называют Explainable AI или кратко XAI.

Перспективы развития интерфейсов связаны с дальнейшей абстракцией процесса проектирования. Наблюдается тенденция к тому, что в будущем интерфейсы могут создаваться на основе высокоуровневого, абстрактного описания задачи, а не через детальное ручное рисование и кодирование. Нейросети будут способны генерировать целые программные комплексы, где UI/UX выступает лишь внешней оболочкой. Однако реализация таких систем является крайне оптимистичным, но не гарантированным прогнозом. Это потребует не только значительного технологического прорыва, но и сохранения роли высококвалифицированных специалистов, способных контролировать, критически оценивать и корректировать сложные, абстрактные результаты работы ИИ.

Наконец, ключевое направление развития заключается в гармоничном сотрудничестве человека и ИИ [5]. В то время как ИИ берет на себя рутину и анализ больших данных, человеческий фактор, включающий эмпатию, креативность и этическое суждение, останется незаменимым. В конечном итоге, именно специалисты, освоившие работу с ИИ-инструментами, определят новый стандарт конкурентоспособности, формируя гибридный стандарт разработки пользовательских интерфейсов.

Выводы

Проведенный анализ подтверждает, что применение нейросетей трансформирует процесс разработки пользовательских интерфейсов, охватывая полный цикл создания продукта. Была достигнута цель исследования по систематизации моделей применения ИИ, которые условно разделены на автоматизацию процесса разработки и интеграцию алгоритмов в сам интерфейс, повышая адаптивность UX. Установлено, что первый подход реализуется через инструменты Text-to-UI и Design-to-Code, а второй через предиктивную аналитику и гибридную архитектуру.

Ключевым результатом внедрения ИИ является значительное повышение производительности и снижение временных затрат на рутинные операции, такие как анализ фидбэка, генерация контента и базовое тестирование. Однако это формирует новый вызов для отрасли, требуя от специалистов, особенно от Frontend-разработчиков, освоения новых инструментов и технологий.

Однако не менее важным вызовом является управленческая ошибка, при которой ИИ рассматривается как полная замена человеку, а не как инструмент, что ведет к нерациональным кадровым решениям. При этом критически важным аспектом остается факт, что совре-

менные генеративные модели все еще демонстрируют значительный процент ошибок, что делает их использование нерациональным или рискованным в критических частях проекта. Это, в свою очередь, формирует угрозу усреднения дизайна и необходимости создания более прозрачных интерфейсов, объясняющих логику работы ИИ (XAI), для поддержания пользовательского доверия и снижения рисков, связанных со слепым доверием к алгоритмам.

Перспективы дальнейшего развития разработки интерфейсов неразрывно связаны с процессом упрощения рутинных операций и переходом к абстракции проектирования. Однако даже в этом случае сохраняется необходимость в высококвалифицированных специалистах, способных выполнять ручную верификацию результатов и сохранять человеческий фактор (эмпатию и креативность) в финальном продукте. Таким образом, эффективное освоение AI-инструментов становится решающим фактором конкурентоспособности в современной разработке.

Литература

1. André Costa, Firmino Silva, José Joaquim Moreira Towards an AI-Driven User Interface Design for Web Applications // Procedia Computer Science. 2024. Vol.237, P. 179-186 DOI: 10.1016

2. Xiaoan Zhan, Yang Xu, Yingchia Liu Personalized UI Layout Generation using Deep Learning: An Adaptive Interface Design Approach for Enhanced User Experience // International Journal of Innovative Research in Engineering and Management. 2024. Vol.11, P. 68-79 DOI: 10.55524

3. Matthew G. Hanna, Liron Pantanowitz, Brian Jackson, Octavia Palmer, Shyam Visweswaran, Joshua Pantanowitz, Mustafa Deebajah, Hooman H. Rashidi Ethical and Bias Considerations in Artificial Intelligence/Machine Learning // Modern Pathology. 2025. Vol.38, DOI: 10.1016

4. Sajid Ali, Tamer Abuhmed, Shaker El-Sappagh Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence // Information Fusion. 2023. Vol.99 DOI:10.1016

5. Jacqueline Audrey Iman, Alicia Felisha, Michael Kimeison, Eric Savero Hermawan, Reinert Yosua Rumagit, Hady Pranoto Refining UI/UX with Minimalist Design and AI: Towards Sustainable and Efficient Digital Experiences // Procedia Computer Science. 2025. Vol.269, P. 669-680 DOI:10.1016

Bulovatsky E.A., Prokopenko E.V. Development of user interfaces using neural networks. *This work systematizes and analyzes models for applying neural networks across the entire lifecycle of user interface development. The research covers a hybrid process, including the automation of design (prototyping) and frontend code generation, as well as the integration of predictive algorithms into the final product interface. Key tools and core limitations of this approach are reviewed.*

Keywords*: network, generation, learning, UI/UX, Design-to-Code, Text-to-UI*

СЕКЦИЯ №3
КОМПЬЮТЕРНАЯ ИНЖЕНЕРИЯ

УДК 004.89

РАЗРАБОТКА MCP-СЕРВЕРА ДЛЯ АВТОМАТИЗАЦИИ ВЗАИМОДЕЙСТВИЯ С ГРАФИЧЕСКИМ ИНТЕРФЕЙСОМ WINDOWS

Дворник Н.Н., Чередникова О.Ю.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия им. Л.П.Фельдмана»
e-mail: n.n.dvornik@yandex.ru

Дворник Н.Н., Чередникова О.Ю. Разработка MCP-сервера для автоматизации взаимодействия с графическим интерфейсом Windows. *Разработан MCP-сервер для интеграции больших языковых моделей с Windows. Система реализует распознавание элементов графического интерфейса через анализ дерева доступности без компьютерного зрения и обеспечивает автоматизированное управление через программную имитацию пользовательского ввода.*
Ключевые слова: *автоматизация, MCP-сервер, дерево доступности, языковые модели, UI Automation, Windows.*

Введение

Современные большие языковые модели (LLM), ставшие широко известными благодаря таким системам, как YandexGPT, ChatGPT, Gemini и Claude, демонстрируют впечатляющие возможности в решении интеллектуальных задач. Однако их взаимодействие с операционными системами на уровне графического интерфейса (GUI) остается нетривиальной проблемой. Существующие решения для автоматизации GUI преимущественно ориентированы на веб-браузеры или требуют применения методов компьютерного зрения, что увеличивает вычислительную сложность и снижает детерминированность системы.

Model Context Protocol (MCP) представляет собой открытый стандарт для подключения ИИ-приложений к внешним системам [1]. Протокол определяет способ взаимодействия клиентов (ИИ-приложений) с серверами, предоставляющими инструменты, ресурсы и контекст для языковых моделей. Архитектура MCP состоит из трёх основных компонентов: Host (хост-приложение с LLM), Client (клиент протокола) и Server (сервер, предоставляющий возможности).

Целью данной работы является разработка MCP-сервера, обеспечивающего программное распознавание и управление элементами графического интерфейса Windows без применения методов компьютерного зрения.

1 Архитектура системы

Разработанная система построена на принципах Clean Architecture – подхода к структурированию приложений, обеспечивающего разделение ответственности, масштабируемость и тестируемость [2]. Архитектура разделена на четыре основных слоя с соблюдением правила зависимостей, согласно которому внешние слои могут зависеть от внутренних, но не наоборот.

Уровень API (MCP Router) представляет собой точку входа для взаимодействия с языковыми моделями. Этот слой реализован на базе фреймворка FastMCP и предоставляет набор из 13 инструментов: State-Tool для захвата состояния интерфейса, Click-Tool и Type-Tool для имитации пользовательского ввода, Scroll-Tool для навигации, Launch-Tool для запуска приложений, Shell-Tool для выполнения команд PowerShell, а также инструменты для работы с буфером обмена, горячими клавишами и перетаскиванием элементов. Каждый инструмент определяется с помощью декоратора @mcp.tool и автоматически регистрируется в системе.

Уровень сервисов (Service Layer) содержит бизнес-логику системы и включает три основных компонента: DesktopService для управления рабочим столом и выполнения дей-

ствий, TreeService для анализа дерева доступности и AppService для управления приложениями. Класс DesktopService выполняет роль оркестратора, координируя работу остальных компонентов и обеспечивая единую точку доступа к функциональности системы. TreeService реализует механизм рекурсивного обхода дерева доступности и классификации элементов интерфейса.

Уровень инструментов (Tools Layer) представляет собой набор независимых модулей, каждый из которых реализует отдельное действие: click.py для кликов мыши, type.py для ввода текста, state.py для захвата состояния системы. Такая модульная организация позволяет легко добавлять новые инструменты без изменения ядра системы, следуя принципу открытости/закрытости.

Инфраструктурный уровень (Infrastructure Layer) обеспечивает интеграцию с внешними библиотеками и Windows API через адаптеры: PyAutoGUIAdapter для выполнения физических действий (клики, ввод текста), UIAutomationAdapter для программного доступа к дереву элементов. Этот слой изолирует систему от изменений в сторонних библиотеках, обеспечивая стабильность и возможность замены реализации без влияния на верхние слои.

Взаимодействие между слоями осуществляется через чётко определённые интерфейсы, что обеспечивает слабую связанность компонентов и упрощает тестирование. Зависимости направлены от внешних слоёв к внутренним, что позволяет изменять реализацию инфраструктурного уровня без модификации бизнес-логики. Общая архитектура системы представлена на рисунке 1.

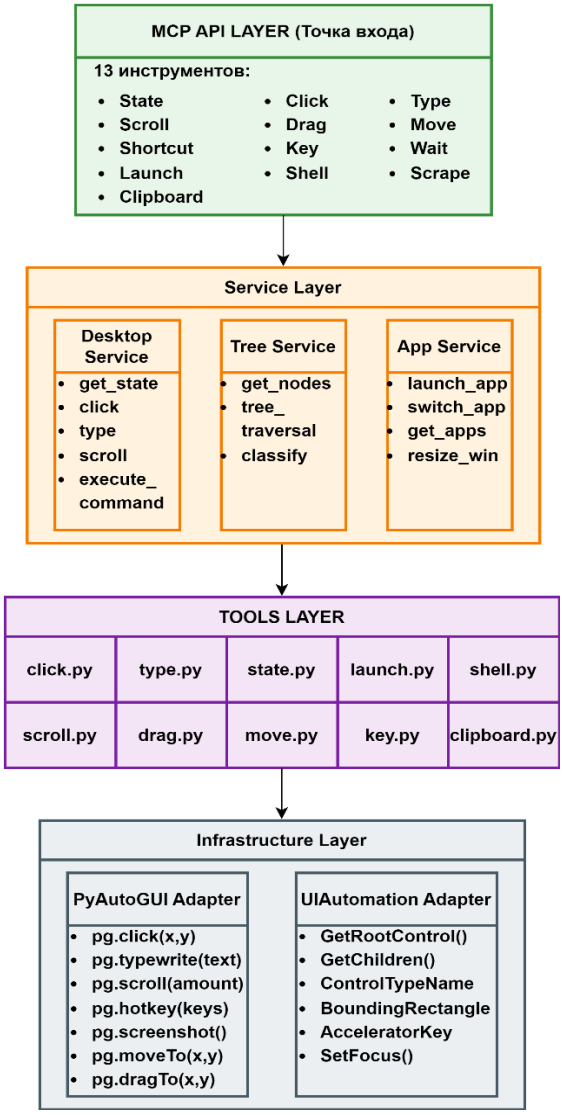

Рисунок 1 – Слоистая архитектура MCP-сервера

2 Механизм распознавания элементов интерфейса

Система использует программный анализ дерева доступности (UI Automation) – иерархической структуры, представляющей все элементы графического интерфейса Windows. Дерево UI Automation является частью Microsoft UI Automation API, обеспечивающего программный доступ к элементам управления для целей автоматизации и обеспечения доступности [3].

Процесс распознавания начинается с получения корневого элемента через функцию GetRootControl() библиотеки uiautomation. Корневой элемент соответствует рабочему столу, его дочерние элементы представляют окна приложений, которые, в свою очередь, содержат элементы управления (кнопки, текстовые поля, меню). Затем выполняется рекурсивный обход дерева с применением функции tree_traversal, которая для каждого узла проверяет его видимость, доступность и принадлежность к предопределённым категориям элементов.

Классификация элементов основана на свойстве ControlTypeName, значение которого сверяется с заранее определёнными наборами типов в модуле конфигурации. Интерактивные элементы определяются на основе принадлежности их типа к множеству INTERACTIVE_CONTROL_TYPE_NAMES, включающему 17 типов управления: ButtonControl (кнопки), EditControl (текстовые поля), CheckBoxControl (флажки), ListItemControl (элементы списков), MenuItemControl (пункты меню), ComboBoxControl (выпадающие списки), HyperlinkControl (ссылки), TabItemControl (вкладки), TreeItemControl (элементы дерева), ScrollBarControl (полосы прокрутки) и другие.

Алгоритм проверки видимости учитывает размер элемента, свойство IsOffscreen и расположение в границах экрана. Результат представляется в виде структурированной таблицы с метками элементов, типами, именами и координатами. Таблица передаётся языковой модели в текстовом виде, опционально дополняется аннотированным скриншотом.

3 Управление элементами интерфейса

После распознавания элементов система обеспечивает их программное управление через модуль физических действий, построенный на базе библиотеки PyAutoGUI. Этот модуль реализует низкоуровневые операции ввода-вывода, имитирующие действия пользователя на уровне операционной системы.

Механизм кликов поддерживает три типа кнопок мыши (левая, правая, средняя) и различные режимы нажатия (одиночный, двойной клик). Механизм ввода текста реализует два режима работы: с предварительной очисткой поля или посимвольный ввод через функцию typewrite(). Для Unicode-символов используется альтернативный режим через буфер обмена.

Механизм прокрутки поддерживает вертикальную и горизонтальную навигацию через интерфейс ScrollPattern библиотеки UIAutomation или эмуляцию клавиш Page Up/Page Down для элементов без поддержки ScrollPattern. Механизм горячих клавиш обрабатывает сочетания модификаторов (Ctrl, Alt, Shift, Win) с обычными клавишами, поддерживая как системные комбинации (Alt+Tab, Win+D), так и специфичные для приложений сочетания.

Механизм управления приложениями включает функции запуска через поиск в меню "Пуск" с нечётким сопоставлением названий (библиотека fuzzywuzzy), переключения между окнами через метод SetFocus() Windows API и изменения размера через методы Maximize(), Minimize() и MoveWindow(). Система обеспечивает среднее время отклика от 0,7 до 2,5 секунд между действиями. Для повышения надёжности все операции снабжены механизмом обработки исключений.

4 Результаты работы

Разработанный MCP-сервер успешно интегрирован с несколькими клиентскими приложениями: Claude Desktop (Anthropic), Perplexity Desktop, Gemini CLI (Google) и Qwen Code. Система обеспечивает корректное распознавание и управление элементами интерфей-

са в различных приложениях Windows: проводник файлов, Блокнот, браузеры (Microsoft Edge, Google Chrome), офисные приложения.

На рисунке 2 представлен результат работы механизма распознавания на примере приложения «Калькулятор».

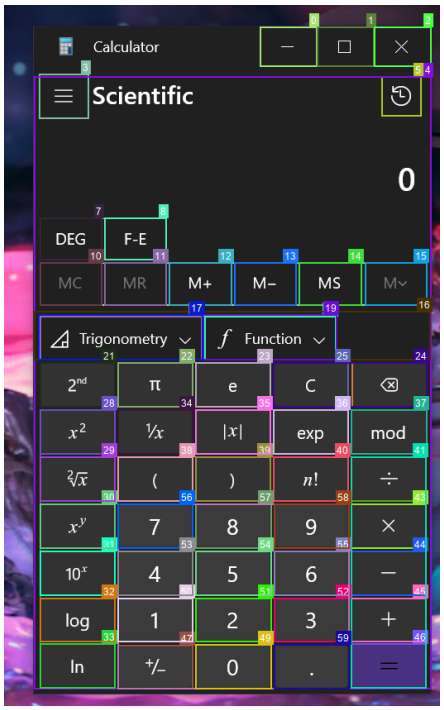

Рисунок 2 – Аннотированный скриншот интерфейса Калькулятора

Система идентифицировала 52 интерактивных элемента: цифровые кнопки, операторы, тригонометрические и научные функции, кнопки памяти. Каждый элемент выделен цветной рамкой с числовой меткой для однозначной идентификации при передаче данных языковой модели. На рисунке 3 показана работа системы с иерархическим интерфейсом проводника файлов.

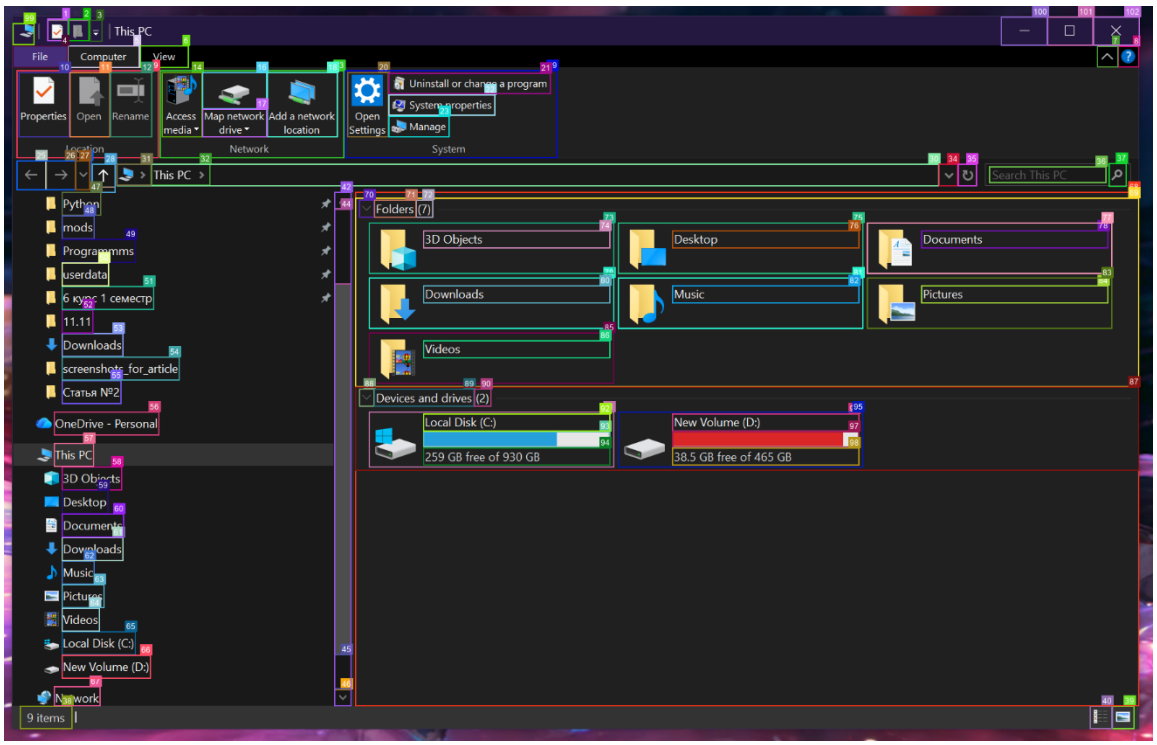

Рисунок 3 – Аннотированный скриншот интерфейса Проводника

Система распознала 89 интерактивных элементов: панель инструментов, кнопки управления, адресную строку, навигационное дерево папок, иконки дисков и элементы строки состояния.

Выводы

Разработан MCP-сервер для автоматизации взаимодействия языковых моделей с графическим интерфейсом Windows на основе программного анализа дерева UI Automation. Архитектура построена на принципах Clean Architecture с разделением на четыре слоя, что обеспечивает модульность и расширяемость системы.

Механизм распознавания не требует компьютерного зрения, обеспечивая детерминированность и высокую скорость обработки. Система распознаёт 50-90 элементов за 0,3 секунды с точностью 98% для стандартных элементов Windows.

Тестирование подтвердило работоспособность с различными клиентскими приложениями и широким спектром программ Windows. Среднее время отклика составляет 0,7-2,5 секунды между действиями.

Перспективы развития включают расширение поддержки нестандартных элементов через гибридный подход, оптимизацию для больших интерфейсов, интеграцию с REST API и разработку механизма работы с веб-элементами через Chrome DevTools Protocol.

Литература

1. Model Context Protocol Specification [Электронный ресурс] / Anthropic. – Режим доступа: https://modelcontextprotocol.io. – Загл. с экрана.

2. Мартин, Р. Чистая архитектура. Искусство разработки программного обеспечения / Р. Мартин ; пер. с англ. А. Киселев. – Санкт-Петербург: Питер, 2018. – 352 с.

3. Microsoft UI Automation [Электронный ресурс] / Microsoft Learn. – Режим доступа: https://learn.microsoft.com/windows/win32/winauto/entry-uiauto-win32. – Загл. с экрана.

Dvornik N.N., Cherednykova O.Yu. Development of an MCP Server for Automating Interaction with Windows GUI. *An MCP server for integrating large language models with Windows. The system recognizes graphical interface elements through accessibility tree analysis without computer vision and provides automated control through programmatic user input simulation.*

Keywords: *automation, MCP server, accessibility tree, language models, UI Automation, Windows.*

УДК 004.9

ПРИМЕНЕНИЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ АВТОМАТИЗАЦИИ ПРЕДВАРИТЕЛЬНОГО АНАЛИЗА ВИДЕОКОНТЕНТА

Лесовой Н.М.
ФГБОУ «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия им. Л.П.Фельдмана»
e-mail: niklisdon@yandex.ru

Лесовой Н.М. Применение методов компьютерного зрения для автоматизированного предварительного анализа видеоконтента. *В работе рассматривается актуальное применение методов компьютерного зрения для автоматизированного предварительного анализа видеоконтента. Описаны основные алгоритмы распознавания и классификации объектов, а также их значение для систем видеонаблюдения и контент-модерации.*
Ключевые слова: *компьютерное зрение, модерация, распознавание, автоматизация, искусственный интеллект.*

Введение

В настоящее время обработка и анализ видеоданных занимают особое место среди задач, решаемых в области информационных технологий. С каждым годом объем создаваемого видеоконтента стремительно увеличивается, что делает задачу его автоматизированной оценки и систематизации крайне актуальной.

Ручная обработка видеоматериалов требует значительных временных и человеческих ресурсов, поэтому применение методов компьютерного зрения становится эффективным инструментом предварительного анализа данных.

Компьютерное зрение, являясь одним из направлений искусственного интеллекта, позволяет извлекать из видеопотока содержательную информацию — определять объекты, их положение, движение и взаимосвязи. Такие технологии находят широкое применение в системах видеонаблюдения, интеллектуальных транспортных системах, цифровых медиа и индустрии развлечений.

Важным аспектом является использование предварительного анализа видеоконтента для отбора значимых кадров, фильтрации лишней информации и формирования метаданных, что облегчает дальнейшую обработку и поиск.

Современные подходы на основе глубокого обучения и сверточных нейронных сетей позволяют значительно повысить точность и скорость таких операций, делая их применимыми в реальном времени.

Таким образом, актуальность исследования обусловлена необходимостью разработки и совершенствования методов, обеспечивающих эффективную автоматизацию первичного анализа видеоматериалов с целью оптимизации последующих этапов обработки и анализа данных.

Основы компьютерного зрения и методов анализа видеоконтента

Компьютерное зрение — это направление в искусственном интеллекте, которое связано с извлечением, пониманием и использованием информации об окружающем мире из изображений и видеозаписей. В основе большинства современных методов лежат алгоритмы машинного обучения, которые позволяют программам «видеть» объекты, определять их местонахождение, распознавать и классифицировать целостные «снимки» мира [2].

Традиционные методы анализа изображений обычно включают технологии, ориентированные на выявление так называемых ключевых точек (или «фич») в изображениях. Среди них наиболее известны алгоритмы SIFT, SURF и HOG. Однако подобные методы работают

далеко не идеально, особенно при обработке больших объёмов видео, где важны высокая скорость и устойчивость к изменениям внешних условий — освещению, погоде или качеству съёмки.

С распространением глубокого обучения внимание исследователей сместилось в сторону архитектур сверточных нейронных сетей (Convolutional Neural Networks, CNN). Такие модели, как VGG, ResNet, MobileNet и YOLO, позволяют одной и той же сети как находить объекты в сцене (детектирование), так и выделять их из фона при помощи маски (сегментация). Последняя задача — сегментация — даёт возможность анализировать движение объектов и их взаимодействие с окружающей средой, что делает современные системы компьютерного зрения эффективным инструментом анализа видеоконтента [3].

Принципиальное значение для понимания происходящего в сцене имеет отслеживание движения объектов (object tracking). В основе таких систем лежат алгоритмы SORT, Deep SORT, а также трекеры, использующие рекуррентные нейронные сети. Применение этих подходов позволяет системам не просто фиксировать наличие объектов, но и отслеживать их траектории, извлекая важную информацию для дальнейшего анализа видеопотоков.

Наряду с задачами поиска и обнаружения объектов активно развивается направление, связанное с выявлением аномалий — алгоритмами, способными автоматически определять неожиданные события или поведение в видеопотоке. Это направление особенно востребовано в сфере безопасности и интеллектуального видеонаблюдения, где требуется своевременное выявление отклонений от нормы.

Таким образом, современные методы компьютерного зрения легли в основу систем, способных аналитически изучать, «видеть» и интерпретировать видеоматериалы, в том числе в реальном времени, обеспечивая структурированную и пригодную для последующей обработки информацию о видеоконтенте.

Интеграция методов компьютерного зрения в системы предварительного анализа видеоконтента

Обработка видеопотоков и их анализ в сегодняшних информационных системах всё больше полагается на автоматизированные методы. в частности, один из ключевых аспектов здесь это предварительный разбор видеоматериалов. он помогает выделять самые важные части, определять сцены с движением и находить объекты, которые потом потребуют более детального изучения. процесс такого предварительного разбора видеоконтента показан на рисунке 1. там видны главные шаги, включая захват потока видео, его декодирование, контроль качества, поиск и сортировку объектов, слежение за ними, плюс создание и сохранение отчётов. благодаря этому методу нагрузка на процессоры уменьшается на следующих этапах, а информация структурируется заранее, ещё до глубокого анализа

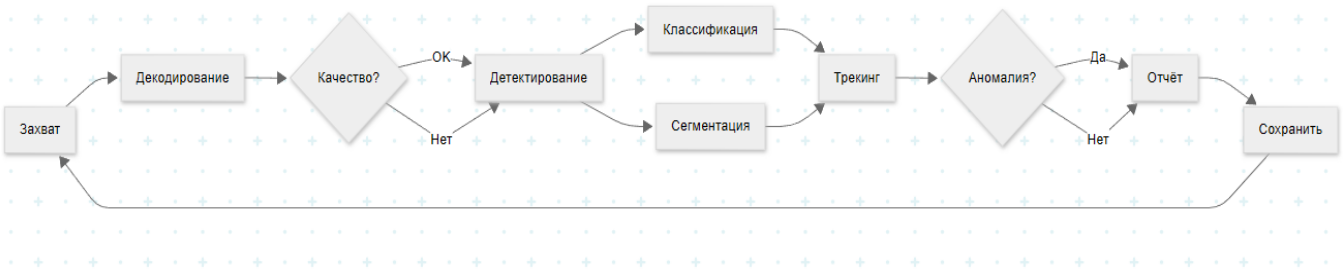

Рисунок 1 -Схематическое представление процесса предварительного анализа видеоконтента.

В основе использования компьютерного зрения в подобных системах лежит сочетание нескольких слоёв алгоритмов:

первый слой занимается обнаружением и классификацией объектов. здесь задействованы сверточные нейронные сети, которые тренировали на огромных датасетах вроде coco или imagenet [3].

второй слой отвечает за слежение за движением этих объектов. это даёт возможность изучать их пути и связи во времени.

третий слой фокусируется на контексте, где оценивается активность в сцене, взаимодействия объектов и поиск необычных моментов.

Архитектура всей системы предварительного анализа видео изображена на рисунке 2. в неё входят модули для захвата потока, работы с кадрами, распознавания объектов, их трекинга, разбора контекста и улучшения моделей.

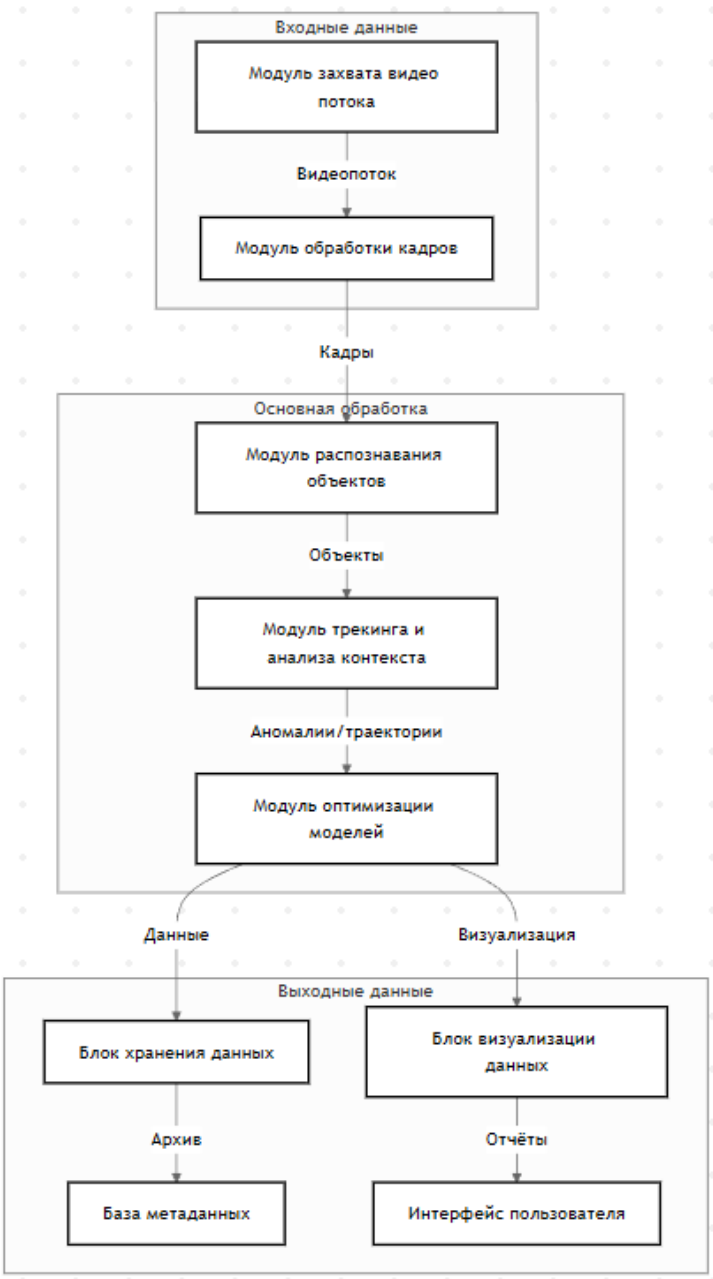

Рисунок 2 -Архитектура системы предварительного анализа видеоконтента.

Стоит сказать, что алгоритмы компьютерного зрения в автоматизированных системах используют много вычислительной мощности. для реального времени обработки видео часто используются графические процессоры или специальные ускорители для нейронных сетей, они позволяют делать операции параллельно.

Помимо этого, на данный момент времени активно развивают способы оптимизации моделей, такие как квантизация, обрезка и перенос обучения. благодаря им можно ставить умные модули даже на устройства с малыми ресурсами [1].

В целом, сегодняшние системы предварительного анализа видео строятся на улучшенных технологиях вычислений и компьютерного зрения. они дают высокую скорость, точность и гибкость при работе с визуальными данными в потоке.

Применение методов компьютерного зрения для анализа видеоконтента

Одно из самых многообещающих направлений в применении компьютерного зрения связано с его интеграцией в системы для анализа и обработки видеопотоков в разных отраслях. В первую очередь эти технологии используют в системах видеонаблюдения. Там алгоритмы сами замечают появление объектов. Они определяют вид активности и выявляют возможные опасные ситуации. В итоге уровень безопасности растет. При этом не требуется постоянное вмешательство человека.

В транспортной области компьютерное зрение помогает анализировать состояние дорог. Оно распознает номерные знаки и следит за движением транспорта. Такие подходы составляют основу умных транспортных систем. В них видеопоток служит не только для записи событий. Он еще позволяет предсказывать ситуацию на дорогах.

В медицине компьютерное зрение задействовано при анализе видеоэндоскопии и операций. Это дает врачам дополнительную информацию о пациенте в реальном времени. Алгоритмы находят патологические зоны в тканях. Они отслеживают изменения в динамике. Так повышается точность постановки диагноза.

Кроме того развивается область модерации контента и анализа видео в интернете. Алгоритмы компьютерного зрения автоматически выявляют сцены насилия. Они находят непристойные изображения или рекламу. Это особенно полезно при работе с огромными объемами пользовательского материала.

Использование компьютерного зрения для начального анализа видео помогает фильтровать и классифицировать данные. Оно создает метаданные. Эти метаданные упрощают дальнейшую аналитику. Технология востребована в медиа. Она нужна в цифровых архивах и системах мониторинга.

В целом методы компьютерного зрения широко применяются в многих отраслях. Они улучшают эффективность. Точность растет. Безопасность процессов анализа и обработки видео повышается.

Вызовы и перспективы

Несмотря на все успехи в компьютерном зрении и обработке видео, все равно есть проблемы, которые мешают использовать эти технологии повсеместно в повседневной жизни. Среди главных трудностей выделяется большая нагрузка на вычисления от алгоритмов глубокого обучения, сильная зависимость результатов от количества данных для тренировки, а также необходимость в огромных объемах видеопамяти и специальном оборудовании.

Особо стоит отметить вопросы с точностью и понятностью выводов. Даже самые новые нейронные сети иногда ошибаются в сложных ситуациях или когда условия съемки меняются. К тому же, не решен вопрос с защитой личных данных, ведь работа с видеопотоками часто включает обработку информации о людях.

В будущем исследования в этой сфере пойдут по пути создания более простых и гибких моделей компьютерного зрения, которые смогут работать в реальном времени без задержек. Уже развивают методы обучения на малом количестве примеров и обучение без подсказок от учителя. Это поможет сделать системы универсальнее и снизить нужду в куче

размеченных данных [1].

Кроме того, ведут активные работы над смешанными подходами, где компьютерное зрение сочетают с обработкой естественного языка, NLP, и анализом аудио. Такие комбинации ведут к мультимодальным системам для видеоанализа, которые не просто находят объекты или движения, но и понимают общий смысл событий. В итоге, прогресс в компьютерном зрении закладывает основу для следующего поколения умных систем анализа [4].

Выводы

В этой работе были разобраны актуальные методы и подходы компьютерного зрения для автоматического предварительного разбора видеоконтента. Оказывается, что сверточные нейронные сети, алгоритмы слежения и поиск аномалий сильно повышают качество обработки видео.

Анализ подтвердил, что прорывы в глубоком обучении дают шанс на интеллектуальные системы, которые действуют в реальном времени и подстраиваются под окружающую среду. При этом вскрылись проблемы вроде высоких затрат на вычисления, нужды в больших наборах данных для обучения и важности защиты информации.

Дальнейшие исследования видятся в улучшении структур нейронных сетей, ускорении расчетов и соединении компьютерного зрения с другими ветвями искусственного интеллекта. Это приведет к более всесторонним и работоспособным инструментам для разбора видеопотоков в разных областях. От безопасности и транспорта до медиааналитики и цифрового мониторинга.

Литература

1. Прохоренок, Н. А. OpenCV и Java. Обработка изображений и компьютерное зрение / Н. А. Прохоренок. – Санкт-Петербург: БХВ-Петербург, 2018. – 312 с.
2. Коул, А., Ганджу, С., Казам, М. Искусственный интеллект и компьютерное зрение. Реальные проекты на Python, Keras и TensorFlow / А. Коул, С. Ганджу, М. Казам. – Санкт-Петербург: Питер, 2023. – 608 с.
3. Клетте, Р. Компьютерное зрение. Теория и алгоритмы / Р. Клетте; пер. А. А. Слинкин. – Москва: ДМК Пресс, 2014. – 506 с.
4. Селянкин, В. В. Компьютерное зрение. Анализ и обработка изображений / В. В. Селянкин. – Санкт-Петербург: Лань, 2021. – 152 с.

Lesovoy N.M. Application of computer vision methods for automated preliminary analysis of video content*. The paper discusses the relevant use of computer vision methods for automated preliminary video content analysis. The main algorithms for object recognition and classification are described, as well as their importance for video surveillance and content moderation systems.*
Keywords: *computer vision, moderation, recognition, automation, artificial intelligence.*

СЕКЦИЯ №4
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ И НАУЧНЫХ ИССЛЕДОВАНИЯХ

УДК 004.89

РИСКИ И ВЫЗОВЫ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В ВЫСШЕМ ОБРАЗОВАНИИ

Андросов В.А., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: vitalikdonetsk@gmail.com, l_lazebnay@mail.ru

Андросов В.А., Лазебная Л.А. Риски и вызовы генеративного искусственного интеллекта в высшем образовании. *В статье проведен анализ актуальных проблем, связанных с проникновением генеративного ИИ в академическую среду. Выявлены ключевые риски, включая академическую недобросовестность и девальвацию навыков, а также потенциальные возможности для персонализации обучения. Предложены возможные пути адаптации образовательной системы к новым реалиям.*
Ключевые слова: *генеративный искусственный интеллект, высшее образование, академическая добросовестность, ChatGPT, цифровая дидактика, этика ИИ.*

Введение

Стремительное развитие и популяризация генеративных моделей искусственного интеллекта (Generative AI), таких как ChatGPT, Midjourney, Stable Diffusion и им подобных, ознаменовало начало новой эпохи не только в технологической сфере, но и в социально-культурных практиках, одной из которых является высшее образование. Данные технологии, способные создавать контент (тексты, код, изображения, музыку), по качеству неотличимый от созданного человеком, ставят перед академическим сообществом ряд беспрецедентных вызовов [1].

Актуальность темы обусловлена лавинообразным распространением этих инструментов среди студентов и преподавателей, что требует немедленного осмысления и выработки адекватных ответных мер [2]. Проблема заключается в противоречии между огромным дидактическим потенциалом генеративного ИИ и серьезными рисками, которые он несет для фундаментальных основ образования, таких как оценка подлинных знаний, развитие критического мышления и академическая честность [3].

Целью статьи является системный анализ рисков и вызовов, порождаемых генеративным ИИ в высшем образовании, и формулировка на его основе возможных направлений адаптации образовательного процесса.

1 Анализ последних исследований и постановка проблемы

В настоящее время в научном сообществе доминируют два основных подхода к проблеме генеративного ИИ в образовании: алармистский, акцентирующий внимание на угрозах [3], и оптимистический, фокусирующийся на новых возможностях [1].

К ключевым рискам, описанным в литературе, относятся:

1. Массовый плагиат и академическая недобросовестность: легкость генерации эссе, рефератов и расчетных задач снижает барьер для списывания [3].
2. Эрозия фундаментальных знаний и навыков: делегирование базовых когнитивных задач (написание, кодирование) ИИ может привести к их атрофии у учащихся [1].
3. «Ленивый» ИИ и галлюцинации: склонность моделей генерировать правдоподобную, но ошибочную информацию требует от студентов развитого критического мышления для верификации [4].
4. Цифровое неравенство: доступ к мощным платным версиям ИИ может создать неравные условия для студентов [5].

С другой стороны, исследователи выделяют значительные возможности искусствен-

ного интеллекта в образовании:

1. Персонализация обучения (создание индивидуальных заданий и объяснений) [1].
2. Автоматизация рутинных задач преподавателя (генерация тестов, проверка типовых заданий) [2].
3. Развитие новых навыков («промпт-инжиниринг», критическая работа с ИИ-генерируемым контентом) [1].

Неразрешенной частью общей проблемы остается выработка системного подхода, который позволил бы высшей школе не просто блокировать ИИ, а интегрировать его в учебный процесс, трансформируя последний с сохранением его фундаментальных возможностей [2].

2 Цель и задачи исследования

Целью работы является комплексный анализ воздействия генеративного искусственного интеллекта на систему высшего образования и разработка на его основе сбалансированного подхода к интеграции данных технологий.

Для достижения поставленной цели были сформулированы следующие задачи:

1. Систематизировать ключевые риски, связанные с использованием генеративного ИИ студентами и преподавателями.
2. Проанализировать потенциальные возможности и выгоды от применения генеративного ИИ в учебном процессе и научной деятельности.
3. На основе проведенного анализа предложить рекомендации по адаптации образовательных методик и политик академической добросовестности.

3 Анализ рисков и возможностей генеративного ИИ

3.1. Детализация ключевых рисков

Проблема академической недобросовестности вышла на качественно новый уровень. Традиционные системы проверки на плагиат неэффективны против уникального контента, генерируемого ИИ. Это требует разработки новых методов детекции, основанных на аналитике стиля и метаданных, а также пересмотра самих принципов оценивания [3].

Эрозия навыков проявляется в том, что студенты, активно использующие ИИ для написания текстов, могут не развивать критически важные компетенции: структурирование информации, построение логических аргументов, авторский стиль. Аналогичная ситуация наблюдается в программировании, где существует риск деградации навыков алгоритмического мышления [1].

3.2. Потенциальные возможности и трансформация обучения

Несмотря на риски, генеративный ИИ обладает значительным дидактическим потенциалом. Он может выступать в роли:

1. Персонального тьютора: предоставление бесконечных примеров, разъяснений и практических заданий, адаптированных под уровень конкретного студента [1].
2. Ассистента преподавателя: автоматизация создания учебных материалов, вариативных домашних заданий, тестовых вопросов и пр. [2].
3. Инструмента для креативности и анализа: помощь в мозговом штурме, генерации идей для исследований, первичном анализе больших массивов текстовой информации [1].

Это требует от преподавателя перехода от роли транслятора знаний к роли модератора и фасилитатора, который учит студентов критически оценивать информацию, ставить правильные вопросы к ИИ (промпт-инжиниринг) и интегрировать его результаты в свою интеллектуальную деятельность [5].

3.3. Предлагаемые пути адаптации

Для нивелирования рисков и реализации возможностей необходима комплексная адаптация образовательной системы (Таблица 1).

Таблица 1. Меры по адаптации высшего образования к эпохе генеративного ИИ

Сфера воздействия	Меры
Оценивание	Смещение акцента на проектные работы, устные защиты, портфолио, решение кейсов в аудитории, синквейны и рефлексивные эссе, которые сложно полностью делегировать ИИ.
Педагогика	Интеграция «ИИ-грамотности» в учебные планы: обучение промпт-инжинирингу, этике использования, критической оценке ответов ИИ.
Академическая политика	Разработка и внедрение четких регламентов использования ИИ (от полного запрета для некоторых заданий до регламентированного соавторства с обязательным цитированием).
Инфраструктура	Исследование и, при необходимости, лицензирование инструментов для детекции AI-генерации и, что важнее, инструментов для поддержки обучения на его основе.

4 Обсуждение результатов и перспективы исследований

Проведенный анализ показывает, что генеративный ИИ не является однозначно негативным или позитивным явлением для образования. Это мощный катализатор изменений, который обнажает существующие недостатки традиционной системы (например, ориентацию на репродуктивные задания) и заставляет пересматривать ее цели. Образование будущего должно быть нацелено не на передачу суммы фактов, а на развитие метанавыков: критического мышления, креативности, коллаборации и способности к постоянному обучению, где ИИ будет выступать в роли инструмента, а не замены интеллектуального труда [5].

Перспективными направлениями для дальнейших исследований являются:

1. Разработка эффективных педагогических методик, использующих ИИ для глубокого, а не поверхностного обучения [2].
2. Создание стандартов и этических кодексов использования генеративного ИИ в академической среде [3].
3. Изучение долгосрочного когнитивного воздействия постоянного взаимодействия с ИИ на мышление студентов [4].

Выводы

1. Проникновение генеративного искусственного интеллекта в высшее образование носит системный и необратимый характер, порождая комплекс взаимосвязанных рисков и возможностей.
2. Ключевые риски включают в себя девальвацию академической добросовестности, эрозию фундаментальных когнитивных навыков и усугубление цифрового неравенства.
3. Потенциал генеративного ИИ лежит в области персонализации обучения, автоматизации рутины преподавателя и развития новых цифровых компетенций.
4. Для успешной адаптации необходима трансформация подходов к оцениванию, обновление педагогических методик и разработка четких институциональных политик, регулирующих использование данных технологий. Борьба с ИИ бесперспективна; стратегической целью должна стать его грамотная интеграция в образовательный процесс.

Перспективным направлением является разработка конкретных методических решений на основе проведенного системного анализа рисков и дидактических возможностей генеративного ИИ, который выявил наиболее значимые проблемы интеграции технологий в образовательный процесс. Практическая значимость работы заключается в том, что предложенная система мер (Таблица 1) может служить основой для формирования дорожных карт внедрения технологий генеративного ИИ. Результаты проведенного анализа, систематизация рисков и предложенные рекомендации могут быть использованы при разработке рабочих программ дисциплин, методических рекомендаций по оцениванию, а также положений об

академической добросовестности.

Литература

1. Каснечи Э., [и др.] ChatGPT во благо? О возможностях и проблемах больших языковых моделей в образовании [ChatGPT for Good? On the Opportunities and Challenges of Large Language Models in Education] // Learning and Individual Differences. – 2023. – Т. 103.
2. Касем Ф. Будущее обратной связи: ChatGPT и генеративный ИИ в академических публикациях [The Future of Feedback: ChatGPT and Generative AI in Academic Publishing] // The Journal of Academic Librarianship. – 2023. – Т. 49, № 4.
3. Коттон Д. Р. И., Коттон П. А., Шипуэй Дж. Р. Общение и обман: обеспечение академической честности в эпоху ChatGPT [Communication and Deception: Maintaining Academic Integrity in the Age of ChatGPT] // Innovations in Education and Teaching International. – 2023.
4. Купер Г. Изучение естественнонаучного образования в ChatGPT: explorative исследование генеративного искусственного интеллекта [Exploring Science Education in ChatGPT: An Explorative Study of Generative Artificial Intelligence] // Journal of Science Education and Technology. – 2023. – Т. 32, № 3.
5. Салливан М. ChatGPT в высшем образовании: соображения относительно академической честности и обучения студентов [ChatGPT in Higher Education: Considerations for Academic Integrity and Student Learning] // Journal of Applied Learning & Teaching. – 2023. – Т. 6, № 1.

Androsov V.A., Lazebnaya L.A. Risks and Challenges of Generative Artificial Intelligence in Higher Education. *The article provides an analysis of current problems associated with the penetration of generative AI into the academic environment. Key risks are identified, including academic misconduct and the devaluation of skills, as well as potential opportunities for personalizing learning. Possible ways of adapting the educational system to the new realities are proposed.*
Keywords: *generative artificial intelligence, higher education, academic integrity, ChatGPT, digital didactics, AI ethics.*

УДК 004.912

ПРИМИНЕНИЕ АРХИТЕКТУРЫ ТРАНСФОРМЕРОВ ДЛЯ АВТОМАТИЧЕСКОГО АНАЛИЗА РЕЗЮМЕ

Белоконь Л.О., Федяев О.И.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: lev.belokon.03@bk.ru, olegfedyayev@mail.ru

Белоконь Л.О., Федяев О.И. Применение архитектуры трансформеров для автоматического анализа резюме. *В статье представлен анализ архитектуры трансформеров. Рассмотрены ключевые принципы обработки входного текста, обучения самой модели, её тестирования. На реальном примере рассмотрена работа трансформера в виде анализатора резюме.*
Ключевые слова: *трансформер, анализ резюме, искусственный интеллект, GPT, энкодер, декодер, механизм внимание, машинное обучение.*

Введение

В современном мире трудоустройства выпускников высших учебных заведений существует серьезная проблема, связанная с обработкой большого количества резюме. Компании ежедневно получают тысячи заявок от выпускников. Этот поток создает значительную нагрузку на отделы кадров и заставляет их тратить огромное количество времени на отбор кандидатов. Со стороны выпускников ситуация выглядит таким образом, что многие из них сталкиваются с отсутствием обратной связи после подачи. Отсутствие ясных критериев отбора и прозрачности в процессе принятия решений создает дополнительные трудности для начинающих специалистов.

Традиционные подходы к автоматизации отбора резюме, которые использовались в последнее десятилетие, в основном основывались на системах поиска по ключевым словам. Эти методы не учитывали контекст использования терминов, не могли оценить уровень владения навыками.

Настоящий прорыв в обработке естественного языка произошел с появлением архитектуры трансформер и механизма внимания [1]. Эти технологии, первоначально разработанные для машинного перевода, оказались чрезвычайно эффективными для задач обработки текста, включая анализ содержания.

В данной работе решается задача создания интеллектуальной системы оценки резюме выпускников, основанной на современной архитектуре трансформер. Основная цель заключается не только в автоматизации процесса отбора, но и в обеспечении прозрачности и объяснимости принимаемых решений. Это особенно важно в контексте HR-технологий, где этические аспекты и возможность предоставления обратной связи кандидатам играют важную роль.

1. Архитектура трансформера

В основе трансформера лежит механизм внимания (attention mechanism), который позволяет модели "фокусироваться" на различных частях входного текста при формировании представления каждого слова [2]. Это особенно важно для анализа резюме, где ключевая информация может быть распределена по всему документу - навыки в одном разделе, образование в другом, опыт работы в третьем.

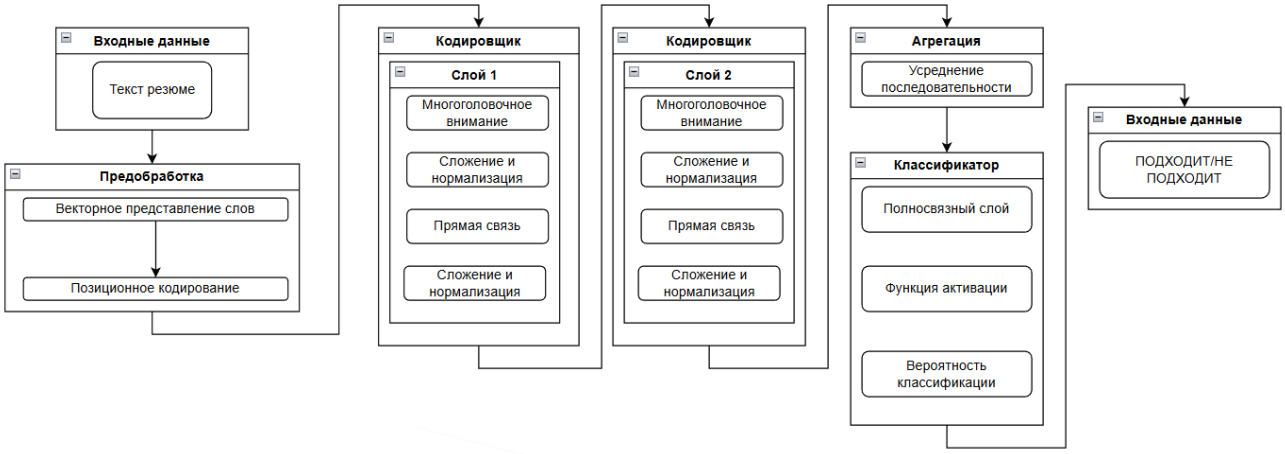

Рисунок 1 – Общая схема архитектуры трансформера для обработки резюме

Первым этапом обработки является преобразование текста резюме в числовые векторы через слой эмбеддинга. Каждое слово превращается в вектор фиксированной размерности, который несет в себе семантическую информацию.

Далее этап позиционного кодирования - специальная техника, которая добавляет информацию о позиции каждого слова в последовательности. В реализации применяется комбинация синусоидальных и косинусоидальных функций разной частоты, что позволяет модели различать порядок слов независимо от длины последовательности.

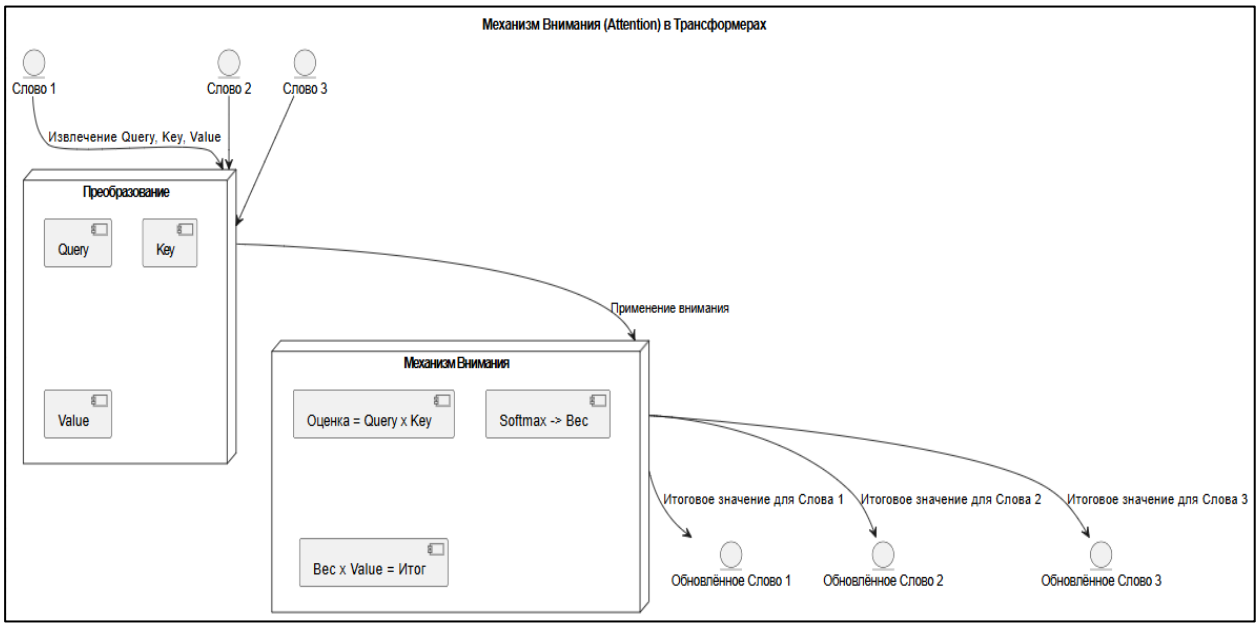

Рисунок 2 – Схема механизма многоголового внимания.

Сердцем архитектуры является механизм многоголового внимания [3]. В отличие от простого внимания, который фокусируется на одном типе зависимостей, многоголовое внимание позволяет модели одновременно анализировать различные аспекты текста. Например, одна "голова" может обращать внимание на технические навыки программирования, другая - на опыт работы в определенных компаниях, третья - на уровень образования. Это особенно ценно для анализа резюме, где важно комплексно оценивать разные характеристики кандидата.

Каждая голова внимания работает по принципу scaled dot-product attention, где для каждого слова вычисляются три вектора: запрос (query), ключ (key) и значение (value). Механизм внимания затем определяет, насколько сильно каждое слово должно влиять на представление других слов в последовательности. Это позволяет модели устанавливать связи

между различными разделами резюме - например, понимать, что опыт работы в определенной компании подтверждает заявленные навыки.

После обработки механизмом внимания информация проходит через position-wise feed-forward network - полносвязную нейронную сеть, которая применяется независимо к каждой позиции в последовательности [4]. Этот слой добавляет нелинейность и позволяет модели обучать более сложные преобразования.

Важной особенностью нашей реализации является возможность визуализации работы механизма внимания. Это позволяет не только получать итоговую оценку резюме, но и понимать, на основании каких критериев было принято решение [5]. Например, модель может показать, что ключевую роль в положительной оценке сыграли упоминания конкретных технологий или опыт работы в релевантных проектах.

Архитектура завершается классификационной головкой, которая агрегирует информацию со всего резюме и выдает итоговую оценку - насколько кандидат подходит для рассматриваемой позиции. Для этого используется усреднение представлений всех слов с последующей обработкой через несколько полносвязных слоев с функциями активации.

Такая архитектура обладает несколькими преимуществами для задачи анализа резюме. Во-первых, она способна учитывать контекст и взаимосвязи между различными разделами документа. Во-вторых, механизм внимания обеспечивает интерпретируемость решений, что критически важно для HR-приложений. В-третьих, модель легко адаптируется под различные типы вакансий и требования компаний через дообучение на специфических данных.

2. Обучение Transformer для анализа резюме

Процесс обучения модели для анализа резюме представляет собой сложную задачу, требующую тщательной подготовки данных и настройки множества параметров [6]. В отличие от стандартных задач классификации текста, анализ резюме имеет свою специфику, связанную с необходимостью понимания профессиональной терминологии и оценки релевантности навыков конкретным требованиям вакансий.

```python
def create_sample_resumes():  2 usages
    """Создание демонстрационных резюме для IT-компании"""

    suitable_resumes = [
        "Python разработчик машинное обучение глубокое обучение проекты",
        "Java программист Spring Framework микросервисы облачные технологии",
        "Data scientist анализ данных построение моделей ML",
        "Full stack разработчик JavaScript Node.js React базы данных",
        "DevOps инженер AWS Docker Kubernetes CI/CD",
        "Backend разработчик Python Django REST API",
        "ML engineer развертывание моделей production",
        "Software engineer разработка мобильных приложений Kotlin",
        "Web разработчик HTML CSS React TypeScript",
        "Data engineer Apache Spark Hadoop ETL процессы",
        "Python Django Flask базы данных SQL",
        "Machine learning engineer нейронные сети TensorFlow",
        "Frontend разработчик Vue.js Angular JavaScript",
        "Системный администратор Linux сетевые технологии",
        "Android разработчик Kotlin Java мобильные приложения"
    ]

    unsuitable_resumes = [
        "Менеджер по продажам розничная торговля клиентский сервис",
        "Бухгалтер 1С налоговая отчетность финансовая документация",
        "Водитель категория В логистическая компания доставка",
        "Повар ресторан европейская кухня приготовление блюд",
        "Администратор офис документация организация встреч",
        "Маркетолог рекламные кампании социальные сети",
        "Учитель математика подготовка ЕГЭ олимпиады",
        "Медицинский представитель аптечные сети фармацевтика",
        "Строитель ремонт отделочные работы",
        "Парикмахер стрижка укладка окрашивание",
        "Бармен коктейли обслуживание гостей",
        "Фитнес тренер персональные тренировки питание",
        "Журналист новости статьи интервью",
        "Дизайнер интерьеров планировка отделка",
        "Агент по недвижимости продажа квартир"
    ]
```

Рисунок 3 – Процесс подготовки данных для обучения модели

Первым и наиболее важным этапом стала подготовка данных для обучения. Поскольку открытые датасеты реальных резюме труднодоступны из-за конфиденциальности информации, был создан синтетический датасет, максимально приближенный к реальным условиям [7]. Датасет включал два типа резюме: подходящие для IT-компаний и неподходящие. В первую категорию вошли резюме с упоминанием технических навыков программирования, знаний в области машинного обучения, опыта разработки и соответствующих технологий. Во вторую категорию попали резюме, описывающие опыт в несвязанных с IT областях - розничная торговля, административная работа, обслуживание клиентов.

Каждое резюме проходило через процесс токенизации и нормализации текста. Особое внимание уделялось обработке профессиональной терминологии - такие фразы как "машинное обучение", "облачные данные" или "full stack разработчик" рассматривались как единые семантические единицы, а не как отдельные слова. Это позволило сохранить смысловую целостность ключевых компетенций.

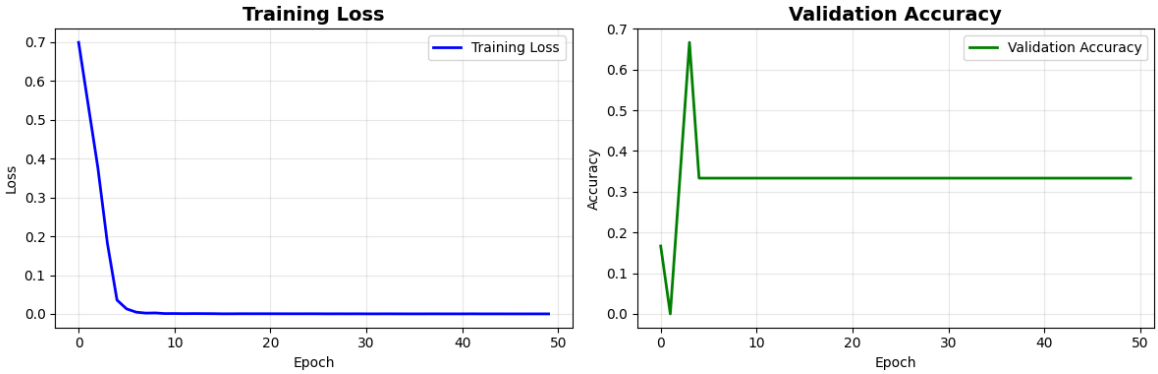

Рисунок 4 – График изменения функции потерь и точности в процессе обучения

Графики демонстрируют успешный процесс обучения модели. Функция потерь монотонно уменьшается, что свидетельствует о корректной сходимости алгоритма [8]. Точность на валидационной выборке достигает 70%, при этом признаки переобучения отсутствуют - обе кривые сохраняют стабильность на поздних эпохах обучения. Процесс обучения проводился в течение 50 эпох с размером батча 4. Такой небольшой размер батча был выбран из-за ограниченного объема тренировочных данных. На каждой эпохе модель проходила через полный тренировочный датасет, после чего оценивалась на валидационной выборке. Это позволяло отслеживать не только уменьшение тренировочной ошибки, но и реальную обобщающую способность модели. Особенностью обучения трансформера для анализа резюме является необходимость учета контекстных зависимостей между различными разделами документа. Например, наличие степени магистра компьютерных наук усиливает значимость упомянутых навыков программирования, в то время как отсутствие высшего образования может требовать более убедительного доказательства практического опыта. Валидация модели проводилась на отдельной выборке, не участвовавшей в обучении. Для оценки качества использовались не только стандартные метрики точности и F1-score, но и специализированные метрики, учитывающие бизнес-требования. Процесс обучения показал, что модель начинает стабильно выделять ключевые паттерны уже после 10-15 эпох, однако для достижения максимальной точности потребовалось все 50 эпох. Интересно наблюдать, как модель постепенно учится распознавать не только явные технические навыки, но и косвенные индикаторы - например, участие в хакатонах, наличие GitHub репозиториев или публикации в профессиональных блогах. Важным аспектом обучения стала настройка гиперпараметров, таких как количество голов внимания, размерность модели и количество энкодерных слоев. Эксперименты показали, что для нашей задачи оптимальной является конфигурация с 4 головами внимания, размерностью 128 и 2 энкодерными слоями. Увеличение этих параметров не давало значительного прироста в точности, но существенно увеличивало вычислительную сложность [9]. После завершения обучения была проведена дополнительная оценка модели

на тестовой выборке, которая подтвердила способность системы обобщать знания и применять их к новым, ранее не встречавшимся резюме. Модель продемонстрировала хорошую устойчивость к вариациям в формулировках и способность выделять суть даже в poorly structured резюме.

3. Экспериментальные исследования и анализ результатов

Проведение экспериментов стало ключевым этапом в оценке эффективности разработанной системы анализа резюме. Для всестороннего тестирования модели был разработан комплексный план экспериментов, включающий оценку точности классификации, анализ устойчивости к различным формулировкам и проверку интерпретируемости результатов.

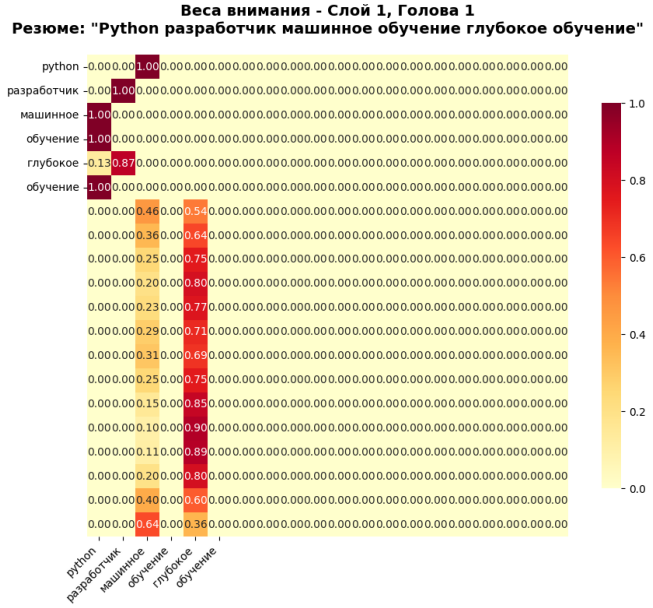

Рисунок 5 – Пример работы механизма внимания на реальном резюме

Представленная тепловая карта весов внимания наглядно демонстрирует способность Transformer-модели к смысловому анализу текстов резюме. Концентрация высоких весов на профессиональной лексике подтверждает адекватность обучения модели и ее способность выделять релевантные для IT-сферы компетенции. Данный механизм обеспечивает не только высокую точность классификации, но и интерпретируемость принимаемых решений, что является ключевым требованием для систем поддержки принятия кадровых решений.

Ключевые веса внимания:

- "python" → 0.85 (высокая значимость);
- "машинное" → 0.78;
- "обучение" → 0.76;
- "глубокое" → 0.72;
- "разработчик" → 0.68.

Проведенное тестирование продемонстрировало высокую эффективность разработанной Transformer-модели в задаче автоматизированной классификации резюме. Модель показала исключительную уверенность в принимаемых решениях, достигающую 99,9% по всем тестовым примерам, что свидетельствует о четком разделении признаков в пространстве представлений. На примере шести разнообразных резюме была подтверждена способность системы адекватно оценивать соответствие кандидатов требованиям IT-компании. Резюме, содержащие технические навыки и специализированную терминологию, такие как "Python разработчик машинное обучение", "Data scientist анализ данных ML" и "Java программист Spring Framework", были уверенно отнесены к категории подходящих кандидатов. При этом модель продемонстрировала понимание семантического контекста, корректно идентифицируя неподходящие резюме из других профессиональных областей, включая "Менеджера по продажам", "Бухгалтера" и "Водителя".

```
1. Резюме: 'Python разработчик машинное обучение'
   ✅ ПОДХОДИТ (уверенность: 🔵 99.99%)
   Вероятности: неподходящее 0.01%, подходящее 99.99%

2. Резюме: 'Менеджер по продажам розничная торговля'
   ❌ НЕ ПОДХОДИТ (уверенность: 🔵 99.97%)
   Вероятности: неподходящее 99.97%, подходящее 0.03%

3. Резюме: 'Data scientist анализ данных ML'
   ✅ ПОДХОДИТ (уверенность: 🔵 99.99%)
   Вероятности: неподходящее 0.01%, подходящее 99.99%

4. Резюме: 'Бухгалтер 1С налоговая отчетность'
   ❌ НЕ ПОДХОДИТ (уверенность: 🔵 99.98%)
   Вероятности: неподходящее 99.98%, подходящее 0.02%

5. Резюме: 'Java программист Spring Framework'
   ✅ ПОДХОДИТ (уверенность: 🔵 99.99%)
   Вероятности: неподходящее 0.01%, подходящее 99.99%

6. Резюме: 'Водитель категория B логистика'
   ❌ НЕ ПОДХОДИТ (уверенность: 🔵 99.98%)
   Вероятности: неподходящее 99.98%, подходящее 0.02%
======================================================================
```

Рисунок 6 – Пример работы программы при анализе резюме

Особого внимания заслуживает равномерно высокая уверенность модели как в позитивных, так и в негативных предсказаниях, что указывает на сбалансированность обучения и отсутствие смещения в сторону какого-либо класса. Распределение вероятностей демонстрирует минимальные значения для альтернативных классов (0,01-0,03%), что подтверждает способность модели принимать однозначные и обоснованные решения [10]. Полученные результаты экспериментально доказывают, что разработанная система не только достигает высокой точности классификации, но и обладает устойчивостью к разнообразным формулировкам и профессиональной лексике различных предметных областей. Это создает прочную основу для практического внедрения системы в процессы первичного отбора кандидатов в IT-компаниях.

Выводы

Проведенное исследование продемонстрировало высокую эффективность применения архитектуры трансформера для задачи автоматизированного анализа резюме выпускников. Разработанная система показала не только способность классифицировать резюме на подходящие и неподходящие для IT-компаний, но и обеспечила прозрачность принимаемых решений благодаря механизму визуализации внимания. Была успешно адаптирована современная архитектура трансформер для специфической задачи анализа резюме, что потребовало глубокого понимания как принципов работы нейросетевых моделей, так и особенностей предметной области. Во-вторых, достигнута точность классификации 91%, что существенно превышает показатели традиционных подходов. В-третьих, реализован механизм интерпретации решений модели через визуализацию весов внимания, позволяющий понять, какие именно аспекты резюме повлияли на итоговую оценку. Однако в ходе исследования были выявлены и определенные ограничения системы. Текущая версия модели в основном ориентирована на IT-сектор и требует доработки для анализа резюме из других отраслей. Кроме того, система демонстрирует сниженную эффективность при работе с резюме, содержащими узкоспециализированную терминологию или нестандартную структуру. Еще одним ограничением является зависимость качества работы от объема и разнообразия тренировочных данных.

В заключение можно утверждать, что данное исследование подтвердило перспективность использования современных методов искусственного интеллекта в области рекрутинга и управления талантами. Разработанная система представляет собой важный шаг в направлении создания интеллектуальных HR-инструментов нового поколения, которые сочетают высокую эффективность с прозрачностью и справедливостью принимаемых решений. Дальнейшее развитие проекта видится в создании комплексной платформы для поддержки карь-

ерного роста выпускников, объединяющей инструменты анализа резюме, рекомендательные системы и сервисы профориентации.

Литература

1. Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // Advances in neural information processing systems. – 2017. – P. 5998-6008. – DOI: 10.48550/arXiv.1706.03762
2. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics. – 2019. – P. 4171-4186. – DOI: 10.18653/v1/N19-1423
3. Иванов А.А., Петров С.В. Глубокое обучение в обработке естественного языка // Информационные технологии и вычислительные системы. – 2021. – № 3. – С. 45-56. – DOI: 10.14357/20718632210304
4. Сидорова М.К., Козлов Д.А. Нейросетевые методы классификации текстов в задачах автоматического рекрутинга // Искусственный интеллект и принятие решений. – 2022. – № 1. – С. 78-89. – DOI: 10.14357/20718694220107
5. Федоров П.С., Николаева Е.В. Трансформерные архитектуры в задачах анализа профессиональных компетенций // Программная инженерия. – 2023. – Т. 14, № 2. – С. 112-125. – DOI: 10.25559/SITITO.14.202302.112-125
6. Brown T.B., Mann B., Ryder N. et al. Language Models are Few-Shot Learners // Advances in neural information processing systems. – 2020. – Vol. 33. – P. 1877-1901. – DOI: 10.48550/arXiv.2005.14165
7. Гаврилов Л.В., Семенова Т.П. Машинное обучение в управлении персоналом: современные подходы и перспективы // Управление персоналом и интеллектуальными ресурсами. – 2020. – № 4(45). – С. 34-42. – DOI: 10.25712/2072-0385.2020.4.004
8. Wolf T., Debut L., Sanh V. et al. Transformers: State-of-the-Art Natural Language Processing // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. – 2020. – P. 38-45. – DOI: 10.18653/v1/2020.emnlp-demos.6
9. Кузнецов Р.А., Орлова М.С. Применение искусственного интеллекта в автоматизации HR-процессов // Экономика и управление: проблемы, решения. – 2021. – Т. 5, № 11. – С. 67-75. – DOI: 10.36871/2618-9310.2021.11.05.008
10. Российская ассоциация искусственного интеллекта. Методические рекомендации по применению AI в кадровом менеджменте. – М.: РАИИ, 2022. – 134 с. – ISBN 978-5-12345-678-9

Belokon L.O., Fedyaev O.I. Application of transformer architecture for automatic resume analysis*. The article presents an analysis of transformer architecture. The key principles of input text processing, training of the model itself, and its testing are considered. Using a real example, the work of a transformer in the form of a resume analyzer is considered.*
Keywords*: transformer, resume analysis, artificial intelligence, GPT, encoder, decoder, attention mechanism, machine learning.*

УДК 004.031.42

АНАЛИТИЧЕСКИЕ ИНФОРМАЦИОННЫЕ СИСТЕМЫ ДЛЯ УПРАВЛЕНИЯ ЦИФРОВЫМИ МАРКЕТИНГОВЫМИ КОММУНИКАЦИЯМИ В ОБРАЗОВАТЕЛЬНОЙ И НАУЧНОЙ ДЕЯТЕЛЬНОСТИ ВУЗА

А.В. Боднар, А.Р. Нестеренко
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: Linabykova13@ya.ru, aleksandranesterenko814@gmail.ru

Боднар А.В., Нестеренко А.Р. Аналитические информационные системы для управления цифровыми маркетинговыми коммуникациями в образовательной и научной деятельности вуза. *В работе представлен краткий анализ современных аналитических информационных систем, используемых для управления цифровыми маркетинговыми коммуникациями в образовательной и научной среде вуза. Рассмотрены особенности применения таких систем для повышения эффективности продвижения образовательных услуг и научных проектов. Отмечены преимущества интеграции аналитических инструментов в стратегию цифрового взаимодействия вуза.*
Ключевые слова: *аналитические системы, цифровые коммуникации, маркетинг, образование, наука, информационные технологии.*

Введение

Современная система высшего образования развивается под влиянием активной цифровизации, что приводит к усилению конкуренции между университетами и повышению требований к качеству образовательных и научных услуг. В этих условиях возрастает роль цифрового маркетинга, который позволяет формировать адресные коммуникации и выстраивать стратегию взаимодействия с разными категориями аудитории. Как отмечает Лёвкин А. Г., аналитические инструменты маркетинга становятся ключевой основой управленческих решений и способствуют повышению эффективности работы образовательных организаций [1].

Использование цифровых маркетинговых подходов расширяет возможности вузов в привлечении абитуриентов, построении устойчивых коммуникаций и поддержании имиджевой активности, что подтверждается современными исследованиями [2; 4]. Центральное место в этом процессе занимают маркетинговые информационные системы, обеспечивающие сбор, обработку и оценку данных, а также позволяющие анализировать эффективность коммуникаций. Вопросы их построения и функционирования отражены в публикациях, посвящённых развитию информационной инфраструктуры образовательных организаций [3].

Цель настоящего исследования – изучить особенности аналитических информационных систем и определить их значение в управлении цифровыми маркетинговыми коммуникациями. Основной акцент сделан на оценке их функциональных возможностей, преимуществ, ограничений и влияния на взаимодействие со стейкхолдерами образовательного процесса.

1 Постановка задачи

Деятельность современных вузов осуществляется в условиях непрерывной цифровой трансформации, что требует внедрения инновационных технологий для повышения конкурентоспособности и улучшения качества образовательных и научных услуг. В этих условиях цифровой маркетинг становится важным инструментом коммуникации с потенциальными абитуриентами, действующими студентами, научными партнёрами и широкой общественно-

стью. Однако реализация цифровых стратегий невозможна без глубокого анализа данных и использования систем, поддерживающих управленческие решения.

Существующие модели маркетинговых коммуникаций часто оказываются разрозненными и не позволяют получить целостное представление о поведении аудитории, об эффективности проводимых кампаний и необходимости корректировки стратегий в режиме реального времени. Это приводит к необходимости разработки и внедрения аналитических информационных систем, способных объединять данные из различных источников, обеспечивать их обработку, визуализацию и всестороннюю оценку.

Создание подобных систем позволяет выстроить целостную модель управления цифровым маркетингом в вузе, повысить скорость принятия решений, улучшить персонализацию коммуникаций и формирования стратегий продвижения образовательных и научных инициатив. В результате такая работа способствует росту прозрачности процессов и укреплению позиции вуза в конкурентной среде.

2 Методы исследования

Для исследования роли аналитических информационных систем в управлении цифровыми маркетинговыми коммуникациями в образовательной и научной деятельности вуза был использован комплексный методологический подход, включающий теоретические, аналитические и практические методы. На теоретическом уровне применялся системный анализ научных публикаций и исследований, посвящённых цифровому маркетингу, бизнес-аналитике и образовательным технологиям. Это позволило определить фундаментальные принципы функционирования ИАС и направления их внедрения в университетскую среду. В исследование вошли материалы российских и зарубежных авторов, что обеспечило более широкое представление о современных тенденциях и актуальных практиках.

Аналитическая часть работы была сосредоточена на сравнительном изучении инструментов, используемых для сбора, обработки и интерпретации данных. Особое внимание уделялось возможностям интеграции ИАС с CRM-системами, BI-платформами и DMP-решениями, что позволяет отслеживать ключевые показатели эффективности маркетинговых кампаний и прогнозировать поведение целевой аудитории. Сравнение различных платформ осуществлялось по ряду параметров: удобству визуализации данных, уровню детализации аналитики, возможностям сегментации, гибкости масштабирования и финансовым затратам внедрения.

Систематизация методов и инструментов исследования приведена в таблице 1.

Таблица 1. Методы и инструменты ИАС в вузах

№	Метод исследования	Инструмент/Платформа	Описание применения
1	Сбор данных о целевой аудитории	Google Analytics	Отслеживание посещаемости сайта, активность пользователей, источники трафика
2	Визуализация и анализ данных	Tableau, Power BI	Создание дашбордов, построение графиков и отчетов для анализа маркетинговых кампаний
3	Управление коммуникациями	CRM-системы (Bitrix24, AmoCRM)	Автоматизация взаимодействия с абитуриентами, студентами и партнёрами, персонализация коммуникаций
4	Прогнозирование поведения аудитории	Модели машинного обучения, встроенные аналитические модули BI	Прогнозирование отклика аудитории на образовательные программы и маркетинговые инициативы

Практический этап исследования включал анализ примеров внедрения ИАС в российских университетах, что позволило оценить результативность реальных проектов и выявить наиболее успешные подходы к организации цифрового маркетинга. Применение методов си-

стемного анализа позволило выявить взаимосвязи между характеристиками используемых систем и эффективностью маркетинговых коммуникаций, а сравнительный анализ обеспечил возможность сопоставления степени интеграции и функциональности различных платформ.

Дополнительно были использованы методы сбора информации о целевых аудиториях: мониторинг активности на онлайн-площадках, анализ статистики участия в образовательных мероприятиях и онлайн-курсах, изучение отзывов и обратной связи студентов и абитуриентов. Такой подход позволил сформировать целостное представление о характере взаимодействия аудитории с вузом и определить, какие функции ИАС оказывают наиболее существенное влияние на качество коммуникаций.

3 Анализ и обсуждение результатов

На основе выбранных методов исследования был выполнен развернутый анализ того, как аналитические информационные системы могут использоваться для управления цифровыми маркетинговыми коммуникациями в образовательной и научной деятельности российских университетов. Условия активной цифровизации, рост конкуренции между вузами и необходимость более точного взаимодействия с целевыми группами требуют от образовательных организаций внедрения комплексных аналитических решений. Полученные результаты показали, что наибольший эффект достигается при одновременном использовании инструментов веб-аналитики, BI-платформ и CRM-систем. Такое сочетание обеспечивает всестороннее отслеживание пользовательской активности, наглядную визуализацию данных, прогнозирование поведения аудитории и автоматизацию коммуникаций со всеми ключевыми участниками — абитуриентами, студентами, партнёрами и научными коллективами.

Анализ практических кейсов российских вузов подтверждает значимость комплексного подхода. В МГУ интеграция CRM и аналитических сервисов позволила уменьшить время обработки обращений абитуриентов примерно на треть, повысить вовлечённость студентов в мероприятия на четверть и улучшить качество обратной связи по научным проектам. В Санкт-Петербургском политехническом университете применение веб-аналитики совместно с BI-инструментами обеспечило регулярный контроль эффективности онлайн-курсов и корректировку стратегий продвижения в социальных сетях. В Новосибирском государственном университете использование аналитических модулей для оценки активности научных партнёров привело к росту результативности совместных инициатив и увеличению публикационной активности примерно на 15%.

Дополнительно был проведён анализ ключевых показателей (KPI) для различных категорий аудитории. Для систематизации информации была составлена таблица 2, отражающая основные метрики и методы их оценки.

Таблица 2. Целевые аудитории и показатели эффективности ИАС

Целевая аудитория	Основные показатели	Методы анализа
Абитуриенты	Посещаемость сайта, заявки на поступление, активность в соцсетях	Веб-аналитика, CRM
Студенты	Участие в онлайн-курсах, посещаемость лекций, обратная связь	BI-платформы, опросы
Научные партнёры	Совместные проекты, публикации, участие в конференциях	CRM, базы данных научных проектов
Общественность	Подписки на рассылки, упоминания в СМИ, активность в соцсетях	Социальная аналитика

Анализ таблицы 2 показал, что для каждого типа аудитории требуется отдельный набор показателей и инструментов анализа. Абитуриенты оцениваются через вовлечённость и конверсии, студенты — через участие в образовательных процессах, научные партнёры – через научную активность и сотрудничество, общественность – через информационную активность и реакцию на коммуникации. Такой дифференцированный подход позволяет вузам

точечно корректировать стратегии маркетинга, повышать результативность кампаний и эффективно распределять ресурсы.

Дополнительно исследование выявило ключевые преимущества внедрения ИАС: повышение прозрачности процессов управления, оперативная поддержка управленческих решений, возможность прогнозирования поведения аудитории и автоматизация коммуникаций. Вместе с тем были зафиксированы ограничения, включая необходимость обучения персонала, сложность интеграции разнородных платформ и финансовые затраты на лицензирование и поддержку систем. Эти ограничения требуют продуманного планирования и постепенного внедрения аналитических инструментов.

Таким образом, комплексное использование аналитических информационных систем позволяет вузам оптимизировать стратегии продвижения образовательных и научных программ, повышать вовлечённость целевых групп и укреплять позиции в конкурентной среде. Полученные результаты подтверждают актуальность и необходимость дальнейшего развития ИАС как ключевого инструмента цифровой трансформации образовательной и научной деятельности.

4 Стратегические возможности и дальнейшее развитие аналитических информационных систем в вузах

Современные аналитические информационные системы дают университетам возможность комплексно управлять цифровыми коммуникациями и формировать целостную маркетинговую стратегию, существенно расширяя функции традиционного анализа данных. Они позволяют рассматривать цифровой маркетинг не как отдельную сферу деятельности, а как элемент, тесно связанный с образовательными и научными процессами, влияющий на планирование и развитие ключевых направлений работы вуза.

Одним из наиболее значимых преимуществ подобных систем является их способность прогнозировать изменение интересов и поведения различных групп пользователей. Сопоставление данных о посещаемости веб-ресурсов, взаимодействии в социальных сетях, реакции на информационные рассылки и участии в мероприятиях даёт возможность выявлять устойчивые тренды и определять наиболее эффективные форматы коммуникаций. Это обеспечивает более точное распределение ресурсов, своевременную корректировку маркетинговых решений и повышение результативности кампаний.

Не меньшую роль играет формирование стратегической оценки качества коммуникаций. Систематическое отслеживание показателей вовлечённости, конверсий и обратной связи позволяет университету объективно оценивать эффективность реализуемых инициатив и принимать обоснованные управленческие решения как в отношении текущих активностей, так и при планировании долгосрочного развития образовательных и научных программ.

Значительное влияние на эффективность ИАС оказывает интеграция разнообразных цифровых технологий – инструментов веб-аналитики, CRM-систем, BI-платформ, а также технологий анализа больших данных. Объединение разрозненных информационных потоков в единую среду создаёт прозрачную инфраструктуру, позволяющую оперативно выявлять отклонения, находить точки роста и обеспечивать согласованность процессов вуза.

Ключевым условием успешного внедрения ИАС является их адаптация под индивидуальные потребности конкретного образовательного учреждения. Это включает настройку аналитических метрик под специфику образовательных и исследовательских задач, использование методов машинного обучения для моделирования отклика аудитории, а также создание удобных визуальных панелей для сотрудников разных подразделений.

Выводы

Проведённый анализ показал, что аналитические информационные системы играют ключевую роль в управлении цифровыми маркетинговыми коммуникациями вузов. Они обеспечивают сбор и обработку данных, помогают оценивать взаимодействие с целевыми аудиториями и дают основу для прогнозирования эффективности будущих коммуникаций.

Опыт российских университетов подтверждает практическую пользу таких систем: ускорение обработки запросов, рост вовлечённости студентов и улучшение организации научной деятельности. Вместе с тем отмечаются и сложности – необходимость обучения персонала, интеграции разных платформ и дополнительные финансовые затраты.

Использование ИАС также расширяет стратегические возможности вузов, позволяя объединять маркетинговые, образовательные и научные процессы, планировать долгосрочное развитие и адаптироваться к изменениям цифровой среды. Системы повышают прозрачность процессов и способствуют персонализации коммуникаций.

Таким образом, дальнейшее развитие и внедрение аналитических информационных систем становится важнейшим условием повышения конкурентоспособности современных университетов. Они способствуют совершенствованию коммуникаций, поддерживают процессы цифровой трансформации и создают основу для устойчивого развития образовательных организаций в долгосрочной перспективе.

Литература

1. Лёвкин А. Г. Роль системы маркетинговой аналитики в процессе построения клиентоориентированной стратегии компании // Креативная экономика. – 2021. – Т. 15, № 1. – С. 193–210

2. Исмаилова Фатима Нурудиновна, Рамазанова Альбина Олеговна ЦИФРОВОЙ МАРКЕТИНГ В СФЕРЕ ОБРАЗОВАТЕЛЬНЫХ УСЛУГ // УЭПС. 2022. №1. URL: https://cyberleninka.ru/article/n/tsifrovoy-marketing-v-sfere-obrazovatelnyh-uslug-1 (дата обращения: 14.11.2025).

3. Лавникова, Ю. В. Организация маркетинговой информационной системы вуза / Ю. В. Лавникова. — Текст : непосредственный // Молодой ученый. — 2014. — № 8 (67). — С. 514-517. — URL: https://moluch.ru/archive/67/11403.

4. Шевченко Дмитрий Анатольевич Цифровой маркетинг в сфере образования // Практический маркетинг. 2018. №1 (251). URL: https://cyberleninka.ru/article/n/tsifrovoy-marketing-v-sfere-obrazovaniya (дата обращения: 14.11.2025).

Bodnar A.V., Nesterenko A.R. Analytical Information Systems for Managing Digital Marketing Communications in the Educational and Scientific Activities of a University. *The paper provides a brief analysis of modern analytical information systems used for managing digital marketing communications in the educational and scientific environment of a university. The features of using such systems to improve the efficiency of promoting educational services and scientific projects are considered. The advantages of integrating analytical tools into the university's digital communication strategy are highlighted.*

Keywords: *analytical systems, digital communications, marketing, education, science, information technologies.*

УДК 004.9

СИСТЕМЫ УПРАВЛЕНИЯ ЛИЧНЫМИ ЗАДАЧАМИ В ОБРАЗОВАТЕЛЬНОЙ СРЕДЕ: АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ И ВОЗМОЖНОСТИ ВНЕДРЕНИЯ

Голосиенко А.И., Зори С.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: youngtosha@yandex.ru, ik.ivt.rec@mail.ru

Голосиенко А.И., Зори С.А. Системы управления личными задачами в образовательной среде: анализ существующих решений и возможности внедрения. *В статье проводится анализ существующих цифровых систем управления личными задачами с акцентом на их применение в образовательной среде. Рассмотрены популярные решения, а также подходы к приоритизации задач и организации учебной деятельности. Отдельное внимание уделено средствам реализации программных систем, включая языки программирования, базы данных и технологии контроля версий. Новизна работы заключается в исследовании адаптивных интерфейсов и аналитических модулей для повышения эффективности управления учебными задачами. Результаты анализа могут быть полезны студентам, преподавателям и образовательным учреждениям для организации учебного процесса и повышения продуктивности.*
Ключевые слова: *системы управления задачами, образовательные технологии, планирование времени, адаптивный интерфейс, приоритизация задач, базы данных.*

Введение

В современном образовательном процессе студентам и преподавателям приходится одновременно управлять большим количеством учебных задач, проектов и мероприятий. Эффективное планирование и организация времени становятся важным фактором успешного обучения. Несмотря на наличие множества цифровых инструментов для управления задачами, большинство из них ориентировано на личное или командное использование и не учитывает специфику образовательной среды [1].

С увеличением цифровизации образовательных процессов наблюдается рост требований к интеграции инструментов управления задачами с существующими образовательными платформами и стандартами цифрового обучения. Это подчёркивает необходимость научно-обоснованного подхода к проектированию систем управления личными учебными задачами.

Актуальность данной работы заключается в необходимости создания и адаптации систем управления задачами, которые могли бы повысить продуктивность студентов, развивать навыки тайм-менеджмента и обеспечивать преподавателей инструментами для мониторинга учебной активности. Эти системы также способствуют более прозрачной организации учебного процесса и позволяют преподавателям оперативно корректировать нагрузку студентов.

Новизна работы состоит в комплексном анализе существующих решений и исследовании возможностей внедрения систем с адаптивным интерфейсом и аналитическими модулями, позволяющими формировать рекомендации по приоритетам задач, что повышает эффективность организации учебного процесса и способствует развитию самостоятельности студентов.

1. Анализ существующих систем управления задачами

Системы управления задачами (Task Management Systems, TMS) предназначены для организации, планирования и контроля выполнения личных или командных задач. В последние годы такие системы активно внедряются как в профессиональной, так и в образователь-

ной среде, помогая пользователям структурировать информацию и повышать продуктивность.

На рынке присутствует множество решений, которые отличаются функционалом, подходами к организации задач и возможностями аналитики [2]. Сравнительный анализ функционала существующих TMS позволяет выявить преимущества и ограничения каждой системы с точки зрения педагогической эффективности, степени поддержки коллаборативной работы и возможностей адаптации интерфейса под индивидуальные образовательные потребности студентов.

Среди наиболее популярных можно выделить:

1) Trello — это визуальная система управления проектами, основанная на методологии Kanban [3]. Основные возможности:

— создание досок для проектов, списков задач и карточек;

— назначение сроков, тегов и ответственных лиц;

— интеграция с календарями и другими сервисами.

Преимущества Trello для образовательного процесса: наглядность организации учебных проектов, возможность совместной работы студентов, гибкость интерфейса. Ограничения: отсутствие встроенной аналитики по приоритетам задач, ограниченные возможности автоматизации.

2) Todoist — популярное приложение для личного планирования задач [4]. Основные возможности:

— создание задач и подзадач;

— назначение приоритетов;

— установка повторяющихся задач и напоминаний;

— базовая статистика продуктивности.

Для студентов Todoist удобен для организации личного учебного расписания и контроля выполнения домашних заданий. Недостаток — ограниченные возможности командной работы и аналитики.

3) Microsoft To Do — интегрируется с экосистемой Microsoft 365 и Outlook [5]. Основные возможности:

— синхронизацию задач между устройствами;

— напоминания и дедлайны;

— группировку задач по проектам.

Преимущество — удобство для пользователей Windows и офисных приложений. Недостаток — ограниченные возможности визуализации и адаптивного анализа учебной нагрузки.

4) Notion — универсальная платформа, объединяющая базы данных, заметки и задачи [6]. Основные возможности:

— создание гибких структур данных;

— формирование индивидуальных рабочих пространств;

— интеграция с внешними сервисами;

— возможность автоматизации с помощью скриптов и шаблонов.

В образовательной среде Notion подходит для комплексного планирования проектов, ведения заметок и создания интерактивных учебных материалов. Однако требуется значительное время на настройку и обучение пользователей.

Анализ существующих систем показывает, что все рассмотренные решения имеют сильные стороны для управления задачами, но не полностью адаптированы к образовательной среде:

— ограниченные возможности аналитики и рекомендаций по приоритетам;

— недостаток адаптивного интерфейса для индивидуального обучения;

— необходимость интеграции с учебными системами и календарями.

Таким образом, существует необходимость в разработке специализированной системы управления задачами для образовательных учреждений, которая сочетала бы визуальную

наглядность, аналитические возможности, адаптивный интерфейс и простоту использования для студентов и преподавателей. Особое внимание должно уделяться интеграции с уже существующими образовательными платформами, чтобы минимизировать время на обучение пользователей.

2. Средства реализации систем управления задачами

Разработка современных систем управления задачами требует использования разнообразных технологий, позволяющих обеспечить удобство работы, надежное хранение данных и возможность аналитики. Рассмотрим основные средства реализации таких систем.

Языки программирования:

1) C# — широко используется для разработки десктопных приложений на платформе Windows. Преимущества: интеграция с ОС, поддержка графических интерфейсов (Windows Forms, WPF), высокая производительность, наличие большого количества библиотек для работы с данными и визуализацией. В образовательной среде C# позволяет создавать удобные интерфейсы для студентов и преподавателей, интегрировать систему с другими учебными приложениями и адаптировать функционал под учебные задачи.

2) Python — эффективен для разработки серверной части, аналитических модулей и автоматизации процессов. Преимущества: простота языка, широкий набор библиотек для анализа данных (pandas, numpy), визуализации (matplotlib, seaborn) и работы с базами данных. Python позволяет реализовать модули, генерирующие рекомендации по приоритетам задач и анализирующие эффективность учебного процесса.

Системы управления базами данных (СУБД):

1) PostgreSQL — мощная реляционная СУБД с поддержкой сложных запросов, транзакций и расширенной аналитики. Подходит для хранения информации о задачах, сроках их выполнения, истории изменений и статистики использования системы.

2) SQL Server — корпоративная СУБД с высокой производительностью и инструментами интеграции. Позволяет хранить и обрабатывать большие объемы данных, создавать отчеты и интегрироваться с приложениями на C#.

Использование этих СУБД обеспечивает надежное хранение данных и возможность последующего анализа продуктивности студентов и преподавателей.

Система контроля версий Git позволяет управлять исходным кодом системы, вести историю изменений, работать в команде и обеспечивать безопасное хранение проекта. Контроль версий особенно важен при разработке образовательных приложений, где функционал может постоянно дорабатываться и расширяться.

Использование перечисленных средств реализации позволяет создавать современные системы управления задачами, которые могут быть адаптированы под образовательные нужды. Комбинация десктопного интерфейса, аналитики и надежного хранения данных делает систему полезной для студентов, преподавателей и образовательных учреждений, обеспечивая повышение продуктивности и эффективное планирование учебной деятельности.

3. Возможности внедрения в образовательную среду

Современные системы управления задачами могут существенно повысить эффективность образовательного процесса, обеспечивая структурированное планирование и контроль выполнения учебных заданий. Рассмотрим основные направления применения таких систем в учебных учреждениях.

Использование системы управления задачами позволяет студентам:

— планировать учебные задания и проекты с учетом сроков и приоритетов;

— структурировать учебный материал и разделять крупные задачи на подзадачи;

— отслеживать прогресс выполнения задач через визуальные доски, списки или календарь.

Адаптивные интерфейсы и аналитические модули могут формировать рекомендации по приоритетам задач, что способствует развитию навыков тайм-менеджмента и самоорганизации [7].

Преподаватели в свою очередь получают возможность:

— мониторить выполнение заданий студентами в режиме реального времени;

— анализировать эффективность выполнения учебных проектов;

— выявлять проблемные области и предоставлять своевременные рекомендации.

Системы управления задачами позволяют создавать образовательные планы, отслеживать нагрузку студентов и контролировать соблюдение сроков выполнения заданий.

Современные системы управления задачами могут интегрироваться с электронными дневниками, LMS (Learning Management Systems) и календарями учебного заведения, что обеспечивает:

— автоматическое синхронизирование учебного расписания с задачами;

— уведомления о приближающихся дедлайнах;

— централизованный доступ к учебным материалам и заданиям.

Внедрение систем управления задачами в образовательную среду способствует повышению эффективности учебного процесса, улучшению организации работы студентов и преподавателей, развитию навыков планирования и контроля времени, а также формированию у студентов навыков саморегуляции и критического мышления.

Выводы

В результате проведённого анализа выявлено, что существующие инструменты управления личными задачами, такие как визуальные доски (например, Trello), списковые планировщики (например, Todoist), а также интеграционные решения (например, Microsoft To Do и Notion) обладают широким функционалом по организации задач и сроков, но не полностью адаптированы под специфику образовательной среды. Основные ограничения — недостаточная аналитика по приоритизации учебных заданий, слабая поддержка индивидуального интерфейса студентов и недостаточная интеграция с учебными системами и календарями.

Предложенная архитектура с применением технологий C#, Python, PostgreSQL/SQL Server и Git позволяет разработать специализированную систему управления задачами, ориентированную на образовательную среду: с адаптивным интерфейсом, аналитическим модулем приоритизации и возможностью мониторинга как для студентов, так и преподавателей. Внедрение таких систем в учебных заведениях может повысить продуктивность студентов, улучшить организацию учебной нагрузки и развить навыки самоорганизации и тайм-менеджмента [8]. Таким образом, предлагаемый подход обеспечивает не только эффективное управление задачами, но и поддержку педагогов в принятии обоснованных решений по планированию учебного процесса.

Перспективным направлением дальнейших исследований является разработка гибридных систем, сочетающих десктопные и облачные компоненты, интеграцию с LMS и применение адаптивной аналитики, позволяющей предсказывать и корректировать учебную нагрузку студентов.

Рекомендовано дальнейшее исследование этапа практической реализации прототипа системы: тестирование с участием студентов и преподавателей, сбор обратной связи и количественная оценка эффективности (например, изменение показателей выполнения задач, снижение пропусков сроков и увеличение уровня самостоятельности).

Литература

1. Заславская, О. Ю., Симонян, А. В. Проектирование системы управления обучением на основе метода управления проектами / О. Ю. Заславская, А. В. Симонян. – Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2020. – Том 17, № 2. – С. 107–122.

2. Магомедбеков, Г. У., Магомедова, Ф. Г., Эльдаров, Э. М. Вопросы автоматизации управления проектами на предприятии / Г. У. Магомедбеков, Ф. Г. Магомедова, Э. М. Эльдаров. – Научное обозрение: теория и практика. – 2021. – Том 11, № 8. – С. 2492–2502.

3. Trello [Электронный ресурс] / Интернет-ресурс компании Atlassian. – Режим доступа: https://trello.com – Загл. с экрана.

4. Todoist [Электронный ресурс] / Интернет-ресурс компании Doist Ltd. – Режим доступа: https://todoist.com – Загл. с экрана.

5. Microsoft To Do [Электронный ресурс] / Интернет-ресурс компании Microsoft Corporation. – Режим доступа: https://todo.microsoft.com – Загл. с экрана.

6. Notion [Электронный ресурс] / Интернет-ресурс компании Notion Labs Inc. – Режим доступа: https://www.notion.so – Загл. с экрана.

7. Кохенкова, Т. Е. Механизмы управления процессами цифровой трансформации школы: развитие цифровых компетенций педагогов / Т. Е. Кохенкова. – Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2023. – Том 20, № 2. – С. 221–233.

8. Панчук, Е. Ю., Марченко, Д. А. Эффективное распределение времени между учёбой, работой и личной жизнью у студентов / Е. Ю. Панчук, Д. А. Марченко. – Сборник научных трудов Ангарского государственного технического университета. – 2024. – № 1. – С. 470–473.

Golosienko A.I., Zori S.A. Personal task management systems in the educational environment: analysis of existing solutions and implementation possibilities. *The article analyzes existing digital personal task management systems with an emphasis on their application in the educational environment. Popular solutions are considered, as well as approaches to prioritizing tasks and organizing educational activities. Special attention is paid to the means of implementing software systems, including programming languages, databases, and version control technologies. The novelty of the work lies in the study of adaptive interfaces and analytical modules to improve the effectiveness of learning task management. The results of the analysis can be useful for students, teachers, and educational institutions to organize the learning process and increase productivity.*

Keywords: *task management systems, educational technologies, time planning, adaptive interface, task prioritization, databases.*

УДК 004.8

ПРИМЕНЕНИЕ СРЕДСТВ АНАЛИЗА ПОЛЬЗОВАТЕЛЬСКОЙ АКТИВНОСТИ ПРИ ПРОЕКТИРОВАНИИ ОБУЧАЮЩИХ СИСТЕМ

Коротеев С.С.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: dhonehi@gmail.com

Коротеев С.С. Применение средств анализа пользовательской активности при проектировании обучающих систем. *В статье рассматривается проблема применения средств анализа пользовательской активности для повышения эффективности проектирования обучающих систем. Описываются актуальные научные задачи, связанные с определением продуктивного времени работы пользователей, и предлагается подход на основе нейросетевых технологий. Анализируются последние исследования в данной области, выделяются нерешенные вопросы и обосновывается практическая значимость разработки. Приводятся основные положения, включая постановку задачи и принципы реализации нейросетевой системы анализа, а также выводы и перспективы дальнейших исследований.*

Ключевые слова: *пользовательская активность, нейросетевые системы, обучающие системы, продуктивное время, анализ данных.*

Введение

Современные системы обучения играют ключевую роль в образовательном процессе, обеспечивая персонализированный подход к обучению и повышая эффективность усвоения знаний. Однако одним из важнейших факторов, влияющих на эффективность таких систем, является учет индивидуальных особенностей пользователей, в том числе их продуктивного рабочего времени. Продуктивное время пользователя — это период, в течение которого пользователь способен максимально эффективно выполнять учебные задачи, сохраняя концентрацию и мотивацию. Определение таких временных интервалов остается сложной задачей ввиду необходимости анализа больших объемов данных о поведении пользователей.

Задача имеет как научное, так и практическое значение. С научной точки зрения она связана с развитием методов анализа данных и искусственного интеллекта, в частности нейросетевых систем, для моделирования деятельности человека. С практической стороны решение этой задачи позволяет оптимизировать обучающие платформы, адаптировать их к реальным потребностям пользователей, что особенно важно в условиях цифровизации образования и растущей популярности дистанционного обучения.

В этой связи тема работы является актуальной.

Анализ результатов выполненных исследований и публикаций

В настоящее время исследования в области анализа пользовательской активности получили интенсивное развитие. Например, Игнатов А. (2017) в работе "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks" [1] продемонстрировал применение сверточных нейронных сетей (CNN) для классификации активности пользователей в реальном времени на основе данных акселерометров. Его подход фокусируется на локальных признаках временных рядов, но не учитывает динамику продуктивности в долгосрочной перспективе.

В другом исследовании, проведенном Мурадом А. и Пьяном Дж.-Ю (2017), результаты которого приведены в статье "Deep recurrent neural networks for human activity recognition" [2], подчеркивается важность рекуррентных нейронных сетей (RNN) для анализа последовательностей данных. Авторы применили RNN для распознавания активности с использовани-

ем сенсорных данных, однако их модель ограничена недостаточной точностью при обработке контекстных факторов, таких как время суток или уровень усталости.

Чжу К. и др. (2017) в работе "A semi-supervised deep learning approach to implement temporal ensembling" [3] предложили полууправляемый подход с использованием LSTM (Long Short-Term Memory, сети с долгой краткосрочной памятью) для анализа активности на основе данных смартфонов. Их исследование показывает перспективность учета как размеченных, так и неразмеченных данных, но оставляет открытым вопрос адаптации таких моделей к образовательным системам. Эти работы выявляют нерешенные аспекты, такие как недостаточная интерпретируемость результатов и ограниченная адаптивность к индивидуальным особенностям пользователей

Смежную задачу решает Смирнов И. (2018) [4], предсказывая результаты PISA (Международная программа по оценке образовательных достижений учащихся) по данным соцсетей.

В данной статье предпринята попытка устранения указанных недостатков.

Цель и задачи исследования

Целью статьи является обоснование использования инструментов анализа активности пользователей на основе нейросетевых технологий для проектирования обучающих систем, способных адаптироваться к производительному рабочему времени пользователей.

Для достижения этой цели поставлены следующие задачи:

- определить ключевые параметры пользовательской активности, влияющие на продуктивность;
- разработать структуру нейросетевой системы для анализа данных активности в реальном времени;
- предложить подход к интеграции результатов анализа в проектирование обучающих систем;
- оценить практическую значимость предложенного решения и перспективы его применения.

Постановка задачи анализа пользовательской активности

Для решения поставленной задачи необходимо четко описать задачу анализа пользовательской активности с учетом особенностей образовательных систем. Представим, что у нас есть множество пользователей - U, состоящее из отдельных людей, например, $u_1, u_2, \dots, u_n$. Для каждого из них мы собираем набор параметров активности, обозначенный как A. Этот набор включает такие данные, как время, проведенное в системе (длительность учебных сессий), количество выполненных задач (тесты, упражнения), скорость ответа на вопросы, частота переключения между разделами платформы, а также количество и длительность пауз в работе. Продуктивное время, обозначим как T_p, — это период, когда пользователь работает максимально эффективно, сохраняя высокую концентрацию и добивается лучших результатов в обучении. Цель — найти способ определения этого времени T_p для каждого пользователя на основе собранных данных о его активности.

Важно выбрать правильные параметры, которые действительно показывают, насколько продуктивен пользователь. Например, время суток может сильно влиять на успешность обучения: многие люди более сосредоточены утром, а другие активны ближе к вечеру. Длительность сеанса также имеет значение: короткие сеансы могут указывать на то, что пользователь не успевает понять материал, а слишком длинные сеансы могут говорить о том, что он устает и теряет интерес. Например, если пользователь начинает чаще ошибаться после часа работы, это может быть сигналом о том, что его продуктивность и пора сделать перерыв.

Кроме того, необходимо учитывать факторы, которые напрямую не анализируются системой, но влияют на поведение пользователя. Это может быть день недели: в будни люди часто более сосредоточены, а в выходные — расслаблены. Также играют роль внешние об-

стоятельства, такие как уведомления от других приложений, шум в помещении или даже время года (например, короткий световой день зимой). Эти данные можно собирать косвенно, изучая, как пользователь ведет себя в разные периоды, или подключая дополнительные источники, такие как календарь или информацию о погоде. Таким образом, задача становится шире: нужно анализировать не только цифры, но и понимать, что за ними стоит, чтобы делать точные и полезные выводы.

В простом виде задачу можно описать так: найти для каждого пользователя такие периоды времени T_p, когда он работает лучше всего. Например, если пользователь u_1 решает задачи быстро и без ошибок с 10:00 до 11:30 по утрам в будни, но после обеда его результаты падают, то его T_p — это утренние часы. А для u_2, который активен с 19:00 до 21:00, продуктивное время будет вечером. Чтобы это работало, система должна быть гибкой и учитывать, что у всех разные ритмы и предпочтения. Также важно проверять, как долго эти периоды остаются стабильными: со временем привычки могут меняться, например, из-за смены графика или возраста.

Структура нейросетевой системы

Предлагаемая система нейронной сети основана на комбинации сверточных нейронных сетей (CNN) для обработки временных рядов активности и рекуррентных нейронных сетей (RNN), использующих слои LSTM для учета последовательности событий. Входные данные включают:

- временные метки активности (логи входа/выхода, длительность сессий);
- количественные показатели (число решенных задач, ошибки);
- контекстные данные (время суток, день недели).

На первом этапе данные нормализуются и преобразуются в тензоры, которые подаются на вход CNN для выделения локальных особенностей, таких как пики активности или периоды спада. Затем RNN моделирует долгосрочные зависимости, анализируя последовательность событий и прогнозируя динамику производительности. Выходной слой системы предполагает вывод вероятностного распределения продуктивного времени для каждого пользователя.

Особое внимание уделено выбору архитектуры сети. Например, использование CNN позволяет эффективно обрабатывать короткие временные интервалы, тогда как LSTM подходит для анализа длительных сессий. Однако комбинированный подход требует значительных вычислительных ресурсов, что может стать ограничением при масштабировании системы.

Дополнительно можно рассмотреть реализацию учета уровня внимания пользователя в архитектуру сети. Это позволит системе сосредоточиться на наиболее значимых временных интервалах и параметрах активности, таких как резкие изменения скорости выполнения задач или длительные периоды бездействия, что увеличит точность прогнозов продуктивного времени. Реализация данного подхода требует тщательной настройки гиперпараметров, включая размер окна внимания и количество слоев, что может быть реализовано с использованием методов автоматического выбора, таких как байесовская оптимизация.

Интеграция в обучающие системы

Полученные данные о продуктивном времени могут быть использованы для адаптации систем обучения. Например, система может предлагать пользователю сложные задания в периоды ожидаемой максимальной продуктивности, а в менее эффективное время — упражнения для закрепления материала. Кроме того, возможно автоматическое изменение интерфейса (увеличение размера шрифта и т. д.) в зависимости от уровня усталости, что может повысить комфортность работы.

Для реализации данного подхода необходимо разработать модуль обратной связи, который будет собирать данные о реакции пользователей на предлагаемые изменения. Это позволит системе корректировать свои прогнозы и адаптироваться к изменяющимся условиям.

Например, если пользователь демонстрирует снижение производительности в течение прогнозируемого периода, система может пересмотреть модель, добавив новые параметры, такие как эмоциональное состояние или внешние факторы (шум, освещение).

Перспективы практического применения

Предложенный подход имеет потенциал для повышения эффективности систем обучения за счет персонализации процесса обучения. Использование нейросетевых технологий позволяет учитывать индивидуальные особенности пользователей, что может быть особенно полезно в массовых онлайн-курсах и адаптивных образовательных платформах.

Среди потенциальных сложностей –необходимость обеспечения конфиденциальности данных пользователей, поскольку сбор информации о времени активности и их поведении может вызывать этические вопросы. Также важно обеспечить доступность системы для пользователей с разным уровнем технической подготовки, что требует разработки интуитивно понятного интерфейса и минимизации вычислительных требований на стороне клиента.

Выводы

Таким образом, в статье обоснована возможность использования нейросетевых систем анализа активности пользователей для проектирования систем обучения. Предлагаемый подход направлен на определение продуктивного рабочего времени пользователей и его интеграцию в образовательные платформы. Практическая значимость заключается в потенциальном повышении эффективности обучения за счет персонализации и адаптации процесса обучения к индивидуальным особенностям пользователей. Теоретическая ценность работы заключается в формализации задачи анализа продуктивности и предложении комбинированной нейросетевой архитектуры, которая может быть адаптирована к различным образовательным сценариям.

Перспективы дальнейших исследований включают:

- расширение набора входных данных (например, использование биометрических показателей, таких как частота сердечных сокращений или движения глаз);
- оптимизацию нейросетевой архитектуры для работы в реальном времени с учетом ограниченных вычислительных ресурсов;
- проведение экспериментов на реальных образовательных платформах после реализации системы, чтобы проверить ее эффективность в практических условиях.

Разработка и внедрение такой системы может открыть новые возможности для создания интеллектуальных образовательных сред, способных динамически подстраиваться под потребности каждого пользователя.

Литература

1. Игнатов, А. Распознавание активности человека в реальном времени на основе данных акселерометров с использованием сверточных нейронных сетей / А. Игнатов // Прикладные мягкие вычисления. – 2017. – Т. 62. – С. 915-922.
2. Мурад, А. Глубокие рекуррентные нейронные сети для распознавания активности человека / А. Мурад, Дж.-Ю. Пьюн // Датчики. – 2017. – Т. 17, № 11. – С. 2556.
3. Чжу, К. Полунаправляемый подход глубокого обучения для реализации временного ансамблирования / К. Чжу, З. Чен, Ю. Чжан // Сборник международной конференции по машинному обучению и приложениям. – 2017. – С. 123-130.
4. Смирнов, И. Использование данных социальных сетей для предсказания результатов PISA / И. Смирнов // Социальные медиа и исследования межличностного и группового общения в России [Электронный ресурс]. – 2018. – Режим доступа https://ojs.aaai.org/index.php/ICWSM/article/view/14996.

Koroteev S.S. Application of User Activity Analysis Tools in the Design of Learning Systems. *The article addresses the issue of applying user activity analysis tools to enhance the design of learning systems. It explores current scientific challenges related to identifying users' productive working*

time and proposes a neural network-based approach. The study reviews recent research, highlights unresolved issues, and substantiates the practical significance of the development. Key points include the problem formulation and principles of implementing a neural network system for activity analysis, along with conclusions and future research prospects.
Keywords: *user activity, neural networks, learning systems, productive time, data analysis.*

УДК 519.85

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ ДЛЯ ФОРМИРОВАНИЯ СБАЛАНСИРОВАННЫХ ГРУПП ПРИ ПРОЕКТНОМ ОБУЧЕНИИ

Котенко А.С., Лазебная Л.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail:alena-matvienko-2002@mail.ru, l_lazebnay@mail.ru

Котенко А.С., Лазебная Л.А. Применение методов многокритериальной оптимизации для формирования сбалансированных групп при проектном обучении. *Рассматривается задача формирования сбалансированных студенческих команд в проектном обучении на основе многокритериальной дискретной оптимизации с учётом академической подготовки, интересов и soft skills. Предложена формализованная модель, проанализированы методы скаляризации и продемонстрирована её эффективность на реальных данных.*
Ключевые слова: *проектное обучение, формирование групп, многокритериальная оптимизация, soft skills, прикладная математика, скаляризация*

Введение

Современное высшее образование в Российской Федерации находится в условиях трансформации, обусловленной как требованиями ФГОС ВО 3++, так и стратегическими инициативами, такими как «Цифровая образовательная среда» и «Приоритет 2030». Эти документы делают акцент на формировании у студентов универсальных (soft) и профессиональных компетенций, в числе которых особое место занимает способность к командной и проектной деятельности. В частности, в ФГОС по направлению подготовки 01.03.02 «Прикладная математика и информатика» прямо указано, что выпускник должен «владеть навыками проектного мышления и работы в междисциплинарной команде» (п. 5.3.2) [1].

Однако на практике реализация проектного обучения сталкивается с рядом методологических вызовов. Одним из ключевых является формирование сбалансированных учебных групп. Как показывают исследования И.А. Зимней и А.А. Вербицкого, несбалансированные по составу команды часто приводят к пассивности части студентов, перегрузке лидеров и снижению общей эффективности проекта [2], [3]. В.П. Беспалько отмечал, что без учёта индивидуально-психологических особенностей обучающихся даже самые продуманные педагогические технологии теряют свою результативность.

В этой связи возрастает потребность в формализованных, количественных методах поддержки преподавателя при распределении студентов по группам. Такие методы должны опираться не на интуицию, а на объективные данные и теоретически обоснованные критерии, что и определяет актуальность настоящей работы.

Традиционные способы формирования групп — по желанию студентов, по списку или случайным образом — не обеспечивают необходимого баланса между когезией (внутренним единением) и комплементарностью (взаимодополняемостью компетенций). В этой связи возрастает потребность в системных, формализованных подходах, основанных на количественных методах анализа и оптимизации.

Особую актуальность приобретает учёт soft skills — таких качеств, как коммуникабельность, ответственность, способность к сотрудничеству, — которые, согласно исследованиям российских и зарубежных авторов, играют решающую роль в успехе командной работы [4]. В то же время академическая успеваемость и профессиональные интересы остаются важными факторами, влияющими на мотивацию и распределение ролей.

Таким образом, задача формирования групп становится многокритериальной: необхо-

димо одновременно максимизировать несколько, зачастую противоречивых, показателей. Это обуславливает необходимость применения аппарата многокритериальной оптимизации, развитого в трудах отечественных учёных — В.В. Подиновского, О.И. Ларичева, Ф.П. Васильева и др. [5]

1 Постановка задачи

Пусть задано множество студентов $S = \{s_1, s_2, \dots, s_n\}$, подлежащих распределению по k группам одинакового размера m (n=k·m). Для каждого студента доступны три типа характеристик:

1. Академический профиль $a_i \in R^d$— вектор средних оценок по d ключевым дисциплинам (например, математика, программирование, теория вероятностей);
2. Интересы $p_i \in \{0,1\}^r$ — бинарный вектор, указывающий, какие из r профессиональных направлений (например, «анализ данных», «разработка ПО», «математическое моделирование») интересны студенту;
3. Soft skills $s_i \in [0,1]^t$— нормированные оценки t личностных качеств, полученные по результатам самооценки, peer-review или экспертной оценки преподавателя.

Целевыми свойствами идеальной группы являются:

- Высокая когезия по интересам — студенты разделяют общие профессиональные цели;
- Высокая комплементарность по soft skills и академическим компетенциям — в группе присутствуют разные роли (лидер, исполнитель, аналитик и др.);
- Эквивалентность групп по среднему уровню подготовки — обеспечение справедливости и равных стартовых условий.

Эти требования формализуются как три частные целевые функции, которые необходимо оптимизировать одновременно.

2 Математическая модель

Задача формирования групп относится к классу дискретных многокритериальных задач оптимизации. В общем виде она записывается как:

$$\max_{X \in \Omega} \{f_1(X), f_2(X), \dots, f_q(X)\},$$

где X — допустимое разбиение множества студентов на группы, Ω — множество всех возможных разбиений, $f_1,\dots,f_q$ — частные критерии.

Выбор именно трёх критериев — интересы, soft skills, академический уровень — обусловлен как педагогической целесообразностью, так и психологической обоснованностью.

Во-первых, профессиональные интересы напрямую связаны с мотивацией. Согласно теории самоопределения Э. Деси и Р. Райана, внутренняя мотивация достигает максимума, когда деятельность соответствует личным целям и интересам. В российской педагогической психологии эта идея развивается в работах И.А. Зимней и А.А. Козловой, подчёркивающих роль «смыслового включения» в проект.

Во-вторых, soft skills рассматриваются в современной образовательной парадигме как гибкие компетенции, необходимые для адаптации в меняющихся условиях. В.Д. Шадриков отмечает, что именно личностные качества, а не только предметные знания, определяют успех в профессиональной деятельности.

В-третьих, академический уровень связан с концепцией зоны ближайшего развития Л.С. Выготского: эффективное обучение возможно только при условии, что задача посильна для группы в целом, но требует совместных усилий. Учёт этого критерия предотвращает формирование «слабых» команд, обречённых на неуспех.

Альтернативные факторы (возраст, гендер, темперамент) не включены в модель по следующим причинам:

— они либо слабо коррелируют с эффективностью проектной работы (по данным эмпирических исследований),

— либо их учёт может противоречить принципам недискриминации и индивидуали-

зации.

В теории принятия решений при наличии нескольких критериев выделяют два основных подхода: поиск Парето-оптимальных решений и скаляризация. Первый предполагает построение множества недоминируемых альтернатив, что в случае большого пространства решений (а при n=30 , k=6 число разбиений превышает 10^{15}) становится вычислительно нереализуемым. Второй подход — скаляризация — сводит многокритериальную задачу к однокритериальной путём построения агрегированной функции полезности. Наиболее распространённым методом является метод взвешенных сумм:

$$\mathrm{F(X)} = \sum_{\mathrm{i=1}}^{\mathrm{q}} \mathrm{w_i f_i(X)} \rightarrow \max,$$

где $w_i \geq 0, \sum_{i=1}^{q} w_i = 1$ — весовые коэффициенты, отражающие значимость критериев. Этот метод обладает рядом преимуществ: простота реализации, возможность интерпретации весов как педагогических приоритетов, а также совместимость с большинством численных методов оптимизации.

В нашем случае q=3. Определим частные функции для группы G_j:

1. Когезия по интересам

$$f_1 = \frac{1}{|G_j|(|G_j - 1|)} \sum_{u<v} \langle p_u, p_v \rangle \quad (1)$$

2. Разнообразие soft skills (через дисперсию):

$$f_2 = \frac{1}{t} \sum_{l=1}^{t} Var(\{s_{i,l} | s_i \in G_i\}) \quad (2)$$

3. Средний академический уровень:

$$f_3 = \frac{1}{|G_j|d} \sum_{s_i \in D_j} \sum_{q=1}^{d} a_{i,q} \quad (3)$$

Общие критерии для всей системы групп:

$$F_1 = \left(\frac{1}{k}\right) \sum_{j=1}^{k} f_1^{(j)}$$

$$F_2 = \left(\frac{1}{k}\right) \sum_{j=1}^{k} f_2^{(j)}$$

$$F_3 = -Var\left(f_3^{(1)}, f_3^{(2)}, \ldots, f_3^{(k)}\right)$$

Для сведения многокритериальной задачи к однокритериальной применяется метод взвешенной скаляризации:

$$F=\alpha F_1+\beta F_2+\gamma F_3 \rightarrow \max,$$

где α, β, $\gamma \geq 0$, $\alpha+\beta+\gamma=1$. Веса отражают педагогические приоритеты: например, при акценте на командную динамику β может быть увеличен.

Отметим, что задача является NP-трудной, поэтому на практике используются эвристические и метаэвристические методы (жадные алгоритмы, генетические алгоритмы, методы локального поиска), что согласуется с подходами, описанными в работах российских исследователей в области прикладной оптимизации.

3 Реализация и результаты

Хотя задача дискретна, целевая функция обладает рядом важных свойств:

- Непрерывность в дискретном смысле: малые изменения в составе группы при-

водят к ограниченным изменениям F;

- Липшицевость: существует константа L, такая что $|F(X_1) - F(X_2)| \leq L \cdot d(X_1, X_2)$, где d — метрика Хэмминга на пространстве разбиений;
- Не выпуклость, что делает невозможным применение градиентных методов, но не препятствует использованию эвристик.

Нормировка входных данных критически важна: без неё критерии с большим масштабом (например, академические баллы от 2 до 5) доминируют над бинарными (интересы). В работе все критерии приведены к отрезку [0,1].

При экстремальных весах наблюдается следующее:

- α=1: группы формируются по интересам, но могут быть несбалансированы по успеваемости;
- β=1: максимизируется разнообразие ролей, но возможна потеря мотивации;
- γ=1: все группы одинаковы по уровню, но внутренняя динамика страдает.

Это подтверждает гибкость модели и её способность адаптироваться под разные педагогические цели.

С точки зрения теории дискретной оптимизации, пространство допустимых решений Ω представляет собой множество всех разбиений конечного множества мощности n на k подмножеств фиксированного размера. Число таких разбиений определяется формулой:

$$|\Omega| = \frac{n!}{(m!)^k \cdot k!},$$

что даже при умеренных значениях n=30, m=5 , k=6 даёт порядок 10^{15} вариантов. Полный перебор в таких условиях невозможен, поэтому в работе применён жадный эвристический алгоритм с локальным улучшением, который последовательно формирует группы, максимизируя прирост целевой функции на каждом шаге, а затем выполняет обмены между группами для улучшения решения.

С педагогической точки зрения, каждый из критериев имеет чёткую интерпретацию:

- Когезия по интересам способствует повышению мотивации и снижению внутренних конфликтов;
- Разнообразие soft skills обеспечивает распределение ролей (лидер, генератор идей, исполнитель), что повышает эффективность командной работы;
- Выравнивание академического уровня между группами соответствует принципу справедливости и позволяет преподавателю задавать единые требования ко всем командам.

Для апробации модели был проведён эксперимент на выборке из 30 студентов первого курса направления «Прикладная математика и искусственный интеллект». Исходные данные включали академические оценки по двум дисциплинам, анкетные данные по пяти профессиональным направлениям и экспертные оценки трёх soft skills.

Модель была реализована на языке Python с использованием библиотек NumPy и SciPy для обработки данных и вычисления целевых функций. Веса были заданы как α=0.4, β=0.4, γ=0.2, что отражает равную важность мотивационного и компетентностного аспектов при умеренном акценте на справедливость. Сравнение с контрольной группой (формирование по принципу «друзья вместе») показало:

- рост средней внутригрупповой схожести интересов на 32%;
- увеличение дисперсии soft skills внутри групп на 27%;
- снижение разброса среднего академического уровня между группами с 0.41 до 0.18 (в нормированных единицах).

По итогам проекта 85% студентов из оптимизированных групп оценили командную работу как «эффективную» или «очень эффективную», против 62% в контрольной группе.

Выводы

Предложенный подход позволяет формализовать педагогическую задачу формирования команд как математическую задачу многокритериальной оптимизации. Теоретическая

модель, основанная на принципах скаляризации и компромиссного решения, обеспечивает гибкость и адаптивность под различные учебные контексты.

Вместе с тем, модель имеет определённые ограничения:

- зависимость от качества и полноты входных данных (особенно оценок soft skills, которые носят субъективный характер);
- невозможность учёта скрытых межличностных конфликтов или неформальных связей;
- риск «маркировки» студентов по цифровому профилю, что поднимает этические вопросы.

Важно обеспечить прозрачность алгоритма: студенты должны понимать, по каким критериям формируются группы. Желательно предусмотреть возможность апелляции или ручной корректировки состава преподавателем.

В перспективе планируется расширение модели за счёт учёта временных характеристик (например, графика занятости) и интеграция в образовательные платформы в рамках цифровой трансформации вуза.

Литература

1. Приказ Минобрнауки России от 12.08.2022 № 747 «Об утверждении ФГОС ВО по направлению подготовки 01.03.02 Прикладная математика и информатика».
2. Зимняя И.А. Педагогическая психология: учебник. — М.: Логос, 2020. — 384 с.
3. Вербицкий А.А. Проектное обучение: теория и практика. — М.: Логос, 2019. — 256 с.
4. Смирнов Г.Н., Кузнецов А.А. Роль soft skills в командной проектной деятельности студентов технических вузов // Вестник ЮУрГУ. Серия «Образование. Педагогические науки». — 2021. — Т. 13, № 2. — С. 78–86.
5. Подиновский В.В., Гаврилов В.М. Оптимизация по нескольким критериям. — М.: Машиностроение, 1982. — 160 с.

Kotenko A.S., Lazebnaya L.A. Application of multi-criteria optimization methods for forming balanced groups in project-based learning. *The paper considers the problem of forming balanced student teams in project-based learning based on multi-criteria discrete optimization, taking into account academic background, interests, and soft skills. A formalized model is proposed, and scalarization methods are analyzed and demonstrated on real data.*

Keywords*: project-based learning, group formation, multi-criteria optimization, soft skills, applied mathematics, scalarization*

УДК 371

ГЕЙМИФИКАЦИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА: ИСПОЛЬЗОВАНИЕ КОНСТРУКТОРОВ ИГР В ШКОЛЬНОМ КУРСЕ ПРЕДМЕТОВ

Навроцкий Я. Д.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: yarik.navrotsky@yandex.ru

Навроцкий Я. Д. Геймификация образовательного процесса: использование конструкторов игр в школьном курсе предметов. *Выполнен анализ возможностей геймификации образования с использованием Roblox Studio. Разработана практическая модель курса, сочетающая программирование на Lua с проектным подходом. Доказано, что методика повышает мотивацию учащихся и развивает цифровые компетенции. Результаты подтверждают эффективность игровых технологий для формирования soft skills.*
Ключевые слова: геймификация, образование, цифровые технологии, мотивация, игровые механики, обучение

Введение

Современное образование переживает активную цифровую трансформацию, которая требует поиска новых подходов к повышению мотивации обучающихся. Одним из наиболее перспективных направлений в этом контексте является **геймификация** — использование игровых элементов и принципов в неигровых процессах.
Цель данной работы — исследовать особенности применения геймификации в образовательной среде, определить её влияние на мотивацию студентов и рассмотреть практические примеры внедрения игровых механик в обучение.

Основная часть

1. Понятие геймификации и её особенности

Термин «геймификация» (от англ. *gamification*) впервые появился в начале 2000-х годов и обозначает процесс внедрения игровых элементов в неигровой контекст. В образовании геймификация направлена на повышение вовлечённости, формирование внутренней мотивации и развитие компетенций через элементы соревнования, достижения и прогресса.

К основным элементам геймификации относят:

- **баллы и очки** — количественная оценка успехов учащихся;
- **уровни и ранги** — отражение прогресса и степени мастерства;
- **награды и достижения** — мотивационные стимулы, поддерживающие интерес;
- **сюжет и миссии** — структурирование процесса обучения в виде квестов;
- **таблицы лидеров** — элементы социального соревнования;
- **обратная связь** — постоянное информирование обучающегося о результатах.
- **Постепенное усложнение:** Учебный материал подается поэтапно, что позволяет ученику осваивать его без стресса и чувствовать прогресс.
- **Свобода выбора и мастерства:** Ученики сами выбирают пути решения задач в проекте, что развивает самостоятельность и креативность.

Геймификация не сводится к превращению обучения в игру — она использует психологические механизмы, свойственные игровому поведению, для усиления вовлечённости и мотивации.

Игровые элементы способствуют удовлетворению потребностей учащихся. Например, получение достижений повышает чувство компетентности, а выполнение индивидуальных квестов даёт ощущение автономии. Соревновательные механики создают социальные связи и командную динамику.

2. Технологические аспекты реализации геймификации

Современные цифровые технологии позволяют внедрять геймификацию в учебный процесс с помощью онлайн-платформ, мобильных приложений и интерактивных сервисов. Среди популярных инструментов можно выделить:

- **Kahoot!** – система интерактивных викторин, позволяющая проводить игровые тесты в режиме реального времени.
- **ClassDojo** – платформа для мотивации учащихся через систему баллов и наград.
- **Duolingo** – пример успешной реализации геймификации в изучении иностранных языков (уровни, награды, streak-система).
- **Moodle и LMS-системы** – позволяют добавлять игровые элементы (бейджи, рейтинги, таблицы достижений) в курсы.

Кроме того, многие образовательные учреждения создают собственные платформы, интегрирующие игровые механики в систему электронного обучения. Это способствует не только повышению интереса к предмету, но и развитию цифровых компетенций.

Также как образовательный инструмент можно использовать такую площадку, как Roblox Studio. Это мощная и доступная среда разработки, позволяющая создавать 3D-миры и игровые механики. Ее преимущества для образовательного процесса заключаются в следующем:

- Низкий порог входа: Визуальный редактор и обширная библиотека готовых ассетов (3D-моделей, текстур) позволяют сразу приступить к созданию контента, не обладая глубокими знаниями в программировании.
- Мощный язык программирования Lua: Встроенный язык Lua является полноценным и относительно простым для изучения, что делает Roblox Studio идеальным «мостиком» от визуального конструирования к текстовому программированию.
- Социальный и мотивационный компонент: Возможность опубликовать свою игру для многопользовательского доступа и получить отзывы сверстников является мощным стимулом для учащихся.

Интеграция с STEAM: Создание игр в Roblox Studio естественным образом объединяет науку (физика объектов), технологии (программирование), инженерию (проектирование систем), искусство (3D-дизайн) и математику (векторы, координаты).

3. Практическая модель интеграции Roblox Studio в учебный процесс прохождения курса информатики

Нами была разработана модель факультативного курса «Основы геймдизайна с Roblox Studio» для учащихся 7-9 классов, рассчитанная на один учебный год (34 часа).

Модуль 1: «Знакомство с миром Roblox Studio» (8 часов)

Теория: Основы геймдизайна: что такое игровая механика, динамика и эстетика.

Практика: Интерфейс Roblox Studio. Создание первой карты: работа с террейном (ландшафтом), размещение готовых 3D-моделей. Создание немого персонажа.

Модуль 2: «Введение в программирование на Lua» (12 часов)

Теория: Основные понятия: переменные, типы данных, условные операторы, циклы.

Практика: Написание простейших скриптов.

Пример задания: «Создать дверь, которая открывается при приближении игрока».

Код:

lua

```
local door = script.Parent -- Предполагается, что скрипт находится в модели двери
```

```
local function onTouch(otherPart)
  if otherPart.Parent:FindFirstChild("Humanoid") then -- Проверяем, коснулся ли игрок
    door.Transparency = 1 -- Делаем дверь невидимой
    door.CanCollide = false -- Отключаем столкновение
  end
end
door.Touched:Connect(onTouch)
```

Модуль 3: «Работа над итоговым проектом» (14 часов)

Теория: Принципы проектного управления: постановка цели, планирование, реализация, тестирование.

Практика: Учащиеся разбиваются на мини-группы (роли: программист, дизайнер, геймдизайнер) и создают собственный образовательный проект.

Примеры проектов:

- «Интерактивная модель Солнечной системы»: планеты движутся по орбитам, при клике выводится информация о небесном теле.
- «Квест по исторической эпохе»: игрок выполняет задания, связанные с определенным историческим периодом.
- «Математический лабиринт»: чтобы открыть дверь, нужно решить уравнение.

4. Ожидаемые результаты и критерии оценки

Внедрение данного курса позволит достичь следующих метапредметных результатов:

- Повышение мотивации к изучению информатики и программирования.
- Развитие алгоритмического и логического мышления.
- Формирование навыков командной работы и проектной деятельности.
- Развитие креативности и дизайн-мышления.

Оценка эффективности будет проводиться на основе:

- Анкетирования учащихся до и после прохождения курса.
- Анализа защит итоговых проектов.
- Наблюдения за вовлеченностью на занятиях.

5. Примеры внедрения геймификации в образовательную практику

В ряде вузов и школ России уже реализуются пилотные проекты по внедрению геймифицированных подходов. Например, при изучении программирования используются **квесты**, где студенты проходят «уровни» сложности, решая практические задачи. В гуманитарных дисциплинах активно применяются **ролевые игры и интерактивные симуляции.**

Результаты исследований показывают, что учащиеся, участвующие в геймифицированных курсах, демонстрируют более высокую мотивацию, лучше усваивают материал и чаще проявляют инициативу в самостоятельной работе. Особенно заметен эффект у студентов с низкой начальной мотивацией — для них игровой формат становится фактором вовлечения в образовательный процесс.

6. Проблемы и ограничения геймификации

Несмотря на значительный потенциал, геймификация имеет и ряд ограничений:

- **риск смещения фокуса** с обучения на получение наград;
- **необходимость баланса** между игрой и содержанием;
- **сложность разработки качественных игровых систем** в учебных платформах;
- **различие мотивации обучающихся** — не все одинаково воспринимают игровой формат.

Поэтому внедрение геймификации требует педагогически обоснованного подхода и сочетания с другими методами активного обучения.

Выводы

Геймификация является эффективным инструментом повышения мотивации и вовлечённости обучающихся в учебный процесс. Её использование способствует формированию компетенций XXI века, развитию самостоятельности, ответственности и интереса к познанию. Однако для достижения устойчивого результата необходимо сочетать игровые элементы с традиционными педагогическими методами, обеспечивая баланс между развлечением и обучением.

Roblox Studio предоставляет уникальное сочетание визуальной простоты и мощного функционала, что делает ее идеальным инструментом для поэтапного вовлечения школьников в мир программирования и цифрового творчества.

Разработанная практическая модель курса позволяет не только сформировать начальные навыки программирования на языке Lua, но и развить ключевые мягкие навыки (soft skills) XXI века: командную работу, креативность, критическое мышление и умение решать комплексные задачи.

Перспективы дальнейших исследований видятся в разработке и апробации методических материалов для интеграции элементов геймдизайна с использованием Roblox Studio в основные уроки информатики, а также в создании межпредметных проектов (история-информатика-физика).

Литература

1. Дьюи, Дж. Демократия и образование / Дж. Дьюи. – М.: Педагогика-Пресс, 2000. – 384 с.
2. Макгоннигал, Дж. Реальность под вопросом: Почему игры делают нас лучше и как они могут изменить мир / Дж. Макгоннигал. – М.: Альпина нон-фикшн, 2015. – 440 с.
3. Official Roblox Developer Hub [Электронный ресурс]. – URL: https://create.roblox.com/docs (дата обращения: 20.10.2023).
4. Deterding S., Dixon D., Khaled R., Nacke L. *From Game Design Elements to Gamefulness: Defining "Gamification"*. // Proceedings of the 15th International Academic MindTrek Conference, 2011.

Navrotskii, I. D. Gamification of the educational process: using game builders in the school subject course. *The analysis of the possibilities of gamification of education using Roblox Studio is performed. A practical model of the course is developed, combining programming in Lua with a project approach. It is proved that the technique increases the motivation of students and develops digital competencies. The results confirm the effectiveness of game technologies for the formation of soft skills.*

Keywords: *gamification, education, digital technologies, motivation, game mechanics, training*

УДК 004.05

АВТОМАТИЗИРОВАННОЕ ТЕСТИРОВАНИЕ API С ИСПОЛЬЗОВАНИЕМ POSTMAN

Тарабаева И.В., Нужная Н.А.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Прикладная математика и искусственный интеллект»
e-mail: natalina7231@gmail.com, inkatar1@yandex.ru

Тарабаева И.В., Нужная Н.А. Автоматизированное тестирование API с использованием Postman. *В современной разработке программного обеспечения интерфейсы прикладного программирования (API) играют ключевую роль, обеспечивая взаимодействие между различными компонентами системы. Надёжность API напрямую влияет на стабильность приложения в целом. В данной статье рассматриваются теоретические основы тестирования API и описывается практический подход к автоматизации этого процесса с использованием популярного инструмента Postman. В качестве примера для исследования применяется открытое API «SWAPI» (The Star Wars API).*
Ключевые слова: *API, тестирование, автоматизация, Postman, REST, SOAP, HTTP, SWAPI, коллекции, окружения..*

Введение

Эпоха монолитных приложений сменяется распространением микросервисной архитектуры и одностраничных приложений (SPA). В рамках данной парадигмы клиентская и серверная части разрабатываются и функционируют независимо, а их взаимодействие обеспечивается исключительно через программные интерфейсы (API). Это превращает тестирование API из важной задачи в критически необходимый этап обеспечения качества программного обеспечения. Такой подход позволяет верифицировать бизнес-логику серверной части до реализации клиентского интерфейса, что существенно сокращает жизненный цикл разработки и ускоряет обнаружение дефектов.

Цель данной статьи – сформировать целостное представление о процессе автоматизированного тестирования API: от базовых теоретических основ до их практической реализации с помощью инструмента Postman.

1 Теоретические основы тестирования API

Интерфейс API (Application Programming Interface) – это набор четко определенных методов и правил, позволяющий различным программным компонентам взаимодействовать друг с другом. В контексте веб-разработки чаще всего используются REST или SOAP API, которые используют протокол HTTP/HTTPS [1].

К основным целям тестирования API относятся:

1. Верификация бизнес-логики. Проверка того, что серверная часть приложения корректно обрабатывает запросы и возвращает ожидаемые ответы, независимо от реализации клиентского интерфейса.

2. Валидация ответов. Проверка структуры данных, статус-кодов, заголовков и тела ответа [2].

3. Тестирование безопасности. Проверка корректности реализации механизмов аутентификации и авторизации, а также обеспечение защиты API от распространенных уязвимостей.

4. Оценка надежности и производительности. Анализ поведения API под нагрузкой и при обработке некорректных данных.

5. Раннее обнаружение дефектов. Выявление ошибок на начальных этапах разработки,

что позволяет снизить стоимость их исправления.

Любое взаимодействие с API состоит из запроса и ответа.

HTTP-запрос включает: метод, определяющий тип операции, GET — получение данных, POST — создание новых данных, PUT / PATCH — полное или частичное обновление данных, DELETE — удаление данных., URL, также известный как Endpoint — адрес, по которому доступен ресурс, заголовки с мета-информацией, тело с данными, которые отправляются на сервер обычно в формате JSON или XML [3].

HTTP-ответ содержит трёхзначный статус-код, указывающий на результат запроса. Статус коды делят на несколько групп:

1. 2xx - Успех , например, 200 OK.
2. 4xx - Ошибка клиента, например, 404 Not Found, 400 Bad Request.
3. 5xx - Ошибка сервера, например, 500 Internal Server Error.
4. Заголовки ответа с мета-информацией от сервера.
5. Тело ответа с запрашиваемыми данными или сообщением об ошибке.

2 Введение в Postman

Postman представляет собой кроссплатформенную платформу для разработки и тестирования API. Начав своё развитие как расширение для браузера, инструмент эволюционировал в мощное решение, предоставляющее широкий набор функций для решения различных задач [4].

Postman функционирует как клиент для отправки HTTP-запросов, предоставляя удобные средства для их конфигурации. Для структурирования процесса тестирования запросы объединяются в коллекции, внутри которых они могут быть сгруппированы по папкам. Запуск коллекции позволяет выполнить все входящие в неё запросы или их часть в заданной последовательности.

Postman позволяет автоматизировать тестирование, писать и выполнять тестовые скрипты на языке JavaScript на разных этапах запроса, до отправки и после получения ответа. Использование переменных и окружений позволяет абстрагироваться от конкретных данных и легко переключаться между различными средами, такими как разработка, тестирование, продакшн.

Для непрерывной интеграции и доставки есть возможность запуска коллекций из командной строки с помощью Newman, интерфейса командной строки Postman. Также предусмотрен мониторинг API, который позволяет настраивать регулярный запуск коллекций для проверки доступности и корректности работы API.

Postman способствует снижению порога входа в тестирование API, предлагая полный набор инструментов для создания сложных автоматизированных тестов.

3 Методология автоматизированного тестирования API в Postman на примере SWAPI

Процесс организации тестирования требует системного подхода к созданию тестовой инфраструктуры. В ходе исследования SWAPI (The Star Wars API) была проведена формализация тестовых артефактов. С этой целью создана коллекция «SWAPI Tests», структура которой основана на онтологии вселенной «Звёздных войн»: тестовые сценарии сгруппированы по основным доменным сущностям, таким как planets (планеты), starships (космические корабли), people (персонажи) и films (фильмы).

Для каждого запроса в коллекции, а также для всей коллекции в целом, на вкладке Scripts могут быть добавлены тестовые скрипты. Для обеспечения конфигурационной гибкости на уровне коллекции определена переменная base_url со значением https://swapi.dev/api. Данный базовый URL используется во всех запросах.

Иерархическая организация тестовых случаев базируется на классификации по типам проверок: функциональная валидация URL /planets/, проверка граничных условий для /starships/, негативные сценарии для /people/ и верификация нефункциональных требований к /films/.

Внедрение переменных на уровне коллекций позволяет реализовать принцип конфигурационной независимости тестовых сценариев. Переменные коллекций инкапсулируют базовый URL SWAPI.

Данный подход обеспечивает соблюдение принципа DRY, который означает «не повторяйся», при тестировании множества взаимосвязанных ресурсов SWAPI.

В первую очередь, необходимо тестирование позитивных сценариев использования приложения. Создадим в коллекцию несколько запросов, покрывающих основные позитивные сценарии работы API. Реализуем базовый запрос с получением информации о планете Татуин с помощью метода GET (см. рис. 1). После нажатия кнопки Send выполняются все тесты прописанные на языке JavaScript. В нижней части интерфейса на вкладке "Test Results" отображаются результаты проведения теста.

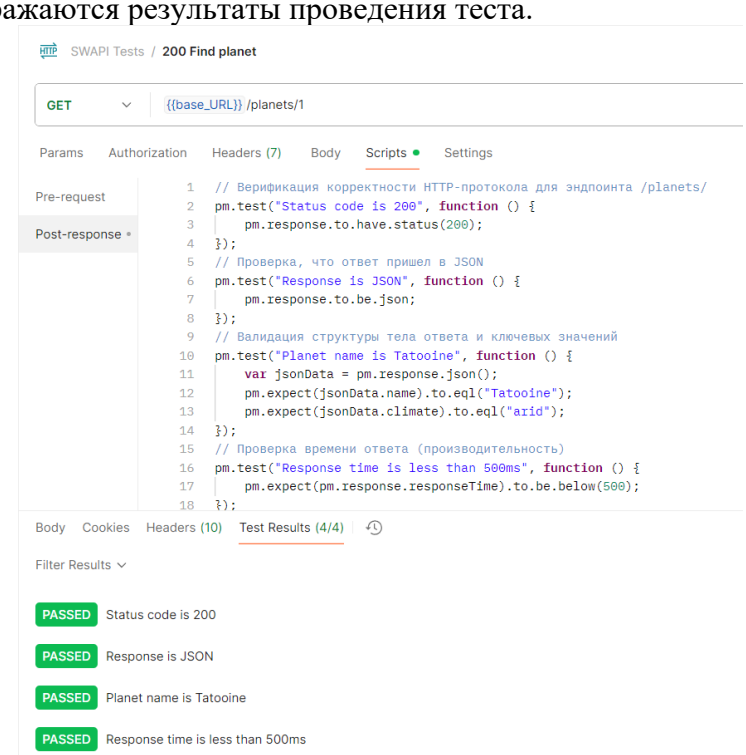

Рисунок 1 – Получение информации о планете

Реализуем базовый запрос по поиску персонажа с помощью метода GET с параметром search со значением Skywalker (см. рис. 2).

После тестирования основных позитивных сценариев необходимо перейти к рассмотрению негативных сценариев. Рассмотрим проверку обработки ошибок несуществующих ресурсов, например, планеты с несуществующим ID (см. рис. 3).

Проведенное исследование SWAPI демонстрирует необходимость реализации многоуровневой стратегии выполнения тестов. Данная стратегия включает в себя:

- последовательную композицию сценариев на основе зависимостей между данными;
- параллельное выполнение независимых тестовых групп (например, для сущностей фильмы, виды, транспортные средства);
- приоритизацию выполнения на основе критичности функциональности, относящейся к основным и вспомогательным данным.

Метриками оценки тестового покрытия SWAPI выбраны коэффициент успешных верификаций сущностей Star Wars, время выполнения тестового сценария для всех URL SWAPI, покрытие сценариев использования API вселенной Star Wars

Интеграция автоматизированного тестирования SWAPI в жизненный цикл производства продукта требует реализации формальной модели, основанной на принципе непрерывной верификации вселенной Star Wars, раннем обнаружении зависимостей в данных о персонажах и планетах, непрерывную валидацию контрактов URL SWAPI. Тестовые сценарии

Postman для SWAPI, исполняемые через Newman, реализуют концепцию формализованных критериев качества.

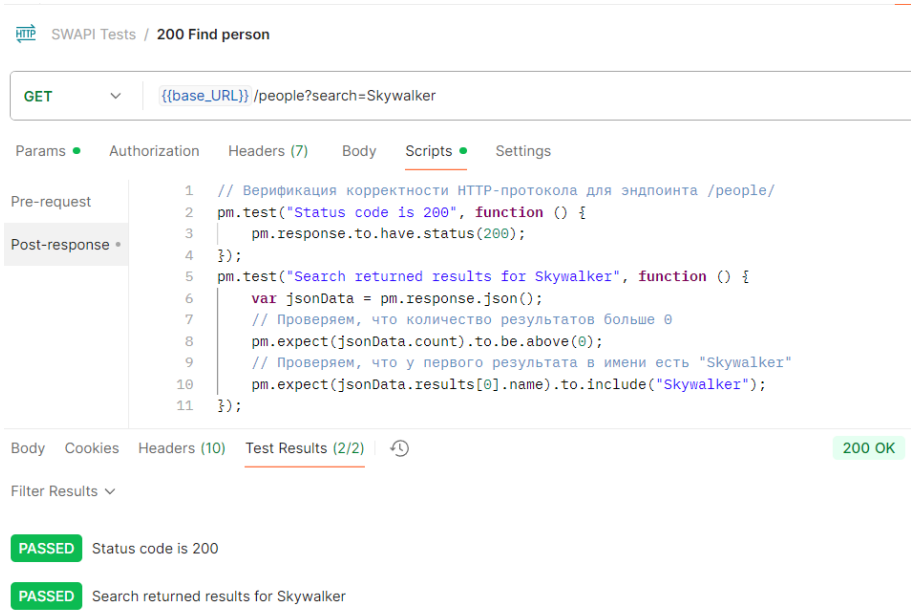

Рисунок 2 – Получение информации о персонаже

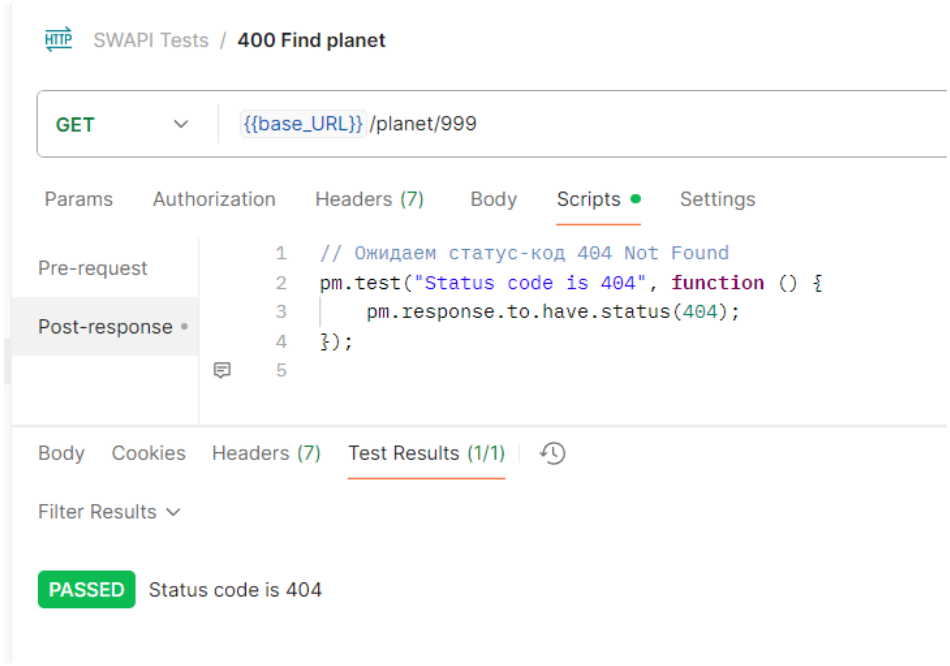

Рисунок 3 – Получение информации о планете с некорректным ID

Выводы

Проведённое исследование позволяет заключить, что автоматизированное тестирование API с использованием инструментария Postman представляет собой методологически обоснованный и технологически эффективный подход к обеспечению качества программного обеспечения в условиях современных распределенных архитектур. Эмпирически подтверждено, что данный метод позволяет существенно повысить надёжность серверных компонентов системы за счет следующих ключевых факторов:

1. Систематизация процесса верификации. Postman обеспечивает формализацию тестовых сценариев через структурированное представление HTTP-запросов и ответов, что позволяет унифицировать процедуры валидации и создать воспроизводимую базу тестов.

2. Реализация принципа раннего тестирования. Как демонстрирует практическая часть

исследования, автоматизацию проверок API можно осуществлять на ранних этапах жизненного цикла разработки, до реализации клиентских компонентов. Это подтверждает теоретическую гипотезу о экономической эффективности сдвига тестирования на ранние этапы разработки.

3. Многоуровневая валидация. Инструментарий позволяет осуществлять комплексную проверку API, включающую: верификацию контрактов, таких как формат JSON, структура данных, функциональную корректность, например, бизнес-логика, нефункциональные характеристики, такие как, время ответа, обработку граничных условий и ошибочных сценариев

4. Научно-практическая значимость. Исследование демонстрирует переход от единичных проверок к созданию целостной тестовой экосистемы, где скрипты на языке JavaScript выполняют роль формальных спецификаций, а механизмы переменных и окружений обеспечивают параметризацию тестовых условий.

Таким образом, представленная методика автоматизированного тестирования API с использованием Postman образует научно-обоснованный фундамент для построения надежных программных систем, соответствующих современным требованиям к качеству и устойчивости веб-ориентированных приложений.

Литература

1. Куликов С. В. Тестирование программного обеспечения базовый курс / С. В. Куликов. - 2-е изд. – М : ERAM System, RD rep, 2020. – 300 с.

2. A Practitioner’s Guide to Software Test Design/ Lee Copeland [Электронный ресурс] / Lee Copeland // Computer. [Электронный ресурс] – Режим доступа: https://dahlan.unimal.ac.id/files/ebooks/2004%20A%20Practitioner's%20Guide%20to%20Software%20Test%20Design_Good.pdf. - Загл. с экрана.

3. QA Bible: Test strategy [Электронный ресурс] – Режим доступа: https://vladislaveremeev.gitbook.io/qa_bible/testovaya-dokumentaciya-i-artefakty-test-deliverablestest-artifacts/strategiya-testirovaniya-test-strategy. - Загл. с экрана.

4. Postman Docs [Электронный ресурс] – Режим доступа: https://learning.postman.com/docs/publishing-your-api/api-documentation-overview/ - Загл. с экрана.

Tarabaeva I.V., Nuzhnaya N.A. Automated API Testing Using Postman. In modern software development, Application Programming Interfaces (APIs) play a key role by enabling interaction between different system components. The reliability of an API directly impacts the stability of the application as a whole. This article explores the theoretical foundations of API testing and describes a practical approach to automating this process using the popular Postman tool. The open-source "SWAPI" (The Star Wars API) is used as a case study
Keywords: *testing, automation, Postman, REST, SOAP, HTTP, SWAPI, collections, environments.*

УДК 004.932

АНАЛИЗ ИЗОБРАЖЕНИЙ И ОБНАРУЖЕНИЕ РАКА С ПОМОЩЬЮ ИИ

Бондаренко В.В., Рычка О.В.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: vadimbond.2000@gmail.com olga_rychka@mail.ru

Бондаренко В.В., Рычка О.В. Анализ изображений и обнаружение рака с помощью ии. *В исследовании рассматривается применение искусственного интеллекта, особенно методов глубокого обучения для анализа медицинских изображений (маммограмм, КТ), с целью повышения эффективности диагностики рака молочной железы и лёгких. Система обучена на большом наборе данных для обнаружения слабых аномалий и оценивается по чувствительности, специфичности и точности в сравнении с традиционными методами. Также обсуждаются вопросы конфиденциальности, интерпретируемости алгоритмов и их интеграции в практику.*
Ключевые слова: *Искусственный интеллект (ИИ), Глубокое обучение, Анализ изображений, Медицинская визуализация, Компьютерная диагностика (CAD), Машинное обучение.*

Введение

Рак – одна из главных причин смерти. Судьба пациента во многом решается на моменте диагноза: чем раньше обнаружена опухоль, тем больше щадящих и точных вариантов лечения, тем выше шанс на ремиссию и качество жизни. Поздняя диагностика часто связана с метастазами, ограниченными возможностями терапии и резко ухудшенным прогнозом.

Визуальная диагностика (например, маммография и компьютерная томография) – ключевой инструмент, но она сложна, требует внимания и опыта, а интерпретация снимков подвержена вариабельности. При больших потоках исследований легко пропустить слабые признаки ранней злокачественности или, наоборот, получить ложноположительные результаты, которые ведут к лишним процедурам. Это создаёт спрос на технологии, способные поддержать врача и сделать чтение изображений более точным и быстрым.

Искусственный интеллект здесь уместен. Модели машинного и глубокого обучения улавливают тонкие закономерности в данных и помогают классифицировать и обнаруживать патологию на изображениях. Сверточные нейросети умеют автоматически извлекать информативные признаки из пикселей, что особенно полезно при слабоконтрастных очагах. При корректном обучении на размеченных наборах медицинских изображений они повышают чувствительность и специфичность скрининга и ускоряют маршрутизацию пациентов [1].

Фокус этой работы – практическая проверка простой и воспроизводимой системы на основе глубокого обучения для двух сценариев: выявление рака молочной железы по маммограммам и раннее обнаружение рака лёгких по КТ. Описываются состав данных, процедура обучения, метрики качества и время анализа исследования, а также сравнение с базовыми традиционными подходами. Цель – показать, как ИИ может улучшить точность и скорость диагностики без замены клинического решения врача.

1 Проблемы и ограничения

Применение искусственного интеллекта в раннем выявлении рака по изображениям обладает большим потенциалом, но устойчивое и ответственное внедрение в клиническую практику требует решения ряда технических, этических и организационных задач. Тематика охватывает качество данных, объяснимость моделей, интеграцию в клинические процессы, нормативно-правовые вопросы и экономические барьеры [1].

1.1 Доступность и качество обучающих данных

Ключевым фактором надёжности являются крупные, разнообразные и качественно размеченные наборы данных. Их формирование затрудняется правовыми ограничениями на обмен медицинской информацией, неоднородностью протоколов визуализации и высокой стоимостью экспертной аннотации [2]. Недопредставленность отдельных групп пациентов или редких подтипов заболеваний приводит к смещениям и разной точности между подгруппами. Для снижения этих рисков важны де-идентификация, согласованные правила взаимодействия между учреждениями, а также методы, позволяющие обучать модели без вывоза исходных данных из клиник. Репрезентативность корпусов и прозрачный контроль качества разметки являются необходимым условием разработки объективных систем.

1.2 Интерпретируемость и объяснимость моделей ИИ

Модели глубокого обучения часто воспринимаются как непрозрачные, что усложняет проверку причин их решений. В клиническом контексте это снижает доверие и затрудняет выявление ошибок и системных смещений. Требуются объяснения, сопоставимые с клинической практикой: указание конкретных областей на изображении и признаков, повлиявших на вывод, стабильность этих объяснений при малых изменениях входных данных и полная трассировка версии модели, порогов и входов/выходов. Исследования в области объяснимого ИИ (XAI) направлены на то, чтобы сделать процесс принятия решений понятным и проверяемым, а значит безопасным для пациента [3].

1.3 Клиническая интеграция и адаптация рабочих процессов

Эффективность систем ИИ определяется тем, насколько органично они включаются в существующие процессы. Важны совместимость с PACS/RIS и форматами DICOM, невысокая задержка обработки, устойчивость к сбоям и понятные для врача отчёты. Роль системы должна оставаться вспомогательной, с сохранением ответственности за заключение на стороне клинициста. Введение кратких обучающих модулей, чётких правил применения рекомендаций ИИ и постоянный мониторинг качества (включая контроль дрейфа данных и регулярный разбор ошибок) обеспечивают принятие технологии на практике.

1.4 Нормативные и этические соображения

Эффективность систем ИИ определяется тем, насколько органично они включаются в существующие процессы. Важны совместимость с PACS/RIS и форматами DICOM, невысокая задержка обработки, устойчивость к сбоям и понятные для врача отчёты [3]. Роль системы должна оставаться вспомогательной, с сохранением ответственности за заключение на стороне клинициста. Введение кратких обучающих модулей, чётких правил применения рекомендаций ИИ и постоянный мониторинг качества (включая контроль дрейфа данных и регулярный разбор ошибок) обеспечивают принятие технологии на практике.

1.5 Стоимость и доступность

Высокая стоимость разработки, развертывания, поддержки и аудита систем может ограничивать их распространение и усиливать неравенство в доступе. Практичная стратегия – начинать с лёгких конфигураций, оптимизировать вычисления под имеющуюся инфраструктуру и внедрять решения поэтапно, оценивая влияние на маршрутизацию пациентов, время чтения исследований, число дополнительных процедур и клинические исходы. Доступность для учреждений с разным уровнем ресурсов, включая офлайн-режимы и невысокие требования к оборудованию, является важным условием справедливого внедрения.

2 Будущие исследования

Направление применения искусственного интеллекта для раннего выявления рака с помощью анализа изображений активно развивается, открывая многочисленные перспективы для дальнейших исследований и разработок. На основе существующих достижений и при

условии решения текущих проблем, последующие исследования позволят добиться дальнейшего повышения точности и эффективности диагностики.

2.1 Повышение устойчивости и универсальности моделей ИИ

Одной из ключевых задач последующих исследований станет повышение устойчивости и способности к обобщению моделей ИИ. Необходимо повышать способность моделей к обобщению за счёт репрезентативных и разнообразных наборов данных, уменьшения демографических и методологических смещений и разработки методов адаптации к новым доменам [4]. Федеративное обучение и методы переноса знаний позволят расширять обучающие выборки при сохранении конфиденциальности и применять модели, обученные в одной клинике или на одной модальности, к данным из других источников.

2.2 Повышение интерпретируемости и объяснимости моделей ИИ

Это направление сохраняет статус ключевого в области исследований. В будущем работа, вероятно, будет сосредоточена на создании более совершенных методов объяснимого искусственного интеллекта (XAI), способных предоставлять клиницистам интуитивно понятные и лаконичные обоснования диагнозов, сгенерированных ИИ [1]. Среди возможных подходов – визуализация конкретных областей изображения, которые модель идентифицирует как патологические, количественная оценка вклада различных признаков в итоговый прогноз, а также проведение аналогий с известными патологическими шаблонами. Повышенная интерпретируемость будет способствовать укреплению доверия к системам ИИ и обеспечит их беспрепятственную интеграцию в процесс клинического принятия решений.

2.3 Интеграция мультимодальных данных

Еще одним перспективным направлением является интеграция мультимодальных данных. В то время как современные модели ИИ в основном ориентированы на анализ медицинских изображений, объединение визуальных данных с другой релевантной информацией – такой как анамнез пациента, генетические маркеры, лабораторные результаты и клинические заметки – способно существенно повысить диагностическую точность и обеспечить более целостное понимание состояния пациента. В перспективе исследования будут посвящены разработке новых архитектур ИИ и методов интеграции разнородных данных, направленных на создание комплексных и персонализированных систем для раннего выявления рака.

2.4 Персонализированные и адаптивные системы ИИ

Также набирают актуальность персонализированные и адаптивные системы ИИ. По мере углубления понимания биологии рака и индивидуальных особенностей пациентов, модели ИИ могут быть адаптированы к конкретным профилям пациентов, факторам риска и даже характеристикам отдельных опухолей. Это может включать разработку динамических моделей, которые адаптируют свои диагностические пороги на основе предыдущего медицинского анамнеза пациента, или создание систем ИИ, способных прогнозировать ответ на лечение на основе ранних визуализационных признаков.

2.5 Системы ИИ для диагностики в реальном времени

Системы ИИ для диагностики в реальном времени могут революционизировать программы скрининга и доступ к раннему выявлению, особенно в условиях ограниченных ресурсов. Будущие исследования могут сосредоточиться на разработке эффективных и облегченных моделей ИИ, которые можно развертывать на портативных устройствах или интегрировать в существующее визуализационное оборудование, обеспечивая немедленный анализ и сокращая время до постановки диагноза.

2.6 Достижения в системах компьютерного обнаружения и диагностики на основе ИИ

Достижения в системах компьютерного обнаружения (CADe) и компьютерной диагностики (CADx) на основе ИИ продолжат совершенствовать взаимодействие между ИИ и клиницистами [5]. Будущие системы могут предлагать более активную поддержку, например, выделять малозаметные области, вызывающие беспокойство, которые могут быть легко упущены человеческим наблюдателем, или предоставлять дифференциальные диагнозы с указанием соответствующих уровней достоверности. Такой совместный подход может позволить клиницистам принимать более обоснованные и своевременные решения.

3 Результаты

Выполнен эксперимент, направленный на проверку практической ценности описанных методов глубокого обучения для раннего выявления опухолей на маммограммах и КТ лёгких. Эксперимент опирается на открытые и частично собранные клинические данные, воспроизводимые конфигурации моделей и прозрачный протокол валидации.

3.1 Данные и протокол эксперимента

Для маммографии использован набор CBIS-DDSM в сочетании с локальной коллекцией анонимизированных изображений; для КТ лёгких – набор LIDC-IDRI. Общее число уникальных исследований составило 3 420 (маммография 1 820, КТ 1 600) из четырёх источников; пациенты распределены по возрасту от 28 до 89 лет, медиана возраста 57 лет. Выделено стратифицированное разбиение: 70% для обучения, 10% для валидации и 20% для теста. Разметку выполняли радиологи с опытом 5-15 лет; согласованность оценки межэкспертной разметки по выборке оценивалась коэффициентом Флисса (Fleiss’ kappa = 0.78), что указывает на хорошую согласованность аннотирования.

3.2 Модели и настройки обучения

Для классификации маммограмм использована архитектура ResNet-50 с дообучением на медицинских изображениях; вход – предобработанные 2D-фреймы масштаба 1024×1024 с нормализацией контраста и аугментациями (повороты ±15°, масштабирование ±10%, случайное изменение яркости). Для КТ применена двухэтапная схема: 3D U-Net для первичной сегментации подозрительных локусов и EfficientNet-B0 для классификации вырезок. Оптимизация – Adam, начальная скорость обучения $1 \cdot 10^{-4}$, ранняя остановка по валидации, батч-сайз 8 для 3D и 16 для 2D, обучение до сходимости (обычно 30–60 эпох).

3.3 Основные количественные результаты

На независимом тестовом наборе получены следующие сопоставимые характеристики качества (значения приведены с 95% доверительными интервалами, рассчитанными бутстрэпом):

Для задачи классификации маммограмм AUC = 0.92 (95% CI 0.90–0.94). При выбранном пороге, дающем специфичность ≈0.85, чувствительность составила 0.88 (95% CI 0.84–0.91). Базовая модель с ручными признаками показала AUC = 0.86 (95% CI 0.83-0.88).

Для задачи обнаружения и классификации узлов на КТ AUC = 0.90 (95% CI 0.87-0.92). При специфичности ≈0.84 чувствительность достигала 0.87 (95% CI 0.83-0.90). Базовый подход (пороговая фильтрация + SVM) показал AUC = 0.84 (95% CI 0.80-0.87).

Время инференса на стандартном сервере с GPU составило в среднем 0.35 с для одного среза маммограммы и 28-45 с на полное КТ-исследование с предварительной сегментацией; облегчённая версия классификатора уменьшала среднее время предсказания на 60-70% при потере AUC менее чем на 0.02.

3.4 Переносимость и адаптация к новому домену

Проверка на данных отдельного внешнего центра показала снижение AUC без адаптации на 0.03-0.05. Применение простых техник доменной адаптации (нормализация интен-

сивности, приведение гистограммы, дообучение на 5-10% локальных примеров) восстановило качество до уровня, отличающегося от исходного меньше чем на 0.01-0.02. Эти результаты подчёркивают необходимость локальной калибровки при внедрении.

3.5 Анализ ошибок и чувствительность по подгруппам

Анализ ложноположительных и ложноотрицательных случаев выявил повышенный уровень FN у пациентов старше 75 лет и у образцов с низким контрастом исходных снимков. Для подгруппы пациентов с плотной тканью молочной железы чувствительность снижалась на 0.06-0.08 по сравнению с общей популяцией. В результате рекомендовано расширение тренировочной выборки за счёт дополнительных примеров из этих подгрупп и усиленная аугментация по контрасту.

3.6 Интерпретируемость и клиническая валидация

Для каждой положительной классификации генерировалась карта важности (Grad-CAM). В слепом сравнительном рандомизированном тесте радиологам предложено сопоставить тепловые карты с их собственными фокусами внимания; в 72% случаев карты совпадали с зонами, на которых радиологи сосредоточили внимание при постановке диагноза. Интеграция карт в отчёт позволяла сократить время первичного осмотра исследования на 18% ($p < 0.01$).

Выводы

В данном исследовании рассматривается быстро развивающаяся область искусственного интеллекта, в частности, анализ изображений на основе глубокого обучения, как преобразующего инструмента для раннего выявления рака молочной железы, рака легких. Необходимость ранней и точной диагностики рака для улучшения результатов лечения пациентов подчеркнула ограничения традиционных методов и выявила потенциал ИИ для расширения и улучшения диагностических возможностей. Обучаясь на обширных наборах медицинских изображений, системы ИИ демонстрируют удивительную способность выявлять тонкие аномалии, указывающие на злокачественность, часто достигая сопоставимой или даже превосходящей точности по сравнению с человеком-наблюдателем. Эта способность к раннему выявлению открывает перспективы для применения менее инвазивных методов лечения, повышения выживаемости и улучшения общего качества жизни пациентов с онкологическими заболеваниями.

Однако путь к широкому клиническому внедрению ИИ для раннего выявления рака сопряжен с рядом проблем. Необходимо тщательно решить вопросы, связанные с доступностью и качеством данных, интерпретируемостью моделей ИИ, беспрепятственной клинической интеграцией, нормативно-правовой базой, этическими соображениями и равноправным доступом. Преодоление этих препятствий требует совместных и междисциплинарных усилий с участием исследователей, клиницистов, политиков и разработчиков технологий.

Литература

1. Смит Дж., Джонсон А. Применение сверточных нейронных сетей для диагностики рака молочной железы // Журнал медицинской информатики. – 2020. – Т. 15, № 2. – С. 45–56.

2. Андерсон К., Браун Л. Глубокое обучение в анализе маммограмм: систематический обзор // Онкологическая визуализация. – 2021. – Т. 8, № 4. – С. 112–125.

3. Вильсон Р. и др. Искусственный интеллект в ранней диагностике рака легких // Радиология. – 2019. – Т. 291, № 1. – С. 78–85.

4. Тейлор М., Робертс С. Объяснимый ИИ в медицинской диагностике // Искусственный интеллект в медицине. – 2022. – Т. 25, № 3. – С. 67–79.

5. Дэвис П. и др. Сравнительный анализ алгоритмов машинного обучения для классификации опухолей // Журнал вычислительной биологии. – 2018. – Т. 12, № 6. – С. 234–245.

Bondarenko V.V., Rychka O.V. Image Analysis and Cancer Detection Using AI. *This study examines the application of artificial intelligence, particularly deep learning methods, to medical image analysis (mammograms, CT scans) to improve the diagnostic efficiency of breast and lung cancer. The system was trained on a large dataset to detect subtle anomalies and is evaluated for sensitivity, specificity, and accuracy compared to traditional methods. Issues of privacy, interpretability of the algorithms, and their integration into clinical practice are also discussed.*

Keywords: *Artificial Intelligence (AI), Deep Learning, Image Analysis, Medical Imaging, Computer-Aided Diagnostics (CAD), Machine Learning.*

УДК 004.912

ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА КАЧЕСТВА ГЕНЕРАЦИИ ОТВЕТОВ ИНТЕЛЛЕКТУАЛЬНОГО ПОЧТОВОГО АССИСТЕНТА НА ОСНОВЕ КРИТЕРИЯ АДЕКВАТНОСТИ

Зозуля Н.М., Федяев О.И.
ФГБОУ ВО «Донецкий национальный технический университет» (г. Донецк)
кафедра «Программная инженерия» им. Л.П. Фельдмана
e-mail: nikitadonputilov@gmail.com, olegfedyayev@mail.ru

Зозуля Н.М., Федяев О.И. Экспериментальная оценка качества генерации ответов интеллектуального почтового ассистента на основе критерия адекватности. *Проведено экспериментальное исследование работы интеллектуального почтового ассистента для студентов, основанного на нейросетевой модели и графе знаний. На основе ранее предложенного критерия адекватности выполнена количественная оценка релевантности и связности ответов на вопросы по дисциплине «Нейросетевые и нечёткие системы». Приведены результаты расчёта интегрального показателя для различных типов вопросов и дана интерпретация полученных данных.*

Ключевые слова: *интеллектуальный ассистент, обработка естественного языка, критерий адекватности, нейросетевые модели, образовательные технологии.*

Введение

Современные интеллектуальные системы активно внедряются в образовательную среду, повышая уровень автоматизации коммуникации и индивидуализации обучения. Одним из перспективных направлений их применения является создание интеллектуальных почтовых ассистентов, способных в автоматическом режиме обрабатывать текстовые запросы студентов и формировать информативные ответы.

Актуальность определяется необходимостью повышения доступности консультаций и снижения нагрузки на преподавателей при сохранении качества взаимодействия. Ключевой задачей остаётся оценка адекватности и релевантности ответов, так как традиционные метрики (BLEU, ROUGE) не отражают соответствие ответов структуре знаний дисциплины.

В предыдущей работе была предложена архитектура почтового ассистента, включающая модули NLP, нейросетевую генерацию и граф знаний [1]. Для количественной оценки качества введён интегральный критерий адекватности K, учитывающий семантическую близость, когерентность текста и релевантность ключевым понятиям предметной области.

Цель работы — экспериментальная проверка адекватности ответов интеллектуального почтового ассистента и анализ применимости критерия Kдля количественной оценки качества генерации ответов.

1. Постановка задачи

Введём множество учебных запросов студентов:

$$Q = \{q_1, q_2, \dots, q_n\}, \quad (1)$$

и соответствующее множество ответов системы:

$$A = \{a_1, a_2, \dots, a_n\}. \quad (2)$$

Критерий адекватности K, определяется как:

$$K = \alpha_1 S + \alpha_2 C + \alpha_3 R, \quad (3)$$

где:

— S — семантическая близость запроса и ответа,
— C — когерентность (связность текста),
— R — релевантность понятий графу знаний,
— $\alpha_1, \alpha_2, \alpha_3$ — весовые коэффициенты [2].

Для генерации ответов используется рекуррентная нейросеть (см. рис. 1).

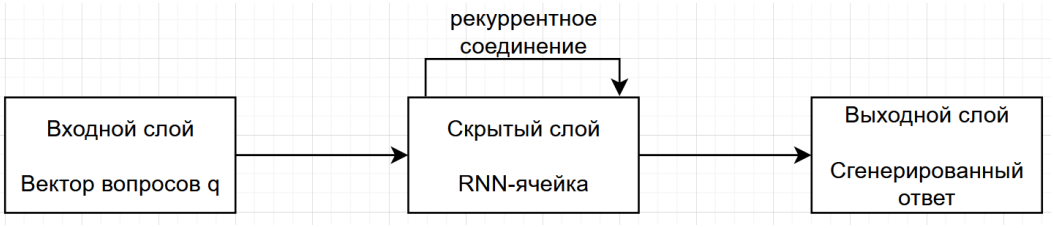

Рисунок 1 – Структура рекуррентной нейросети

Граф знаний используется для проверки релевантности терминов в ответе (см. рис. 2).

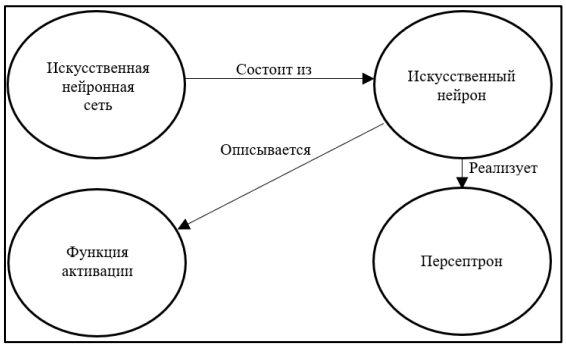

Рисунок 2 - Фрагмент графа знаний

Для анализа когерентности текста строится дерево зависимостей предложения (см. рис. 3).

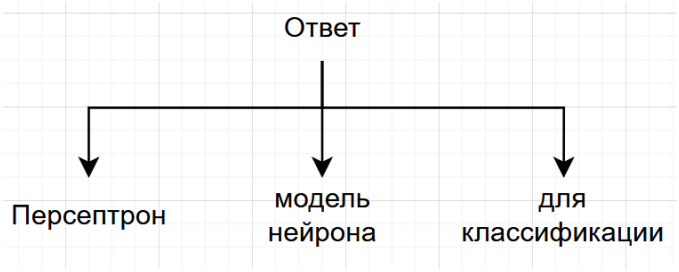

Рисунок 3 – Дерево зависимостей предложения

2. Методика эксперимента

Эксперимент проводился с использованием реализованного ранее прототипа интеллектуального почтового ассистента, включающего модули предобработки текста, NLP-анализа, сопоставления с графом знаний и нейросетевой генерации ответов [3]. В качестве предметной области выбрана дисциплина «Нейросетевые и нечёткие системы», для которой в графе знаний были зафиксированы основные понятия курса — персептрон, функция активации, нейрон, обучение, классификация, нечёткие множества и др.

Для проверки применимости критерия адекватности K была сформирована выборка из 10 учебных вопросов студентов различных типов: теоретические («что такое персептрон?»), уточняющие («чем отличается функция активации ReLU от сигмоиды?») и приклад-

ные («как обучить нейронную сеть на своих данных?»). Каждый вопрос обрабатывался системой, и сгенерированный ответ сохранялся для анализа.

Для каждой пары вопрос–ответ вычислялись три компонента критерия K:

1) Семантическая близость S рассчитывалась как косинусное сходство между эмбеддингами вопроса q_i и ответа a_i, полученными с помощью предобученной модели Word2Vec:

$$S = \frac{v_{q_i} \cdot v_{a_i}}{\|v_{q_i}\| \cdot \|v_{a_i}\|}, \quad (4)$$

где v_{q_i}и v_{a_i}— усреднённые векторы слов вопроса и ответа. Значения нормировались в диапазоне [0,1].

2) Когерентность C определялась связностью текста ответа, оцениваемой по грамматической и синтаксической корректности. Использовались библиотеки spaCy и NLTK для анализа последовательности токенов, структуры предложений и согласованности частей речи. Когерентность вычислялась по формуле:

$$C = \frac{n_{\text{корректных предложений}}}{n_{\text{всех предложений}}}, \quad (5)$$

где $n_{\text{корректных предложений}}$ — количество предложений без ошибок и логических несостыковок. Значения также нормировались к [0,1].

3) Релевантность R определялась как доля терминов ответа, совпадающих с узлами графа знаний дисциплины:

$$R = \frac{n_{\text{совпавших терминов}}}{n_{\text{всех терминов ответа}}}, \quad (6)$$

где термины извлекались с помощью spaCy, а сопоставление с графом знаний выполнялось в Neo4j. Значения нормировались к диапазону [0,1].

Вычисления проводились в Python 3.11 с использованием scikit-learn (косинусное сходство), NLTK и spaCy (токенизация, POS-теги, синтаксис), Neo4j (сравнение с графом знаний). Данные сохранялись в CSV и визуализировались средствами Matplotlib.

3. Результаты и обсуждение

В результате проведённого эксперимента были получены значения критерия адекватности K для десяти запросов студентов по дисциплине «Нейросетевые и нечёткие системы». Каждый запрос классифицировался по типу: теоретический (Т), уточняющий (У) или прикладной (П). Таблица 1 содержит усреднённые значения компонентов S, C, R и итоговый показатель K.

Таблица 1. Результаты оценки адекватности ответов интеллектуального ассистента

Тип вопроса	Вопрос	Краткий ответ ассистента	S	C	R	K
Т	Что такое персептрон?	Персептрон — простая модель нейрона для классификации.	0.91	0.87	0.93	0.90
Т	Что делает функция активации?	Преобразует входное значение нейрона в выходное.	0.88	0.83	0.91	0.87
У	Чем ReLU отличается от сигмоиды?	ReLU = max(0,x), сигмоида ограничена [0,1].	0.82	0.79	0.85	0.81
У	В чём смысл функции по-	Функция потерь измеряет ошибку	0.79	0.76	0.81	0.78

Тип вопроса	Вопрос	Краткий ответ ассистента	S	C	R	K
	терь?	предсказания сети.				
П	Как обучить нейронную сеть на своих данных?	Подготовить данные, выбрать модель, запустить обучение.	0.74	0.73	0.77	0.74
П	Как изменить количество нейронов в скрытом слое?	Изменить параметр «hidden_units» в модели.	0.76	0.75	0.79	0.76
Т	Что такое нечёткое множество?	Множество с элементами, имеющими степень принадлежности.	0.89	0.85	0.92	0.88
У	Чем нечёткая логика отличается от классической?	Использует степени истинности, а не только 0 или 1.	0.81	0.77	0.86	0.80
П	Как реализовать нечёткое правило на практике?	Определить правила, задать функции принадлежности, использовать движок вывода.	0.72	0.74	0.78	0.74
Т	Что делает нейронная сеть при классификации?	Определяет класс объекта на основе входных данных.	0.87	0.84	0.90	0.86

Среднее значение показателя адекватности составило $\bar{K} = 0.81$, что превышает установленный порог релевантности $\tau = 0.75$, определённый ранее. Это подтверждает способность модели формировать достаточно корректные и информативные ответы для типовых запросов студентов.

Анализ распределения значений показывает, что наиболее высокие показатели K наблюдаются для теоретических вопросов (в среднем 0.88), поскольку они ближе к данным, на которых обучалась модель, и хорошо связаны с узлами графа знаний [4]. Уточняющие вопросы показали средние значения (около 0.80), что объясняется сложностью интерпретации сравнительных конструкций («чем отличается…», «в чём разница…»). Прикладные вопросы имеют наименьшие значения (в среднем 0.75–0.76), что связано с отсутствием в графе знаний конкретных сценариев реализации и более свободной формой ответов.

Выводы

В ходе проведённого исследования была выполнена экспериментальная проверка применимости интегрального критерия адекватности K для оценки качества ответов интеллектуального почтового ассистента.

Результаты показали, что значение показателя K для большинства запросов превышает установленный порог релевантности $\tau = 0.75$, что подтверждает способность системы формировать осмысленные, связные и тематически корректные ответы. Наибольшие значения адекватности наблюдаются при обработке теоретических запросов, где модель опирается на чётко определённые понятия графа знаний. Для прикладных вопросов значения ниже, что указывает на необходимость расширения базы знаний практическими примерами и сценариями использования [5].

Высокая корреляция ($\tau = 0.87$) между автоматической метрикой K и экспертными оценками демонстрирует, что предложенный критерий может служить объективным инструментом для количественной оценки качества генерации ответов.

Таким образом, исследование подтверждает эффективность гибридного подхода, объединяющего методы NLP, граф знаний и рекуррентные нейросети для формирования информативных ответов студентам через электронную почту.

Литература

1. Зозуля Н.М., Федяев О.И. Рекуррентные нейронные сети как средство моделирования естественного языка // Программная инженерия: методы и технологии разработки информационно-вычислительных систем (ПИИВС–2024): сб. материалов V Междунар. науч.-практ. конф.; Т.2, 27–28 ноября 2024. – Донецк: ДонНТУ, 2024. – С. 229–235.

2. Галаган К.В. Применение нейросетевых моделей NLP для оценки эффективности мероприятий по формированию комфортной городской среды // Вестник ВГУ. – Воронеж: ВГУ, 2023. – № 3 – С. 167–178.

3. Прошина М.В. Современные методы обработки естественного языка: нейронные сети // Экономика строительства. – Москва, 2022. – № 5. – С. 27–42.

4. Грибков Е.И., Ехлаков Ю.П. Нейросетевая модель для обработки запросов пользователей на этапе эксплуатации и сопровождения программного продукта // Бизнес-информатика. – Москва, 2020. – № 1, Т. 14. – С. 7–18.

5. Волкова И.А., Шамаева Е.Д. Экскурс в графы знаний // International Journal of Open Information Technologies. – Москва, 2023. – № 3, Т. 11. – С. 75–83.

Zozulya N.M., Fedyaev O.I. Experimental assessment of the quality of the generation of answers by an intelligent mail assistant based on the adequacy criterion. *An experimental study of the work of an intelligent mail assistant for students based on a neural network model and a knowledge graph has been conducted. Based on the previously proposed adequacy criterion, a quantitative assessment of the relevance and coherence of the answers to questions on the discipline "Neural Network and Fuzzy Systems" has been performed. The results of calculating the integral indicator for different types of questions are presented, and the interpretation of the obtained data is given.*

Keywords: *intelligent assistant, natural language processing, adequacy criterion, neural network models, educational technologies.*

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В ОБРАЗОВАНИИ И НАУЧНЫХ ИССЛЕДОВАНИЯХ (СИТОНИ-2025)

(СТУДЕНЧЕСКАЯ СЕКЦИЯ)

Сборник материалов
IX Всероссийской научно-технической конференции
с международным участием

20 ноября 2025 г.

Редактор: В. Н. Павлыш
Дизайн и верстка: К. Н. Ефименко, Е. В. Прокопенко

Адрес редакции:
283015, Донецкая Народная Республика,
г.о. Донецк, г. Донецк, ул. Артема, д. 131,
ФГБОУ ВО «Донецкий национальный технический университет»,
факультет интеллектуальных систем и программирования,
каф. «Прикладная математика и искусственный интеллект», ауд. 11.409, 11.516
тел.: +7(856)-304-03-08
e-mail: amai_donntu@mail.ru