BIG DATA
and Advanced Analytics



XII Международная научно-практическая конференция

В двух частях

Часть 2

Министерство образования Республики Беларусь
Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»
Национальная академия наук Беларуси
Объединенный институт проблем информатики
МИРЭА – Российский технологический университет
ФГАОУ ВО «Северо-Кавказский федеральный университет»
Евразийский национальный университет имени Л. Н. Гумилева

BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ

BIG DATA AND ADVANCED ANALYTICS

СБОРНИК НАУЧНЫХ СТАТЕЙ
XII МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

(Республика Беларусь, Минск, 23 апреля 2026 года)

В двух частях

Часть 2

Минск БГУИР 2026

УДК 004.6
ББК 32.973
Б59

Редакционная коллегия:

В.А. Богуш, доктор физико-математических наук, профессор, ректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», Республика Беларусь;

С.К. Дик, кандидат физико-математических наук, доцент, депутат Палаты представителей Национального собрания Республики Беларусь седьмого созыва;

Т.В. Казак, доктор психологических наук, профессор, заведующий кафедрой инженерной психологии и эргономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», Республика Беларусь

Рецензенты:

С.Н. Касанин, заместитель генерального директора по научной работе государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси»;

Д.В. Лихачевский, кандидат технических наук, доцент, заместитель начальника государственного учреждения «Секретариат Наблюдательного совета Парка высоких технологий»

Б59 **BIG DATA** и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. XII Междунар. науч.-практ. конф. (Респ. Беларусь, Минск, 23 апр. 2026 г.) : в 2 ч. Ч. 2 / редкол.: В.А. Богуш [и др.]. – Минск : БГУИР, 2026. – 354 с.

ISBN 978-985-543-878-7 (ч. 2).

В сборнике опубликованы результаты научных исследований и разработок в области BIG DATA and Advanced Analytics для оптимизации IT- и бизнес-решений, а также тематических исследований в области медицины, образования и экологии.

УДК 004.6
ББК 32.973

ISBN 978-985-543-878-7 (ч. 2)
ISBN 978-985-543-876-3

© УО «Белорусский государственный университет информатики и радиоэлектроники», 2026

ОРГАНИЗАЦИОННЫЙ КОМИТЕТ КОНФЕРЕНЦИИ



Председатель
Богуш В.А., ректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р физ.-мат. наук, профессор

ЧЛЕНЫ ОРГАНИЗАЦИОННОГО КОМИТЕТА



Араш Косари, доцент кафедры электротехники и информационных технологий «Иранская исследовательская организация по науке и технологиям», канд. техн. наук, доцент, Исламская Республика Иран



Батура М.П., научный руководитель НИЛ 6.3 учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р техн. наук, профессор, академик Международной академии наук высшей школы, заслуженный работник образования Республики Беларусь



Давыдов М.В., первый проректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент



Дик С.К., заведующий кафедрой инженерной и компьютерной графики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. физ.-мат. наук, доцент



Исмаилова А.А., PhD, ассоциированный профессор, и.о. профессора ГОП «Информационные системы», НАО «Казахский агротехнический исследовательский университет имени С.Сейфулина», Республика Казахстан



Казак Т.В., заведующий кафедрой инженерной психологии и эргономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», д-р психол. наук, профессор



Касанин С.Н., заместитель генерального директора по научной работе Государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси»



Кругликов С.В., главный научный сотрудник Государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси»



Кудж С.А., ректор федерального государственного бюджетного образовательного учреждения высшего образования «МИРЭА – Российский технологический университет», д-р техн. наук



Лихачевский Д.В., заместитель начальника государственного учреждения «Секретариат Наблюдательного совета Парка высоких технологий», канд. техн. наук, доцент



Мезилов Г.А., ректор «Инженерно-технологического университета Туркменистана имени Огуз-хана», д-р техн. наук



Пархименко В.А., заведующий кафедрой экономики учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. экон. наук, доцент



Петренко В.И., заведующий кафедрой организации и технологии защиты информации ФГАО ВО «Северо-Кавказский федеральный университет», канд. техн. наук, доцент



Пискун Г.А., заместитель декана факультета компьютерного проектирования учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент



Рябова А.Н., начальник Государственного учреждения «Секретариат Наблюдательного совета Парка высоких технологий»



Стемпицкий Р.А., проректор по научной работе учреждения образования «Белорусский государственный университет информатики радиоэлектроники», канд. техн. наук, доцент



Тебуева Т.Б., заведующая кафедрой прикладной математики и компьютерной безопасности ФГАОУ ВО «Северо-Кавказский федеральный университет (г. Ставрополь), д-р физ.-мат. наук, доцент, Российская Федерация



Тузиков А.В., заведующий лабораторией математической кибернетики Государственного научного учреждения «Объединенный институт проблем информатики Национальной академии наук Беларуси», д-р физ.-мат. наук, профессор



Тусупов Д.А., заведующий кафедрой информационных систем Евразийского национального университета им. Л.Н. Гумилева, д-р физ.-мат. наук, профессор, Республика Казахстан



Ульянов Н.И., декан факультета компьютерных систем и сетей учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент



Целых А.А., заместитель директора Института компьютерных технологий и информационной безопасности Южного федерального университета, канд. техн. наук, доцент, Российская Федерация



Шнейдеров Е.Н., проректор учреждения образования «Белорусский государственный университет информатики и радиоэлектроники», канд. техн. наук, доцент

ОРГАНИЗАТОРЫ КОНФЕРЕНЦИИ

Министерство образования Республики Беларусь

Учреждение образования «Белорусский государственный университет информатики и радиоэлектроники»

Национальная академия наук Беларуси

Государственное научное учреждение «Объединенный институт проблем информатики Национальной академии наук Беларуси»

МИРЭА – Российский технологический университет, Российская Федерация

Университет инженерных технологий Туркменистана имени Огуз-хана, Туркменистан

ФГАОУ ВО «Северо-Кавказский федеральный университет», Российская Федерация

Евразийский национальный университет имени Л. Н. Гумилева, Республика Казахстан

СПОНСОР КОНФЕРЕНЦИИ

Министерство образования Республики Беларусь

Резидент Парка высоких технологий КБК ТехЛаб
Команда опытных разработчиков и инженеров, которые создают современные технологические решения для бизнеса. Помогают компаниям трансформировать идеи в эффективные цифровые продукты.

Резидент Парка высоких технологий ISsoft
Компания ISsoft – один из крупнейших белорусских разработчиков IT решений для рынков США и Западной Европы. Основана в Минске в 2004 г. как дочерняя компания корпорации Coherent Solutions, Inc. (США). Резидент Парка Высоких Технологий с 2007 года.

ОГЛАВЛЕНИЕ

Э.М. Аксень Подходы к моделированию оптимального распределения ресурсов по регионам 12

Л.И. Архипова, Л.Ф. Медведева Методологический подход к бесшовной интеграции генеративного искусственного интеллекта с проектами цифровой трансформации бизнеса 22

А.В. Беглов, Л.С. Хорошко Анализ и рефакторинг библиотеки обменно-корреляционных функционалов пакета OPENMX 31

А.А. Исмаилова, Р.Н. Тажиева Архитектура потоковой обработки и хранения д анных биогазовых установок на основе технологий Big Data 38

К.А. Афанасенко, Н.А. Жиляк Интеллектуальное прогнозирование загрузки облачной инфраструктуры с использованием модели CHRONOS-2 и методов анализа временных рядов 45

А.И. Лёвина, А.С. Дубгорн Методика выявления структурных недостатков логистических сетей на основе анализа данных с использованием имитационного моделирования 51

Е.И. Бавбель Применение методов big data в разработке адаптивных систем управления элеронами летательных аппаратов 62

Е.А. Алуев, М.А. Булычева семантическое слияние данных на основе графов знаний: интеграция структурированных и неструктурированных данных для раскрытия контекстного интеллекта 70

А.А. Карпук, Л.С. Лазута Модификация алгоритмов нормализации таблиц баз данных для применения в системах Big Data 80

А.А. Радкевич Принципы и алгоритмы построения нейросетевых систем управления координатами нижнего уровня в мобильной робототехнике 85

А.В. Бакулев, М.А. Бакулева Разработка информационной системы автоматизированного сбора сведений о посещаемости учащихся на основе алгоритмов распознавания и идентификации 95

Д.В. Лихачевский, С.А. Мигалевич, В.О. Агапова Интеллектуальные алгоритмы координации и управления устройствами «Умного дома» 100

Н.А. Шиманский Методы машинного обучения с использованием данных OPENMX и AFLOW для оптимизации вычислительного эксперимента 105

Г.А. Пискун, В.Ф. Алексеев, Д.Г. Ершов Разработка программного комплекса для инвентаризации и мониторинга состояния основных средств организации 110

Г.А. Пискун, Д.Г. Ершов Система учета основных средств организации на основе технологий data matrix и платформ Big Data 115

М.С. Печёнов, В.М. Бондарик Модель оценки пропускной способности и безопасной нагрузки лабораторий жидкостной цитологии при внедрении ИИ-ассистентов 123

N.L. Kisialeu, A.M. Nedzved Automated cardiac rhythm classification from 12-lead ECG using attention-enhanced deep neural networks 128

С.А. Мигалевич Экспертные и гибридные модели принятия решений в интеллектуальных системах контроля и оценки знаний 137

Н.Л. Мирончик Измерение системного финансового стресса и анализ его связи с разрывом выпуска и инфляцией: var-подход на данных Беларуси 142

В.И. Борисенко Гибридная модель оптимизации бизнес-решений на основе больших языковых моделей 151

Е.А. Курлюк, Н.А. Ларченко, А.В. Бойдич, М.В. Давыдов Исследование влияния ориентации документа на точность извлечения данных мультимодальной моделью PADDLEOCR-VL. 158

Т.А. Васяева, М.В. Душа Разработка нейросетевого консультанта для студентов на базе больших языковых моделей 166

К.С. Крез, М.А. Кривоносова, А.Р. Шипуль, А.С. Гугалев Transformer-архитектура: как изменила подход к разработке систем выявления текстовых заимствований 174

К.С. Крез, Е.Н. Шнейдеров, В.И. Голушко Алгоритмы сравнения текстовой информации 181

А.Ю. Яцкевич Применение системы электронного обучения для изучения психологических характеристик студентов (на примере склонности к риску) 188

С.Т. Веллиева, Н.М. Реджепов, А. Гурбанов Архитектура познания: трансформация образовательных систем в эпоху интеллектуальной аналитики и больших данных 192

В.А. Вишняков Цифровые двойники на нейронных сетях в электронной терапии пациентов с неврологическими заболеваниями 195

Д.А. Молдасейтов, Б.Е. Таныкпаева Разработка интеллектуального мобильного приложения для изучения иностранных языков с использованием нейросетевых API 199

С.С. Шайымов, А.Б. Чарыев, Б.Н. Атанепесов Big Data и продвинутая аналитика переписывают правила делового мира 207

Н.С. Глушаченко, Д.В. Деменковец Способы аугментации данных при реализации системы поддержки принятия решений для диагностики нарушений равновесия и работы голеностопа 210

Т.А. Васяева, П.Г. Мельник Повышение эффективности метода потенциальных полей в групповом управлении за счёт полигонального заполнения и распределённого поиска пути 215

Т.А. Васяева, М.А. Рябко Исследование методов компьютерной обработки медицинских изображений 223

К.А. Кошлатый, И.В. Лихтар, А.Ю. Ефремова Исследование возможностей сервиса Yandex Load Testing для оценки производительности веб-сервисов 232

O.P. Berdiyeva Application of big data and advanced analytics for optimization of IT solutions 239

В.О. Тихоненко, Е.Д. Ракицкий, Ю.А. Ёщик, А.В. Воробей Концепция цифровых двойников и их применение для моделирования климата 241

А.И. Селезнев Элементы обработки данных в системах управления беспилотным автомобилем 249

А.А. Гайсенок, Г.А. Хацкевич Применение технологий big data в современном рекрутинге 253

М.М. Тарасов, Л.Ю. Костылева Проектирование и разработка чат-бота для генерации рекомендаций по инклюзивному образованию с применением RAG-технологий 256

Е.И. Бавбель, В.Ф. Алексеев Data-driven подход к оценке напряжённо-деформированного состояния силового каркаса БПЛА из полимерных композитов 264

Л.Д. Черемисинова Оптимизация двухуровневых представлений логических схем с учетом энергопотребления 271

Г.Г. Григорьева, Н.С. Давыдова, В.Е. Васюк, М.В. Давыдов, К.И. Татарко Анализ биомеханических признаков движений для прогнозирования двигательной одаренности детей 279

А.А. Навроцкий, Л.Ю. Шилин, Д.Д. Хмыз, К.Д. Сморщик Модели и методы глубокого обучения для детектирования объектов в интеллектуальных системах обработки визуальных данных 285

Е.Р. Адамовский, Р.П. Богуш Формирование обучающей выборки для нейронных сетей при пространственно-временном обнаружении дыма на видео 289

П.А. Шлык Применение машинного обучения для распознавания активности и улучшения качества позиционирования в помещении 298

В.Г. Евдокимов, А.А. Навроцкий Динамические цифровые двойники как инструмент снижения социальных рисков больших данных в условиях Индустрии 4.0 304

Е.В. Степанец Роль информационных технологий в современном научном познании 310

Y. Liashchevich, Cai Yuxin Classic and modern encryption algorithms for big data processing 315

У.А. Пателина, О.Н. Шкор Использование технологий Big Data и концепции BYOD для оценки влияния мобильных устройств на производительность труда в организациях 320

А.С. Довгий, К.В. Василевский Графовая нейронная сеть на основе адаптивного уравнения реакции-адвекции-диффузии для задачи классификации молекулярных графов 324

Н.А. Ларченко, Е.А. Курлюк, А.Ч. Ганиев, М.В. Давыдов Разработка и экспериментальная оценка модели учёта физиологического состояния пользователя на основе ЭКГ с использованием методов Big Data и продвинутой аналитики в HCI-системах 335

Е.М. Косарева, Д.В. Лихачевский, Д.О. Кожевников Мультимодальная система биометрической идентификации на основе комбинирования физиологических и поведенческих признаков 346

УДК 330.46

ПОДХОДЫ К МОДЕЛИРОВАНИЮ ОПТИМАЛЬНОГО РАСПРЕДЕЛЕНИЯ РЕСУРСОВ ПО РЕГИОНАМ



Э.М. Аксень
Заведующий кафедрой экономики и управления,
доктор экономических наук, профессор
УО «Белорусский государственный экономический университет»,
eaksen@bseu.by

Э.М. Аксень

Окончил факультет прикладной математики Белорусского государственного университета. Область научных интересов связана с разработкой математических и инструментальных методов в экономике.

Аннотация. В статье представлены подходы к моделированию оптимальной динамики распределения ресурсов по регионам для социально-экономической системы в дискретном и непрерывном времени. Методика предполагает построение межвременного интегрального социально-экономического показателя и максимизацию его значения с учетом ограничений на объясняющие факторы. При этом учитывается также запаздывание влияния объясняющих факторов на результирующие показатели.

Ключевые слова: динамика, интегральный показатель, оптимизация, моделирование, социально-экономическая система, регион, эластичность, весовой коэффициент.

Введение. Общеизвестно, что выделяемые на развитие регионов ресурсы на динамику социально-экономических показателей, таких как рождаемость, количество безработных, обеспеченность жильем и др. При этом указанное влияние имеет место с запаздыванием, что также следует учитывать при моделировании. В данной статье предложены подходы к построению экономико-математических моделей в дискретном и непрерывном времени, направленных на определение динамики оптимального распределения ресурсов между регионами с учетом того, что выделяемые ресурсы влияют на социально-экономические показатели с некоторым запаздыванием.

Подходы к моделированию влияния объясняющих факторов на социально-экономические показатели. Пусть n – число регионов в рассматриваемой социально-экономической системе. В частности, в случае, когда в качестве социально-экономической системы выступает Республика Беларусь, а регионы – это области Республики Беларусь, а также город Минск, $n = 7$. Через m обозначим число фигурирующих в модели ресурсов. Для обозначения значения j-го ресурса в i-м регионе для t-го периода времени ($i = \overline{1,n}$, $j = \overline{1,m}$) будем использовать $x_{ij}(t)$.

Пусть s – число социально-экономических показателей. Значение k-го социально-экономического показателя в i-м регионе для t-го периода времени ($i = \overline{1,n}$, $k = \overline{1,s}$) обозначим через $y_{ik}(t)$. В соответствии с предлагаемой нами методикой имеет место зависимость (с запаздыванием) показателей $y_{ik}(t)$ от показателей $x_{ij}(t)$.

Для того, чтобы учесть запаздывание влияния объясняющих факторов $x_{ij}(t)$ на результирующие показатели $y_{ik}(t)$, определим ненаблюдаемые показатели $\tilde{x}_{ijk}(t)$ по следующей рекуррентной формуле:

$$\tilde{x}_{ijk}(t) = \tilde{x}_{ijk}(t-1) \cdot \left[\frac{x_{ij}(t)}{\tilde{x}_{ijk}(t-1)} \right]^{\gamma_{jk}}, i = \overline{1,n}, \quad j = \overline{1,m}, \ k = \overline{1,s}, \tag{1}$$

где γ_{jk} – параметры, которые мы будем оценивать с помощью фактических данных.

Прологарифмируем формулу (1):

$$\ln\left[\tilde{x}_{ijk}(t)\right] = \ln\left[\tilde{x}_{ijk}(t-1)\right] + \gamma_{jk} \cdot \left\{ \ln\left[x_{ij}(t)\right] - \ln\left[\tilde{x}_{ijk}(t-1)\right] \right\}, \quad i = \overline{1,n}, \quad j = \overline{1,m}, \ k = \overline{1,s}. \tag{2}$$

Замечание 1. Формула (2) – это формула для получения прогнозных значений методом экспоненциального сглаживания (для прологарифмированных данных).

Непрерывновременной аналог соотношения (2) имеет вид:

$$\frac{d}{dt}\ln\left[\tilde{x}_{ijk}(t)\right] = \gamma_{jk} \cdot \left\{ \ln\left[x_{ij}(t)\right] - \ln\left[\tilde{x}_{ijk}(t)\right] \right\}, \quad i = \overline{1,n}, \quad j = \overline{1,m}, \ k = \overline{1,s}. \tag{3}$$

Замечание 2. В данной статье при описании подходов моделирования в непрерывном времени буква *t* обозначает момент (а не период) времени. Соответственно, в случаях, когда показатель x_{ij} (либо $y_{ik}(t)$) – это поток (а не запас), то $x_{ij}(t)$ – это (при непрерывновременном моделировании) интенсивность соответствующего потока. Например, если при дискретновременном моделировании $x_{ij}(t)$ – это выпуск продукции за период *t*, то при непрерывновременном моделировании $x_{ij}(t)$ – это интенсивность (т.е. скорость) выпуска в момент времени *t*.

Отметим, что формулы (1)–(3) дают возможность найти траектории $\tilde{x}_{ijk}(t)$ и $\ln\left[\tilde{x}_{ijk}(t)\right]$ с помощью заданного значения $\tilde{x}_{ijk}(t_0)$ в начальный момент времени t_0, а именно можно показать, что

$$\tilde{x}_{ijk}(t) = \left[\tilde{x}_{ijk}(t_0)\right]^{\left(1-\gamma_{ik}\right)^{t-t_0}} \prod_{\tau=t_0+1}^{t} \left[x_{ij}(\tau)\right]^{\gamma_{jk}\left(1-\gamma_{jk}\right)^{t-\tau}}, \tag{4}$$

$$\ln\left[\tilde{x}_{ijk}(t)\right] = \gamma_{jk} \cdot \sum_{\tau=t_0+1}^{t} \left(1-\gamma_{jk}\right)^{t-\tau} \ln\left[x_{ij}(\tau)\right] + \left(1-\gamma_{jk}\right)^{t-t_0} \ln\left[\tilde{x}_{ijk}(t_0)\right], \tag{5}$$

$$\ln\left[\tilde{x}_{ijk}(t)\right] = \ln\left[\tilde{x}_{ijk}(t_0)\right] e^{-\gamma_{jk}\cdot(t-t_0)} + \gamma_{jk} \int_{t_0}^{t} \ln\left[x_{ij}(\tau)\right] e^{-\gamma_{jk}\cdot(t-\tau)} d\tau, \tag{6}$$

$i = \overline{1,n}, \quad j = \overline{1,m}, \ k = \overline{1,s}, \ t \geq t_0.$

Через $\hat{y}_{ik}(t)$ обозначим прогнозное значение *k*-го социально-экономического показателя в *i*-м регионе для *t*-го периода времени ($i = \overline{1,n}, \ k = \overline{1,s}$). Будем использовать следующую формулу для нахождения прогнозных значений $\hat{y}_{ik}(t)$:

$$\hat{y}_{ik}(t) = a_{ik} \prod_{j=1}^{m} \tilde{x}_{ijk}^{b_{jk}}(t), \qquad i = \overline{1,n}, \qquad k = \overline{1,s}, \tag{7}$$

где a_{ik} и b_{jk} – параметры, которые мы будем оценивать с помощью фактических данных.

Прологарифмировав формулу (7), получим:

$$\ln\left[\hat{y}_{ik}(t)\right] = \ln a_{ik} + \sum_{j=1}^{m} b_{jk} \cdot \ln\left[\tilde{x}_{ijk}(t)\right], \qquad i = \overline{1,n}, \qquad k = \overline{1,s}. \tag{8}$$

Замечание 3. Формула (8) – это формула для нахождения прогнозных значений в рамках следующей панельной модели с фиксированным эффектом [1, с. 362–367] для прологарифмированных данных:

$$\ln\left[y_{ik}(t)\right] = \ln \alpha_{ik} + \sum_{j=1}^{m} \beta_{jk} \cdot \ln\left[\tilde{x}_{ijk}(t)\right] + \varepsilon_{ik}(t), \qquad i = \overline{1,n}, \qquad k = \overline{1,s}, \tag{9}$$

где α_{ik}, β_{jk} – теоретические аналоги коэффициентов a_{ik} и b_{jk}, $\varepsilon_{ik}(t)$ – случайные отклонения с нулевым математическим ожиданием.

Через T_0 обозначим последний период времени, для которого известны значения всех показателей x_{ij} и y_{ik}. В соответствии с формулами (2) и (8) значения параметров γ_{ijk}, a_{ik} и b_{jk} ($i = \overline{1,n}$, $j = \overline{1,m}$, $k = \overline{1,s}$) будем находить методом наименьших квадратов с помощью решения следующих оптимизационных задач:

$$\sum_{t=t_0}^{T_0} \sum_{i=1}^{n} \left[\ln y_{ik}(t) - \ln \hat{y}_{ik}(t)\right]^2 \to \min, \tag{10}$$

$$\ln\left[\tilde{x}_{ijk}(t)\right] = \ln\left[\tilde{x}_{ijk}(t-1)\right] + \gamma_{jk} \cdot \left\{\ln\left[x_{ij}(t)\right] - \ln\left[\tilde{x}_{ijk}(t-1)\right]\right\}, \qquad t = \overline{t_0+1,T}, \tag{11}$$

$$\tilde{x}_{ijk}(t_0) = x_{ij}(t_0), \qquad i = \overline{1,n}, \qquad j = \overline{1,m}, \qquad k = \overline{1,s}, \tag{12}$$

$$\ln\left[\hat{y}_{ik}(t)\right] = \ln a_{ik} + \sum_{j=1}^{m} b_{jk} \cdot \ln\left[\tilde{x}_{ijk}(t)\right], \; i = \overline{1,n}, \qquad k = \overline{1,s}, \quad t = \overline{t_0,T}. \tag{13}$$

При этом для каждого социально-экономического показателя *y* решается своя задача (10)–(13). В этих задачах $x_{ij}(t)$ и $y_{ik}(t)$, $t = \overline{t_0,T_0}$, – известные (фактические) значения, $\ln a_{ik}$, b_{jk}, γ_{jk} – переменные.

При моделировании в непрерывном времени для оценки параметров решаются задачи, аналогичные приведенной ваше, а именно задачи следующего вида:

$$\sum_{l=0}^{L} \sum_{i=1}^{n} \left[\ln y_{ik}(t_l) - \ln \hat{y}_{ik}(t_l)\right]^2 \to \min, \tag{14}$$

$$\ln\left[\tilde{x}_{ijk}\left(t_l\right)\right]=\ln\left[\tilde{x}_{ijk}\left(t_0\right)\right]e^{-\gamma_{jk}\cdot\left(t_l-t_0\right)}+\frac{\gamma_{jk}\Delta t}{2}\sum_{q=1}^{l}\left\{\ln\left[x_{ij}\left(t_{q-1}\right)\right]e^{-\gamma_{jk}\cdot\left(t_l-t_{q-1}\right)}+\ln\left[x_{ij}\left(t_q\right)\right]e^{-\gamma_{jk}\cdot\left(t_l-t_q\right)}\right\}, \tag{15}$$

$$\tilde{x}_{ijk}(t_0)=x_{ij}(t_0), \qquad i=\overline{1,n} \quad , \qquad j=\overline{1,m}, \quad k=\overline{1,s}, \tag{16}$$

$$\ln\left[\hat{y}_{ik}(t)\right]=\ln a_{ik}+\sum_{j=1}^{m}b_{jk}\cdot\ln\left[\tilde{x}_{ijk}(t)\right],\ i=\overline{1,n} \quad , \qquad k=\overline{1,s}, \quad l=\overline{0,L} \ . \tag{17}$$

Здесь $\left(t_0,t_1,\ldots,t_L\right)$ – разбиение временного интервала $\left[t_0,T_0\right]$ (при этом считается, что фактические значения факторов x_{ij} и показателей y_{ik} известны в моменты времени t_l , $l=\overline{0,L}$), через Δt обозначена длина подынтервала разбиения (т.е. $\Delta t:=t_l-t_{l-1}$), а соотношения (15) – линейные аппроксимации равенств (6) с использованием метода трапеций (для численного интегрирования).

Подходы к построению межвременного интегрального социально-экономического показателя. Введем интегральный показатель в периоде *t* по формуле:

$$Y\left(t\right)=\prod_{i=1}^{n}\prod_{k=1}^{s}\left[y_{ik}(t)\right]^{\alpha_{ik}(t)}, \tag{18}$$

где $\alpha_{ik}(t)$ – весовой коэффициент для *k*-го показателя в *i*-м регионе в периоде *t*, причем

$$\sum_{i=1}^{n}\sum_{k=1}^{s}\alpha_{ik}(t)=1, \tag{19}$$

Замечание 4. Интегральный показатель (18) представляет собой средневзвешенное геометрическое значение социально-экономических показателей $y_{ik}(t)$ в периоде *t* (по регионам и по видам показателей *y*).

Замечание 5. Весовые коэффициенты $\alpha_{ik}(t)$ могут быть оценены экспертным образом, в частности, на основе экспертной информации об эластичностях замещения (*k*-го социально-экономического показателя в *i*-м регионе $y_{ik}(t)$ на *q*-й социально-экономический показатель в *l*-м регионе $y_{lq}(t)$) при постоянном значении интегрального социально-экономического показателя $Y\left(t\right)$(см. [2]).

Введем межвременной интегральный показатель для временного интервала $\left[T_1,T_2\right]$ (состоящего из периодов, начиная с периода T_1по период T_2) по формуле:

$$Y\left(T_1,T_2\right):=\prod_{t=T_1}^{T_2}\left[Y(t)\right]^{\varphi(t)}, \tag{20}$$

где $\varphi(t)$ – весовые коэффициенты для периодов *t*, причем

$$\sum_{t=T_1}^{T_2}\varphi(t)=1. \tag{21}$$

С помощью формулы (20) можно показать [2], что отношение $\frac{\varphi(t)}{\varphi(\tau)}$ равно точечной эластичности $e(t,\tau)$ замещения $Y(t)$ на $Y(\tau)$, т.е. оно показывает на сколько процентов

должно увеличиться значение интегрального социально-экономического показателя $Y(\tau)$ в периоде τ при уменьшении на 1% указанного показателя $Y(t)$ в периоде t так, чтобы значение межвременного интегрального социально-экономического показателя $Y\left(T_1, T_2\right)$ не изменилось.

Обозначим «краткосрочные» эластичности $e(\tau+1,\tau)$ через $\delta(\tau)$:

$$\delta(\tau) := e(\tau+1,\tau), \qquad \tau = \overline{T_1, T_2 - 1}. \tag{22}$$

В частном случае, когда «краткосрочные» эластичности $\delta(\tau)$ постоянны (т.е. не зависят от τ), можно показать [2], что

$$\varphi(t) = \begin{cases} \delta^{t-T_1} \dfrac{1-\delta}{1-\delta^{T_2-T_1+1}} & \text{при } \delta \neq 1, \\ \dfrac{1}{T_2 - T_1 + 1} & \text{при } \delta = 1, \end{cases} \tag{23}$$

$t = \overline{T_1, T_2}$.

Формула (23) может быть использована для вычисления коэффициентов $\varphi(t)$. При этом «краткосрочная» эластичность δ может быть оценена экспертным образом.

Обобщим формулу (20) для моделирования в непрерывном времени. Разобьем временной интервал $\left[T_1, T_2\right]$ на подынтервалы $\left[t_{l-1}, t_l\right]$, $l = \overline{1, L}$ (где $t_0 = T_1$, $t_L = T_2$). Обозначим разбиение $\left(t_0, t_1, \ldots, t_L\right)$ отрезка $\left[T_1, T_2\right]$ через Θ, а диаметр разбиения Θ через $\mu\left(\Theta\right)$. (Напомним, что диаметр разбиения – это максимальная длина подынтервалов $\left[t_{l-1}, t_l\right]$, $l = \overline{1, L}$, т.е. $\mu\left(\Theta\right) = \max\limits_{l=\overline{1,L}} \Delta t_l$, где $\Delta t_l := t_l - t_{l-1}$.)

Напомним также, что разметкой разбиения $\Theta := \left(t_0, t_1, \ldots, t_L\right)$ называется последовательность $\tilde{\Theta} := \left(\tilde{t}_1, \ldots, \tilde{t}_L\right)$ точек $\tilde{t}_l \in \left[t_{l-1}, t_l\right]$, $l = \overline{1, L}$. Размеченное разбиение – это разбиение вместе с разметкой, т.е. это пара $\hat{\Theta} := \left\langle \Theta, \tilde{\Theta} \right\rangle$, а диаметр $\mu\left(\hat{\Theta}\right)$ размеченного разбиения $\hat{\Theta}$ – это диаметр $\mu\left(\Theta\right)$ соответствующего неразмеченного разбиения Θ.

Пусть $\varphi\left(t\right)$ – весовые коэффициенты для моментов времени t, причем (по аналогии с равенством (21) для дискретновременного случая)

$$\int_{T_1}^{T_2} \varphi\left(t\right) dt = 1. \tag{24}$$

По аналгии с формулой (20) введем межвременной интегральный показатель для временного интервала $\left[T_1, T_2\right]$ по формуле:

$$Y\left(T_1, T_2\right) := \lim_{\mu(\Theta)\to 0} \prod_{l=1}^{L} \left[Y(\tilde{t}_l)\right]^{\varphi(\tilde{t}_l)\Delta t_l}, \tag{25}$$

где $\Delta t_l := t_l - t_{l-1}$, $l = \overline{1, L}$.

Замечание 6. В соответствии с формулой (25) $Y\left(T_1, T_2\right)$ – это значение, такое что для любого положительного числа ε найдется такое положительное число δ, что для

любого размеченного разбиения $\hat{\Theta}$ отрезка $\left[T_1, T_2\right]$ с диаметром $\mu\left(\hat{\Theta}\right) \le \delta$ имеет место неравенство:

$$\left| \prod_{l=1}^{L} \left[Y(\tilde{t}_l) \right]^{\varphi(\tilde{t}_l)\Delta t_l} - Y\left(T_1, T_2\right) \right| \le \varepsilon. \tag{26}$$

Заметим, что

$$\prod_{l=1}^{L} \left[Y(\tilde{t}_l) \right]^{\varphi(\tilde{t}_l)\Delta t_l} = \exp\left(\sum_{l=1}^{L} \varphi\left(\tilde{t}_l\right) \ln\left[Y(\tilde{t}_l) \right] \Delta t_l \right). \tag{27}$$

Следовательно,

$$\lim_{\mu(\Theta)\to 0} \prod_{l=1}^{L} \left[Y(\tilde{t}_l) \right]^{\varphi(\tilde{t}_l)\Delta t_l} = \exp\left(\int_{T_1}^{T_2} \varphi(t) \ln\left[Y(t) \right] dt \right). \tag{28}$$

Из равенств (25) и (28) получим:

$$Y\left(T_1, T_2\right) = \exp\left(\int_{T_1}^{T_2} \varphi(t) \ln\left[Y(t) \right] dt \right). \tag{25}$$

Определим эластичность замещения интегрального показателя $Y(t)$ в периоде $\left[t_1 - \Delta t, t_1 + \Delta t\right]$ этим же показателем в периоде $\left[t_2 - \Delta t, t_2 + \Delta t\right]$ при относительном изменении η_1 значения указанного показателя в периоде $\left[t_1 - \Delta t, t_1 + \Delta t\right]$ по формуле:

$$e\left(t_1, t_2; \eta_1, \Delta t\right) := -\frac{\eta_2}{\eta_1}, \tag{30}$$

где η_2 определяется из условия:

$$\begin{aligned} &\int_{t_1-\Delta t}^{t_1+\Delta t} \varphi(t) \ln\left[Y(t)\left(1+\eta_1\right) \right] dt + \int_{t_2-\Delta t}^{t_2+\Delta t} \varphi(t) \ln\left[Y(t)\left(1+\eta_2\right) \right] dt = \\ &= \int_{t_1-\Delta t}^{t_1+\Delta t} \varphi(t) \ln\left[Y(t) \right] dt + \int_{t_2-\Delta t}^{t_2+\Delta t} \varphi(t) \ln\left[Y(t) \right] dt. \end{aligned} \tag{31}$$

Эластичность замещения (30) показывает, во сколько раз относительное изменение η_2 значения показателя $Y(t)$ в периоде $\left[t_2 - \Delta t, t_2 + \Delta t\right]$ должно отличаться от относительного изменения η_1 этого же показателя в периоде $\left[t_1 - \Delta t, t_1 + \Delta t\right]$, так чтобы значение межвременного показателя $Y\left(T_1, T_2\right)$ (определенного по формуле (29)) осталось прежним.

Определим мгновенную эластичность замещения показателя $Y(t)$ в момент времени t_1 этим же показателем в момент времени t_2 при относительном изменении η_1 указанного показателя в момент времени t_1 как предел эластичности (30) при $\Delta t \to 0$:

$$e\left(t_1, t_2; \eta_1\right) := \lim_{\Delta t \to 0} e\left(t_1, t_2; \eta_1, \Delta t\right). \tag{32}$$

Из соотношений (30)–(32) следует, что

$$e(t_1,t_2;\eta_1) := -\frac{\eta_2}{\eta_1}, \tag{33}$$

где η_2 определяется из условия:

$$\varphi(t_1)\ln(1+\eta_1)+\varphi(t_2)\ln(1+\eta_2)=0. \tag{34}$$

Определим точечную мгновенную эластичность замещения (по отношению к межвременному показателю (29)) показателя $Y(t)$ в момент времени t_1 этим же показателем в момент времени t_2 как предел мгновенной эластичности (33) при $\eta_1 \to 0$:

$$e(t_1,t_2) := \lim_{\eta_1 \to 0} e(t_1,t_2;\eta_1). \tag{35}$$

Точечная мгновенная эластичность замещения (35) показывает на сколько процентов (приблизительно) должно увеличится значение показателя $Y(t)$ в «малом» периоде времени $[t_2-\Delta t, t_2+\Delta t]$ при уменьшении значения указанного показателя на 1% в «малом» периоде времени $[t_1-\Delta t, t_1+\Delta t]$, так чтобы значение межвременного показателя $Y(T_1,T_2)$ (определенного по формуле (29)) осталось прежним.

Предположим, что эластичность $e(t_1,t_2)$ зависит только от разности между моментами времени t_1 и t_2, и обозначим через δ эластичность $e(t+1,t)$:

$$\delta := e(t+1,t) \tag{36}$$

(которая в силу указанного предположения не зависит от выбора момента времени *t*).

С помощью формул (33)–(36) можно показать [3], что

$$\varphi(t)=\begin{cases} -\dfrac{\ln\delta}{1-\delta^{T_2-T_1}}\delta^{t-T_1} & \text{при} \quad \delta \neq 1, \\ \dfrac{1}{T_2-T_1} & \text{при} \quad \delta = 1, \end{cases} \tag{37}$$

$t \in [T_1,T_2]$.

Задача максимизации межвременного интегрального социально-экономического показателя. Предположим, что в *t*-м периоде в целом в социально-экономической системе суммарное взвешенное значение *j*-го ресурса (объясняющего фактора в более общем случае) не должно превышать планового значения $B_j(t)$, т.е.

$$\sum_{i=1}^{n} c_{ij}(t)x_{ij}(t) \le B_j(t), \qquad j=\overline{1,m},\ t=\overline{T_1,T_2}, \tag{38}$$

где $c_{ij}(t)$ – известные весовые коэффициенты (сумма которых может отличаться от единицы), $[T_1,T_2]$– период планирования.

Следовательно, в задаче максимизации межвременного интегрального показателя (20) для планового периода с T_1-го периода по T_2-й период должны учитываться условия (38), а также соотношения (4) и (7) (либо эквивалентные им соотношения (5), (8)).

Несложно заметить, что оптимизационная задача с прологарифмированной целевой функцией и прологарифмированными (некоторыми) ограничениями эквивалентна исходной задаче. Прологарифмировав формулу (20) (и используя соотношения (5), (8)),

будем иметь оптимизационную задачу, равносильную задаче максимизации прогнозного значения межвременного интегрального показателя (20):

$$\ln \hat{Y}\left(T_1, T_2\right) = \sum_{t=T_1}^{T_2} \varphi(t) \ln \hat{Y}(t) \to \max, \tag{39}$$

$$\ln \hat{Y}(t) = \sum_{i=1}^{n} \sum_{k=1}^{s} \alpha_{ik}(t) \ln\left[\hat{y}_{ik}(t)\right], \tag{40}$$

$$\ln\left[\hat{y}_{ik}(t)\right] = \ln a_{ik} + \sum_{j=1}^{m} b_{jk} \cdot \ln\left[\tilde{x}_{ijk}(t)\right]. \; i = \overline{1,n}, \qquad k = \overline{1,s}, \quad t = \overline{T_1, T_2}, \tag{41}$$

$$\ln\left[\tilde{x}_{ijk}(t)\right] = \gamma_{jk} \cdot \sum_{\tau=T_1}^{t} \left(1-\gamma_{jk}\right)^{t-\tau} \ln\left[x_{ij}(\tau)\right] + \left(1-\gamma_{jk}\right)^{t-t_0} \ln\left[\tilde{x}_{ijk}\left(T_0-1\right)\right], \tag{42}$$

$$i = \overline{1,n}, \quad j = \overline{1,m}, \; k = \overline{1,s}, \; t = \overline{T_1, T_2},$$

$$\sum_{i=1}^{n} c_{ij}(t) x_{ij}(t) \le B_j(t), \qquad j = \overline{1,m}, t = \overline{T_1, T_2}, \tag{43}$$

$$x_{ij}(t) \ge 0, \quad i = \overline{1,n} \quad , \quad j = \overline{1,m}, \quad t = \overline{T_1, T_2}. \tag{44}$$

Можно показать [2], что динамическую оптимизационную задачу (39)–(44) можно разбить на $T_2 - T_1 + 1$ более простых статических задач для каждого в отдельности периода $t = \overline{T_1, T_2}$:

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \beta_{ij}(t) \ln\left[x_{ij}(t)\right] \to \max, \tag{45}$$

$$\sum_{i=1}^{n} c_{ij}(t) x_{ij}(t) \le B_j(t), \qquad j = \overline{1,m}, \tag{46}$$

$$x_{ij}(t) \ge 0, \quad i = \overline{1,n} \quad , \quad j = \overline{1,m} \tag{47}$$

где

$$\beta_{ij}(t) := \sum_{k=1}^{s} b_{jk} \gamma_{jk} \sum_{\tau=t}^{T_2} \alpha_{ik}(\tau) \varphi(\tau) \left(1-\gamma_{jk}\right)^{\tau-t}. \tag{48}$$

С использованием метода множителей Лагранжа, получим формулу для оптимальных решений задач (41)–(44) (см. [2]):

$$x_{ij}(t) = w_{ij}(t) B_j(t), \qquad i = \overline{1,n}, \qquad j = \overline{1,m}, \; t = \overline{T_1 + 1, T_2} \tag{49}$$

где

$$w_{ij}(t) := \frac{\beta_{ij}(t)}{c_{ij}(t) \sum_{l=1}^{n} \beta_{lj}(t)}, \qquad i = \overline{1,n}, \qquad j = \overline{1,m} \qquad t = \overline{T_1 + 1, T_2} \tag{50}$$

Замечание 7. Из равенства (49) следует, что (в рамках данной модели) распределение долей ресурсов по регионам не зависит от объемов ресурсов в целом в социально-экономической системе.

В случае, когда весовые коэффициенты $\alpha_{ik}(t)$ и «краткосрочные» эластичности $\delta(t)$ постоянны (т.е. не зависят от периода времени *t*), можно показать [2], что

$$w_{ij}(t) = \frac{\tilde{\beta}_{ij}(t)}{c_{ij}(t)\sum_{l=1}^{n}\tilde{\beta}_{lj}(t)}, \qquad i=\overline{1,n}, \quad j=\overline{1,m} \quad t=\overline{T_1,T_2}. \tag{51}$$

где

$$\tilde{\beta}_{ij}(t) := \sum_{k=1}^{s} b_{jk}\alpha_{ik}\gamma_{jk}\frac{1-\left[\delta\left(1-\gamma_{jk}\right)\right]^{T_2-t+1}}{1-\delta\left(1-\gamma_{jk}\right)}, \quad i=\overline{1,n}, \qquad j=\overline{1,m},\ t=\overline{T_1,T_2}. \tag{52}$$

В силу формул (6), (7), (18), (20) непрерывновременной аналог оптимизационной задачи (39)–(44) имеет вид:

$$\ln\hat{Y}\left(T_1,T_2\right) = \int_{T_1}^{T_2}\varphi(t)\ln\hat{Y}(t)dt \to \max \tag{53}$$

$$\ln\hat{Y}(t) = \sum_{i=1}^{n}\sum_{k=1}^{s}\alpha_{ik}(t)\ln\left[\hat{y}_{ik}(t)\right], \quad t\in\left[T_1,T_2\right], \tag{54}$$

$$\ln\left[\hat{y}_{ik}(t)\right] = \ln a_{ik} + \sum_{j=1}^{m} b_{jk}\cdot\ln\left[\tilde{x}_{jk}(t)\right], \qquad i=\overline{1,n}, \qquad k=\overline{1,s}, \qquad t\in\left[T_1,T_2\right], \tag{55}$$

$$\tilde{x}_{ijk}(t) = \tilde{x}_{ijk}(T_1)e^{-\gamma_{jk}\cdot(t-t_0)} + \gamma_{jk}\int_{T_1}^{t}x_{ij}(\tau)e^{-\gamma_{jk}\cdot(t-\tau)}d\tau, \quad i=\overline{1,n}, \quad j=\overline{1,m}, \ k=\overline{1,s}, \ t\in\left[T_1,T_2\right], \tag{56}$$

$$\sum_{i=1}^{n}c_{ij}(t)x_{ij}(t) \le B_j(t), \qquad j=\overline{1,m}, \qquad t\in\left[T_1,T_2\right], \tag{57}$$

$$x_{ij}(t)\ge 0, \quad i=\overline{1,n} \quad , \quad j=\overline{1,m}, \quad t\in\left[T_1,T_2\right]. \tag{58}$$

Отметим, что в силу зависимостей (54)–(56) в качестве переменных этой динамической оптимизационной задачи в конечном счете выступают траектории $x_{ij}(t)$ ($i=\overline{1,n}$, $j=\overline{1,m}$, $t\in\left[T_1,T_2\right]$), описывающие динамику распределения ресурсов по регионам.

Можно показать [3], что решение непрерывновременной динамической оптимизационной задачи (53)–(58) также сводится к решению более простых статических задач (45)–(47) для каждого в отдельности значения $t\in\left[T_0,T\right]$. При этом в непрерывновременном случае:

$$\beta_{ij}(t) := \sum_{k=1}^{s} b_{jk}\gamma_{jk}\int_{t}^{T_2}\varphi(\tau)\alpha_{ik}(\tau)e^{-\gamma_{jk}\cdot(\tau-t)}d\tau, \qquad i=\overline{1,n} \quad , \quad j=\overline{1,m}, \qquad t\in\left[T_1,T_2\right]. \tag{59}$$

Оптимальные решения задач (53)–(58) в непрерывновременной случае также имеют вид (49), (50) (но при этом коэффициенты $\beta_{ij}(t)$ определены формулой (59), а не (48)).

В случае, когда весовые коэффициенты $\alpha_{ik}(t)$ и «краткосрочные» эластичности $\delta(t)$ постоянны (т.е. не зависят от периода времени *t*), можно показать [3], что

$$w_{ij}(t) := \frac{\tilde{\beta}_{ij}(t)}{c_{ij}(t)\sum_{l=1}^{n}\tilde{\beta}_{lj}(t)}, \qquad i=\overline{1,n}, \quad j=\overline{1,m}, \quad t\in\left[T_1,T_2\right]. \tag{60}$$

где

$$\tilde{\beta}_{ij}(t) := \sum_{k=1}^{s}\frac{b_{jk}\alpha_{ik}\gamma_{jk}}{\rho+\gamma_{jk}}\left[1-e^{-(\rho+\gamma_{jk})(T_2-t)}\right], \qquad i=\overline{1,n}, \quad j=\overline{1,m}, \quad t\in\left[T_1,T_2\right], \tag{61}$$

$$\rho := \ln\delta. \tag{62}$$

Оптимальное распределение ресурсов при бесконечном горизонте планирования. Рассмотрим случай, когда весовые коэффициенты $\alpha_{ik}(t)$ и «краткосрочные» эластичности $\delta(t)$ постоянны. Перейдя к пределу в формулах (51), (52) при $T_2 \to +\infty$, получим:

$$w_{ij}^{\infty} = \frac{\tilde{\beta}_{ij}}{c_{ij}\sum_{l=1}^{n}\tilde{\beta}_{lj}}, \qquad i=\overline{1,n}, \quad j=\overline{1,m}. \tag{63}$$

$$\tilde{\beta}_{ij}^{\infty} = \sum_{k=1}^{s}\frac{b_{jk}\alpha_{ik}\gamma_{jk}}{1-\delta\left(1-\gamma_{jk}\right)}, \quad i=\overline{1,n}, \quad j=\overline{1,m}. \tag{64}$$

В непрерывновременном случае перейдя к пределу в формуле (61) при $T_2 \to +\infty$, получим:

$$\tilde{\beta}_{ij}^{\infty} := \sum_{k=1}^{s}\frac{b_{jk}\alpha_{ik}\gamma_{jk}}{\rho+\gamma_{jk}}, \qquad i=\overline{1,n}, \quad j=\overline{1,m}. \tag{65}$$

Замечание 8. В силу формул (63)–(65) при бесконечном горизонте планирования (и при указанных выше допущениях) оптимальная структура распределения ресурсов по регионам постоянна (т.е. не зависит от периода *t*).

Заключение. В статье предложена методика построения межвременного интегрального социально-экономического показателя с использованием эластичностей замещения. В конечном счете значение указанного интегрального показателя определяется динамикой показателей выделяемых ресурсов по регионам. В соответствии с разработанной нами методикой ресурсы должны быть распределены по регионам таким образом, чтобы это приводило к максимизации указанного интегрального показателя. В частности, сделан вывод о том, что распределение долей ресурсов не зависит от объемов ресурсов в целом по национальной экономике. Также рассмотрен случай с бесконечным горизонтом планирования и получены формулы для расчета долей ресурсов по регионам. Данная методика позволяет улучшить динамику распределения выделяемых ресурсов по регионам, способствуя более эффективному развитию социально-экономической системы.

Список литературы

[1] Магнус, Я.Р. Эконометрика / Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий. – М.: Дело, 2004. – 576 с.

[2] Литвинович, А.А. Методика моделирования оптимального распределения показателей результативности жилищной политики с учетом запаздывания / А.А. Литвинович, М.М. Ерёменко, Э.М. Аксень // Белорус. экон. журн. – 2024. – № 2. – С. 81–97.

[3] Литвинович, А.А. Моделирование оптимального распределения ресурсов в сфере жилищной политики в непрерывном времени с учетом запаздывания / А.А. Литвинович, М.М. Еременко, Э.М. Аксень // Экономика, моделирование, прогнозирование : сб. науч. тр. / НИЭИ Мин-ва экономики Респ. Беларусь ; редкол.: М.К. Кравцов (гл. ред.) [и др.]. – Минск, 2024. – Вып. 18. – 316 с. – С. 271–284.

Авторский вклад

Аксень Эрнест Маврициевич – постановка задачи исследования в общем случае, методика построения межвременного интегрального показателя для непрерывновременного моделирования, методика оценивания параметров моделей, методика аналитического решения задачи максимизации межвременного интегрального показателя при ограничениях на ресурсы.

APPROACHES TO MODELING OPTIMAL RESOURCE DISTRIBUTION AMONG REGIONS

E.M. Aksen
Head of the Department of Economics and Management
at Belarus State Economic University, Doctor of Science in Economics, Professor

Abstract. This article presents approaches to modeling the optimal dynamics of resource distribution across regions for a socio-economic system in discrete and continuous time. The methodology involves constructing an intertemporal integral socioeconomic indicator and maximizing its value, taking into account constraints on explanatory factors. The lag in the influence of explanatory factors on the resulting indicators is also taken into account.

Keywords: dynamics, integral indicator, optimization, modeling, socio-economic system, region, elasticity, weighting coefficient.

УДК 005.591.6

МЕТОДОЛОГИЧЕСКИЙ ПОДХОД К БЕСШОВНОЙ ИНТЕГРАЦИИ ГЕНЕРАТИВНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА С ПРОЕКТАМИ ЦИФРОВОЙ ТРАНСФОРМАЦИИ БИЗНЕСА





Л.И. Архипова
Доцент кафедры экономики Белорусского государственного университета информатики и радиоэлектроники, кандидат экономических наук, доцент
l.arkhipova@mail.com

Л.Ф. Медведева
Доцент кафедры управления экономическими системами Академии управления при Президенте Республики Беларусь, кандидат экономических наук, доцент
medvedevaL15@mail.ru

Л.И. Архипова

Окончила Белорусский государственный университет и Академию управления при Президенте Республики Беларусь. Имеет многолетний опыт работы в реальном секторе экономики (НПО «Интеграл»).

Работает в Белорусском государственном университете информатики и радиоэлектроники.

Проводит научные исследования в области инновационных технологий и цифровой трансформации бизнеса

Л.Ф. Медведева

Окончила Белорусский государственный университет. Имеет практический опыт работы в радиоэлектронной отрасли экономики. Сфера научных интересов: синергетический подход в управлении, менеджмент устойчивого развития общества в условиях цифровой трансформации бизнеса.

.

Аннотация: в статье рассматривается проблема цифровой трансформации бизнеса с интеграцией инструментов генеративного искусственного интеллекта (GenAI). Авторы предлагают методологический подход, основанный на предварительной оценке цифровой зрелости бизнес-модели, ИТ-архитектуры, технологического стека и организационной готовности. Обосновывается необходимость органичного, «бесшовного» встраивания GenAI в существующие бизнес-процессы в отличие от практики фрагментарного внедрения. Представлена трёхуровневая модель интеграции, учитывающая специфику B2B-сегмента: длинный цикл сделки, множественность лиц, принимающих решения, и высокую ценность клиентских отношений.

К**лючевые слова**: генеративный искусственный интеллект (GenAI), цифровая зрелость, B2B-взаимодействия, фрагментарный подход, методологический подход интеграции ИИ.

Введение. Современный этап цифровой трансформации характеризуется переходом от автоматизации отдельных операций к комплексной перестройке бизнес-моделей под влиянием технологий искусственного интеллекта.

Особенно остро этот процесс проявляется в B2B-сегменте, где сложность взаимодействий, длительность цикла продаж и высокая стоимость сделки предъявляют повышенные требования к точности прогнозирования, качеству клиентского опыта и эффективности операционных процессов.

Как показывает практика, значительная часть проектов цифровизации B2B-компаний не достигает поставленных целей.

Причины кроются не столько в несовершенстве технологий, сколько в методологических просчётах: попытках внедрения «цифры» без оценки реальной готовности бизнеса, фрагментарной автоматизации, отсутствии связи между технологическими инициативами и стратегическими целями.

С появлением генеративного искусственного интеллекта (GenAI) эта проблема усугубилась: многие компании стремятся «добавить AI» в свои процессы без понимания того, как эта технология должна быть встроена в архитектуру, какие данные для этого необходимы и как изменится роль сотрудников.

Ключевые аспекты внедрения ИИ в действующую систему. Задача состоит в том, чтобы выработать методологический подход (т.е. методы достижения цели на основе знаний об организации деятельности объекта) к комплексной цифровизации бизнеса, обеспечивающий органичную, бесшовную интеграцию инструментов GenAI с действующей ИТ-архитектурой.

Ключевая идея данного подхода заключается в том, что успешное внедрение GenAI невозможно без предварительной оценки цифровой зрелости бизнес-модели, ИТ-архитектуры, технологического стека и готовности руководства. Такой подход позволяет не только избежать типичных ошибок, но и создать основу для устойчивого конкурентного преимущества.

Рассмотрим «фундамент» построения такой интеграции ИТ-решений и цифровой зрелости.

Понятие цифровой зрелости (digital maturity) является важным в дискуссиях о трансформации.

В отличие от ранних концепций, фокусировавшихся на уровне автоматизации, современные подходы учитывают комплекс факторов, что особенно важно для B2B-взаимодействий. Это выражается в высокой степени понимания клиентских потребностей и опыта, гибкости каналов продаж и способности предлагать персонализированные решения.

Анализ научных источников позволяет выделить четыре ключевых измерения цифровой зрелости, значимых для интеграции GenAI в B2B контексте (рисунок 1) [1, 2]:

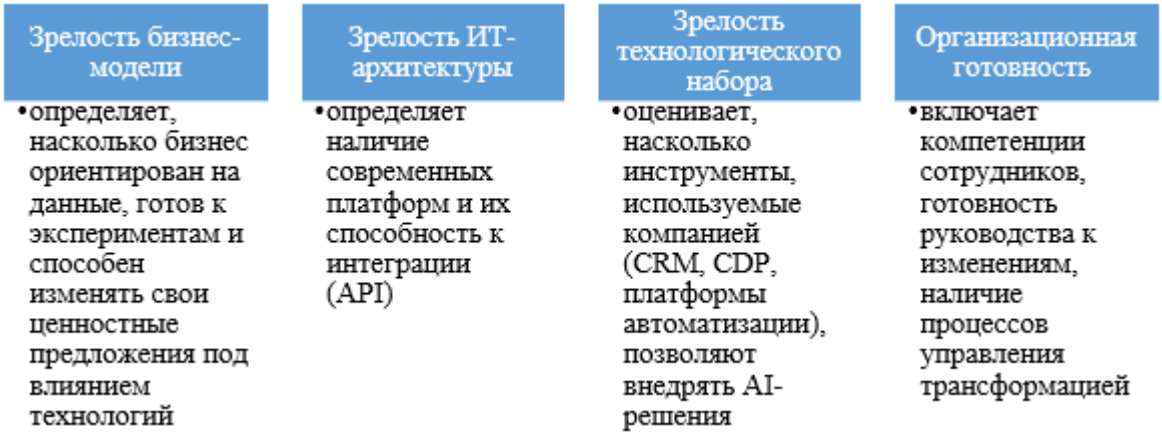


Рисунок 1. Четыре ключевых измерения цифровой зрелости, значимых для GenAI

Особое значение приобретает развитие навыков «промпт-инжиниринга» и работы с AI-ассистентами, а также формирование культуры принятия решений на основе данных (data-driven culture).

Современный ландшафт применения генеративного искусственного интеллекта охватывает практически все функциональные области бизнеса. Он перестаёт быть узкоспециализированным инструментом и превращается в сквозную технологию, способную воздействовать на всю цепочку создания ценности. Возможности ИИ выходят за пределы отдельных функций, охватывая маркетинг, продажи, производство, управление персоналом и другие области. В таблице 1 представлены бизнес-процессы маркетинга, скорость и качество которых в значительной степени определяются использованием инструментов ИИ.

Таблица 1. Применение GenAI в маркетинговой деятельности

Направление	Практическое применение GenAI
Маркетинг и реклама	– генерация текстов – посты для соцсетей, статьи, email-рассылки; – создание изображений и видео – рекламные креативы, баннеры; – SEO-оптимизация – генерация ключевых слов, мета-описаний; – персонализация контента – адаптация под сегменты аудитории; – сценарное планирование – A/B-тестирование гипотез;
Продажи и клиентский сервис	– генерация коммерческих предложений (КП) и писем; – чат-боты и голосовые ассистенты для клиентов (24/7); – анализ диалогов: выявление возражений, оценка тональности; – создание скриптов продаж и обучения продавцов; – персонализация рекомендаций на основе истории клиента;
Контент и креативные индустрии	– написание текстов, статьи, сценарии, книги, поэзия; – создание изображений, иллюстраций, 3D-моделей; – видеопроизводство: генерация анимации, монтаж, субтитры; – музыка и звук – создание треков, озвучка, синтез речи; – дизайн интерфейсов – генерация макетов, UI/UX.

Однако реализация даже самых очевидных сценариев требует системной подготовки, без которой инвестиции в технологии GenAI могут оказаться неэффективными.

Следует отметить, что B2B-взаимодействия имеют ряд особенностей, определяющих требования к цифровым решениям и, соответственно, интеграции с ИИ [1-3]:

Длинный цикл продажи (сделки). От момента генерации лида до заключения контракта могут проходить месяцы, что требует системы управления долгосрочными отношениями.

Множественность лиц, принимающих решения (ЛПР). В сделке, как правило, участвуют несколько заинтересованных лиц с разными ролями, интересами и критериями принятия решений. GenAI может анализировать коммуникации и предлагать персонализированные аргументы для каждой роли.

Высокая стоимость сделки, поэтому важна точность рекомендаций, качество контента и экспертиза (ошибки в продажах или обслуживании обходятся бизнесу дорого).

Сложность продуктов и решений. B2B-продукты часто требуют глубокой технической экспертизы, предпродажной подготовки (pre-sales) и индивидуализации коммерческих предложений.

Именно в этом контексте GenAI открывает новые возможности в автоматизации, создании сложных коммерческих предложений, проведении интеллектуального анализа коммуникаций для выявления «проблемных точек», персонализации взаимодействий, прогнозировании оттока клиентов, а также оценки вероятности подписания контракта (сделки).

Зарубежные специалисты отмечают, что ИИ значительно ускоряет жизненный цикл (развитие) цифровых технологий. Так, хорошо известные кривые Gartner Hype Cycle проходят период от верхней точки кривой (пик) до выхода на практическое внедрение (plato productivity) уже не за 5-10, а за 2-5 лет [4].

На основе анализа внедрения ИИ в B2B-компаниях, представленных в отчетах известных консалтинговых компаний (включая данные из отчётов McKinsey), можно выделить типичные проблемы, возникающие при фрагментарном подходе внедрения инструментов GenAI [5, 6]:

– Отсутствие единой стратегии. GenAI внедряется в отдельных подразделениях (например, маркетинге или клиентском сервисе) без связи с общей архитектурой и бизнес-целями.

– Недостаток данных и их качество. Модели требуют качественных, структурированных и актуальных данных, но во многих B2B-компаниях данные разрознены, хранятся в разных системах, не нормализованы.

– Интеграционные разрывы. ИИ-инструменты работают в изоляции, их выводы не интегрируются в CRM и процессы принятия решений.

– Культурное неприятие. Сотрудники воспринимают ИИ как угрозу, а не как инструмент усиления их компетенций; отсутствует обучение и поддержка.

– Недооценка управленческой составляющей. Внедрение GenAI требует новых компетенций менеджмента: управления экспериментами, оценки рисков (галлюцинации, предвзятость алгоритмов), пересмотра KPI.

Речь в данном случае идет не о компаниях, рожденных цифровыми (digital born), а об организациях, которые находятся на стадии цифровой трансформации и начинающих использовать технологии и инструменты ИИ в своей деятельности (Indie Tools). Здесь больше подойдет классификация, которая в 2025 году представлена в отчете Chiefmartec&MartechTride: Indie Tools (независимые инструменты – компании); Chalkenger Planforms (новые встроенные платформы); Incumbent Platforms (действующие доминирующие платформы) [4, 7].

С целью выяснения текущего состояния и возможности внедрения инструментов GenAI в различных секторах B2B в Республике Беларусь, был проведен опрос специалистов и руководителей различных предприятий Республики Беларусь (энергетический комплекс, промышленное производство, связь, услуги, транспорт и логистика). В опросе участвовало

64 респондента (30% – топ-менеджеры, 44% – руководители среднего звена, 18% – специалисты и др.). Опрос проводился по *четырем блокам*: осведомлённость и опыт использования ИИ на предприятии; отношение к внедрению ИИ; барьеры и потребности; готовность к изменениям.

Результаты проведенного опроса по блоку 1«Осведомлённость и опыт использования ИИ» представлены на рисунках 2 – 4.



Рисунок 2. Оценка уровня понимания возможностей ИИ



Рисунок 3. Оценка уровня использования ИИ на предприятии



Рисунок 4. Оценка категорий решаемых задач

В данном блоке респонденты отвечали также на вопрос – пользуются ли они лично инструментами на основе ИИ (например, ChatGPT, голосовыми помощниками, нейросетями для обработки фото/текста). Более 30% респондентов ответили, что пользуются регулярно (почти каждый день) и почти столько же – около 30% пользуются иногда (раз в неделю), 6% ответов – пользуются постоянно. Эти ответы показывают, что специалисты и руководители начинают самостоятельно изучать и применять на практике новые технологии со встроенным ИИ. Однако такая практика не может быть признана «абсолютно» положительной, так как это может привести к «разрозненным» результатам и данным, на базе которых принимаются управленческие и технические решения. Именно поэтому нужен комплексный подход внедрения ИИ – с регламентами и ограничениями, предотвращающими этические и экономические риски.

Результаты проведенного опроса по блоку 2 «*Отношение к внедрению ИИ*» представлены в таблице 2. Ответы оценивались по шкале от **1 до 5**, где **1** – полностью не согласен, **5** – полностью согласен.

Таблица 2. Отношение к внедрению ИИ на предприятии

№	Утверждение	Оценка (к-во)				
		1	2	3	4	5
1	ИИ поможет повысить производительность моего труда		8	20	18	18
2	ИИ может выполнять рутинные задачи лучше, чем человек		8	16	24	10
3	Я доверяю результатам работы ИИ, если он правильно обучен		8	24	20	10
4	Внедрение ИИ несёт угрозу моей занятости (может лишить работы)	**24**	**20**	20		
5	Руководство предприятия уделяет достаточно внимания внедрению современных технологий		16	18	10	20
6	Я готов проходить обучение, чтобы освоить новые инструменты ИИ				24	40
7	Внедрение ИИ сделает мою работу более интересной, избавив от рутины		6	16	32	10
8	Я боюсь, что не смогу освоить новые технологии из-за недостатка знаний или возраста	**28**	**20**	12	4	
9	Компания должна обеспечить обучение всех сотрудников перед внедрением ИИ		12	12	12	28
10	Я готов передать часть своих текущих задач ИИ, если это будет чётко регламентировано и безопасно		8	12	24	20

Выводы по результатам анализа ответов респондентов на вопросы блока 2:

– Сотрудники ожидают, что ИИ несет для них угрозу потери рабочего места (п.4), а также опасения, что они не смогут освоить новые технологии и инструменты (п.8).

– Достаточно большое количество респондентов нейтрально относятся к внедрению новых технологий на базе ИИ – поставили оценку 3. Это, как раз те сотрудники, которые готовы учиться и работать по новым регламентам и стандартам на своих рабочих местах.

– Респонденты, оценившие текущую ситуацию на 4 и 5 баллов, фактически готовы к внедрению ИИ в бизнес-процессы и передать часть текущих задач ИИ.

Результаты проведенного опроса по блоку 3 «Барьеры и потребности» представлены на рисунке 5.

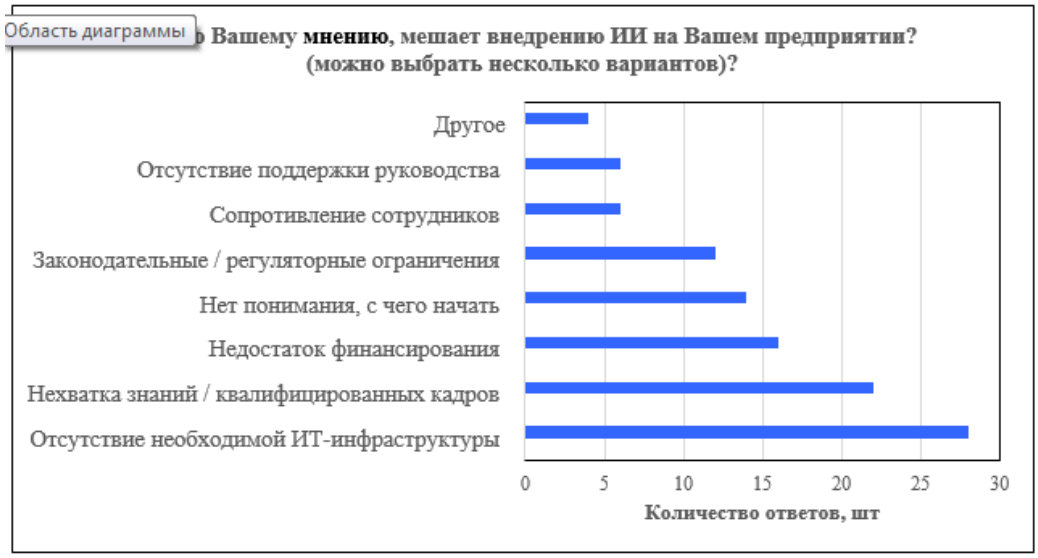


Рисунок 5. Барьеры на пути к внедрению ИИ и их преодоление

Ответы на вопрос «Какая поддержка потребовалась бы Вам лично для эффективной работы с ИИ» подтверждают, что большинство сотрудников готовы к обучению, однако они ожидают наличия разработанных инструкций и регламентов, а лучше – доступа к готовым ИТ-решениям с интегрированным ИИ.

Результаты опроса по блоку 4 «Готовность к изменениям и предложения» при ответе на вопрос «Оцените по шкале 1–5, насколько Вы лично готовы к тому, что в ближайшие 2–3 года Ваша работа изменится под влиянием ИИ (1 – совсем не готов, 5 – полностью готов) более 50% респондентов дали ответ – 3. То есть, при достаточно позитивном отношении к внедрению ИИ, сомнения присутствуют. Главное – чтобы они не превратились в «глубокое» сопротивление. Основной вывод, сделанный по результатам проведенного опроса, состоит в следующем: существует проблема «фрагментарного внедрения» GenAI, когда на предприятии отсутствует стратегия внедрения ИИ и системное обучение сотрудников, что, практически, совпадет с выводами международных консалтинговых агентств, изучающих аналогичные проблемы, связанные с системной интеграцией инструментов ИИ в действующие бизнес-процессы [3–7].

На основании проведенного исследования и анализа разработан проект методологии, представленный в виде пошагового руководства по комплексной интеграции GenAI в действующие бизнес-процессы на основе оценки их цифровой зрелости. Предлагаемый подход основан на последовательной реализации трёх этапов, каждый из которых предполагает оценку и «дошлифовку» соответствующих факторов и аспектов цифровой зрелости (Таблица 3).

Таблица 3. Пошаговое руководство по комплексной интеграции GenAI в действующие бизнес-процессы на основе оценки цифровой зрелости

Этап 1. Диагностика цифровой зрелости *Цель: определение стартовой позиции и выявление «узких мест», которые могут заблокировать интеграцию GenAI*	
Оценка зрелости бизнес-модели	Анализ того, насколько клиентоцентрична компания, используются ли данные для принятия решений, существует ли культура экспериментов.
Оценка ИТ-архитектуры	Выявление степени интеграции между CRM, ERP, маркетинговыми платформами, качеством данных, наличием API-интерфейсов.
Оценка технологического стека	Аудит используемых инструментов на предмет совместимости с II-решениями, возможностью работы с Big Data, наличием CDP.
Оценка организационной готовности	Уровень компетенций сотрудников, отношение руководства к инновациям, наличие процессов.
Этап 2. Проектирование бесшовной интеграции GenAI *Задача*: *разработка бизнес-модели, в которой инструменты GenAI встраиваются в бизнес-процессы как органичная их часть, а не как внешняя надстройка*	
AI как помощник сотруднику, а не его замена	GenAI автоматизирует рутинные задачи (подготовка писем, итоги переговоров/summary, первичный анализ данных), освобождая время сотрудников для стратегической работы, креативности и сложных коммуникаций.
Единый клиентский профиль (Single Customer View)	GenAI работает на основе централизованных данных из CDP/CRM, что обеспечивает контекстную персонализацию и снижает риск «галлюцинаций».
Встраивание в существующие интерфейсы	Инструменты GenAI интегрируются в привычную для сотрудников среду: CRM, мессенджеры, офисные пакеты, колл-центры.
Цикл обратной связи	Предусматривается механизм сбора обратной связи от сотрудников и клиентов для «дообучения» моделей и коррекции сценариев.
Этап 3. Поэтапное внедрение с масштабированием *Диагностика – проектирование бесшовной интеграции – поэтапное внедрение с масштабированием*	
Пилотный проект	Выбирается один бизнес-процесс с высокой добавленной ценностью (например, автоматизация подготовки коммерческих предложений или анализ качества переговоров). Критерии успеха чётко определены (KPI).
Обучение и адаптация	На основе пилотного проекта корректируются подходы, «дообучаются» модели, разрабатываются инструкции для сотрудников.
Масштабирование	Успешные практики распространяются на другие подразделения и процессы, постепенно формируется единая ИИ-экосистема.

Предложенная методология позволяет системно подойти к трансформации, обеспечивая органичное встраивание ИИ-инструментов в существующие процессы и архитектуру.

Ключевыми условиями успеха выступают: наличие единого клиентского профиля на основе качественных данных, интеграция ИИ-инструментов в привычные должностные обязанности сотрудников, развитие управленческих компетенций и формирование культуры ответственного использования ИИ.

Заключение. Проведённое исследование позволяет сделать вывод, что успешная интеграция генеративного искусственного интеллекта в B2B-бизнес возможна только при условии комплексного подхода, включающего предварительную оценку цифровой зрелости бизнес-модели, ИТ-архитектуры, технологического стека и организационной готовности.

Фрагментарное внедрение GenAI, ориентированное исключительно на технологическую составляющую, как правило, не даёт устойчивых результатов и может привести к дополнительным издержкам и организационному сопротивлению.

В условиях, когда генеративный искусственный интеллект перестаёт быть «технологией будущего» и становится инструментом повседневной работы, способность компаний к его системному, осознанному и этичному внедрению превращается в ключевой фактор конкурентоспособности. Комплексный подход, основанный на предварительной диагностике цифровой зрелости и бесшовной интеграции ИИ, позволяет не только снизить риски, но и получить максимальную отдачу от инвестиций в цифровую трансформацию, включая ИИ.

Список литературы

[1] Achieving Digital Maturity Adapting Your Company to a Changing World SUMMER 2017 RESEARCH REPORT By y Gerald C. Kane, Doug Palmer, Anh Nguyen Phillips, David Kiron, and Natasha Buckley / Deloite/ University Press. MIT Sloan Management Review [Electronic resource]. – Mode of access https://www.the-digital-insurer.com/wp-content/uploads/2017/11/1105-59180-MITSMR-Deloitte-Digital-Report-2017.pdf

[2] Arkhipova L.I., Digital maturity categories in business: integrating and assessment BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сб. науч. ст. XI Междунар. науч.-практ. конф. (Республика Беларусь, Минск, 23–24 апреля 2025 года) / редкол.: В. А. Богуш [и др.]. – Минск : БГУИР, 2025. – С.365-371

[3] Brinker, S. 2023 Marketing Technology Landscape Supergraphic: 11,038 solutions searchable on martechmap.com [Electronic resource]. – Mode of access: https://chiefmartec.com/2023/05/2023-marketing-technology-landscape-supergraphic-11038-solutions-searchable-on-martechmap

[4] Martech for 2025 by Scott Brinker and Frans RiemersmaScott Chiefmartec & MartechTride [Electronic resource]. – Mode of access: https://chiefmartec.com/wp-content/uploads/2024/12/martech-for-2025-report.pdf

[5] McKinsey Technology Trends Outlook 2025 [Electronic resource] / McKinsey & Company. – Mode of access: https://www.mckinsey.com/capabilities/tech-and-ИИ/our-insights/the-top-trends-in-tech.

[6] McKinsey & Company. (2024). The state of AI in 2024: GenAI's breakout moment. [Electronic resource]. – Mode of access: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024

[7] Brinker, S., Martech-2030. Five Trends in Marketing Technology for the Decade of the Augmented Markete/Scott Brinker, Jason Baldwin [Electronic resource]. – Mode of access: https://chiefmartec.com/wp-content/uploads/2021/08/martech-2030-brinker-baldwin.pdf

Авторский вклад

Архипова Лариса Ивановна – постановка задачи исследования, разработка методологии внедрения инструментов генеративного искусственного интеллекта на предприятии. Разработка анкет для исследования и обработка полученных результатов.

Медведева Людмила Феофановна – постановка задачи исследования, разработка методологии внедрения инструментов генеративного искусственного интеллекта на предприятии. Проведение опроса и обработка полученных результатов.

A METHODOLOGICAL APPROACH TO SEAMLESS INTEGRATION OF GENERATIVE ARTIFICIAL INTELLIGENCE WITH DIGITAL BUSINESS TRANSFORMATION PROJECTS

L.I. Arkhipova
Associate professor, PhD, Belarussian State University of Informatics and Radioelectronics

L.F. Medvedeva
Associate professor, PhD, Academy of Management under the President of the Republic of Belarus

Abstract: This article examines the problem of digital business transformation with the integration of generative artificial intelligence (GenAI) tools. The authors propose a methodological approach based on a preliminary assessment of the digital maturity of the business model, IT architecture, technology stack, and organizational readiness. They substantiate the need for organic, seamless integration of GenAI into existing business processes, as opposed to the practice of piecemeal implementation. A three-tier integration model is presented, taking into account the specific characteristics of the B2B segment: long transaction cycles, multiple decision-makers, and the high value of customer relationships.

Keywords: generative artificial intelligence (GenAI), digital maturity, B2B interactions, fragmented approach, methodological approach to AI integration.

УДК 544.183.25:004.422.8

АНАЛИЗ И РЕФАКТОРИНГ БИБЛИОТЕКИ ОБМЕННО-КОРРЕЛЯЦИОННЫХ ФУНКЦИОНАЛОВ ПАКЕТА OPENMX



А.В. Баглов
Старший научный сотрудник НИЛ энергоэффективных материалов и технологий физического факультета БГУ; научный сотрудник Центра 4.11 НИЧ БГУИР
baglov@bsu.by



Л.С. Хорошко
Ведущий научный сотрудник НИЛ энергоэффективных материалов и технологий физического факультета БГУ; ведущий научный сотрудник Центра 4.11 НИЧ БГУИР, канд. физ.-мат. наук, доцент
khoroshko@bsu.by

А.В. Баглов

Окончил Белорусский государственный университет информатики и радиоэлектроники. Проводит научные исследования в области квантово-механического моделирования структурных и электронных свойств перспективных материалов.

Л.С. Хорошко

Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с управляемым синтезом наноматериалов и их исследованием их свойств.

Аннотация. Проведен рефакторинг кода обменно-корреляционных функционалов пакета OpenMX, устранивший недочеты оригинала и оптимизировавший производительность. Созданная кроссплатформенная библиотека на ANSI C обеспечивает ускорение до 4 раз в среде Linux и сокращение объема исходного кода в 2,5 раза. Работа закладывает основу для дальнейшей интеграции современных функционалов, необходимых в теоретической физике твердого тела, наноматериаловедении и для развития материаловедческих ИИ-моделей с использованием пакета OpenMX. Открытый характер лицензии пакета OpenMX и

самодостаточность разработанной библиотеки способствуют ее интеграции в международное научное сообщество и использованию в образовательном процессе.

Ключевые слова: моделирование, *ab initio*, OpenMX, теория функционала плотности, программирование, ANSI, C.

Введение. Исследование свойств материалов путем численного моделирования является актуальным и востребованным научным направлением. В современной вычислительной физике твердого тела и наноматериаловедении для исследования структурных и электронных свойств объектов различной размерности наибольшее распространение получили так называемые *ab initio* (из первых принципов) методы, в основе которых лежат приближенные методы квантовой механики для решения многоэлектронной задачи [1].

Результаты, получаемые в процессе *ab initio* моделирования, позволяют значительно расширить, уточнить и дополнить данные, полученные экспериментальным путем [2].

Высокая достоверность предсказания кристаллической и электронной структуры различных соединений, оценка их стабильности и получение термодинамических характеристик делает эти методы незаменимым инструментом для компьютерного дизайна новых перспективных материалов для передовых областей науки и техники [3].

Центральное место среди *ab initio* подходов занимает теория функционала плотности (ТФП), сочетающая высокую точность и разумную вычислительную сложность, что позволяет исследовать системы различной размерности [4].

В рамках ТФП основное энергетическое состояние системы описывается через распределение электронной плотности в ней, а ключевым объектом выступает функционал, связывающий данное распределение с полной энергией системы [5].

В формализме Кона-Шэма данный функционал включает кинетическую энергию невзаимодействующих электронов, кулоновское взаимодействие, внешний потенциал и обменно-корреляционный член, включающий сложные квантовые эффекты межэлектронного взаимодействия. Ввиду отсутствия точного аналитического решения многоэлектронной задачи разработано множество обменно-корреляционных функционалов [6]. Проведение *ab initio* исследований реализуется с помощью специализированных научных пакетов, доступных как на коммерческой основе, так и бесплатно. Авторы данной работы в своих исследованиях активно задействуют пакет OpenMX (Open source package for Material eXplorer), бесплатно распространяемый под лицензией GPLv3, в котором ТФП сочетается с теорией псевдопотенциала [7–14].

Данный пакет был разработан в качестве высокоэффективного инструмента, оптимизирован для параллельной работы на большом числе вычислительных узлов, что позволяет проводить численное моделирование электронных и структурных свойств систем, содержащих сотни и тысячи атомов, в том числе металлорганических каркасных структур, наношнуров, нанотрубок, нанолент, наночастиц и биомолекул. Такая эффективность достигается, среди прочего, использованием в качестве базиса для описания электронных состояний численных псевдоатомных орбиталей, снижающих вычислительные затраты при сохранении высокой точности, необходимой для описания сложных комплексных наносистем.

Особенностью данного пакета является малое количество доступных обменно-корреляционных функционалов, которое обусловлено его научной специализацией. Поскольку архитектура пакета ориентирована на максимальную вычислительную эффективность при моделировании крупномасштабных систем с сохранением физической достоверности и численной точности, разработчики пакета сфокусировались на использовании наиболее широко применяемых обменно-корреляционных функционалов.

К настоящему моменту такой подход несколько устарел и ограничивает точность проводимых исследований при использовании данного пакета.

Целью данной работы является анализ и рефакторинг исходного кода пакета OpenMX, отвечающего за реализацию ТФП и разработка библиотеки, унифицирующей и расширяющей стандартный набор обменно-корреляционных функционалов.

Анализ текущих возможностей пакета. Согласно руководству пользователя в пакете OpenMX в настоящее время для использования доступны обменно-корреляционные функционалы, обозначаемые как LDA, LSDA-CA, LSDA-PW и GGA-PBE [15]. Большинство обозначений имеют составную структуру, где первая (левая) часть обозначает приближение, а вторая (правая) часть – аббревиатуру фамилий авторов, предложивших конкретную параметризацию (обменно-корреляционный функционал). LDA – Local Density Approximation – соответствует приближению локальной плотности (ПЛП), а LSDA – Local Spin Density Approximation – приближению локальной спиновой плотности (ПЛСП) соответственно. GGA – Generalized Gradient Approximation – соответствует обобщенному градиентному приближению, доступному в спин-неполяризованном и спин-поляризованном вариантах. Аббревиатуры CA, PW и PBE расшифровываются как Ceperley and Alder, Perdew and Wang, и Perdew and Burke and Ernzerhof соответственно [16–18]. Отметим, что в пакете OpenMX LDA эквивалентен спин-неполяризованному LSDA-CA. Также важно отметить, что в вариантах LDA/LSDA-CA, не глядя на аббревиатуру, используется репараметризация Perdew and Zunger [19].

За программную реализацию перечисленных обменно-корреляционных функционалов в пакете OpenMX отвечают следующие файлы: XC_CA_LSDA.c, XC_Ceperly_Alder.c, XC_EX.c, XC_PBE.c и XC_PW92C.c.

Анализ исходного кода показывает, что разработчики OpenMX адаптировали часть кода проекта SIESTA [20]. Поскольку пакет OpenMX написан на языке C, код, изначально созданный на Fortran, был перенесен с соответствующей адаптацией синтаксиса.

С нашей точки зрения представляется целесообразным выделение кода, связанного с обменно-корреляционными функционалами, во внутреннюю библиотеку, что позволит инкапсулировать доступные обменно-корреляционные функционалы от основного кода пакета и одновременно сохранит его архитектуру. Также инкапсуляция кода позволит существенно упростить дальнейшую разработку, сопровождение и поддержку подсистемы обменно-корреляционных функционалов.

Обсуждение требований к разработке. Для успешной реализации внутренней библиотеки обменно-корреляционных функционалов в пакете OpenMX необходимо рассмотреть ряд вопросов.

Поскольку пакет предназначен для научных исследований, то целевыми системами для его установки являются серверы суперкомпьютерных комплексов и вычислительные кластеры университетов и научных лабораторий. Анализ рейтинг суперкомпьютеров СНГ «Топ-50» показывает, что в составе суперкомпьютеров используются процессоры преимущественно 2014–2021 гг. [21]. Локальные кластеры научных лабораторий могут быть еще старше и неоднороднее. Таким образом следует обеспечить возможность сборки модифицированного пакета OpenMX на оборудовании более чем десятилетней давности.

С практической точки зрения необходимо обеспечить воспроизводимость результатов вычислений пакетом OpenMX до и после разработки библиотеки.

Независимость библиотеки обменно-корреляционных функционалов от основного кода также позволит оценить изменения в производительности непосредственно данного кода, поскольку оценка этих изменений путем анализа производительности самого пакета является нетривиальной задачей ввиду большого количества прочих вычислений.

Данный пакет также используется в рамках педагогической деятельности. Регулярно проводятся т.н. воркшопы для подготовки исследователей и студентов. Также известно о его использовании составе лекций и практикумов по курсу «Введение в физику твердого тела» Орегонского университета [22], и спецкурсов на физическом факультете БГУ. Библиотека обменно-корреляционных функционалов также может быть использована

студентами теоретических специальностей в качестве автономного инструмента для численного моделирования таких фундаментальных объектов как однородный электронный газ и вигнеровские кристаллы, а также осваивания приближенных методов теории многих частиц без необходимости развертывания полного программного комплекса OpenMX.

Поскольку сам пакет ожидаемо ориентирован на Unix-подобные операционные системы, следует обеспечить возможность компиляции библиотеки и в нативном окружении операционной системы Windows для использования в педагогической практике. Это позволит закрепить навыки программирования у студентов и улучшит понимания современных методов описания электронной структуры конденсированных сред.

Обсуждение результатов. Код обменно-корреляционных функционалов, как и сам пакет OpenMX, написан на языке программирования C.

В качестве стандарта авторы использовали ANSI, который также гарантированно поддерживается в компиляторе MSVC. Однако возможность компиляции кода в текущей версии в Windows ограничена использованием заголовочного файла openmx_common.h, содержащего ряд структур и макросов.

Для тестирования воспроизводимости и производительности на разных платформах мы модифицировали исходный код таким образом, чтобы исключить только привязку к указанному заголовочному файлу.

Непосредственно сам исходный код обменно-корреляционных функционалов написан в императивном стиле, что и следует из его сути. Каждый функционал реализован в отдельной функции, за исключением PBE, в котором обменная энергия вычисляется также через отдельную функцию, вызываемую в рамках выполнения основной.

В реализации разработчиков OpenMX мы обнаружили достаточно большое количество неиспользуемых переменных, что не влияет на функциональность, но затрудняет читаемость и анализ кода. По всей видимости это является следствием переноса с Fortran и незавершённой оптимизации производительности для компиляторов начала 2000-х годов, о чем говорит вынесение результатов части некоторых вычислений в локальные переменные и отсутствие ручной размотки циклов из двух итераций.

Поскольку ключевая идея ТФП – это связь между энергией и распределением электронной плотности, то во все функционалы входит фундаментальный параметр теории однородного электронного газа – радиус Вигнера-Зейтца. Данный параметр характеризует плотность системы через объем, приходящийся на один электрон.

Поскольку эти функции вызываются многократно (порядка 10^5–10^6 раз на каждую итерацию цикла самосогласования), в процессе рефакторинга следует уделить внимание их оптимизации. Исходно авторы для нахождения радиуса Вигнера-Зейтца использовали функцию pow, возводящую основание в степень 1/3.

В операционной среде Linux оптимизатор компиляторов GCC, Clang и ICC с уровня -O2 обеспечивал замену функции на более быструю функцию взятия кубического корня cbrt и ее встраивание, в то время как MSVC ограничивался только встраиванием функции. Части кода, использующие ветвления, были перестроены таким образом, чтобы первой стояла ветвь, которая будет наиболее вероятно исполняться. Учитывая, что глубина вложенности не превышает трех уровней, такой код не является проблемой для предсказателя ветвлений современных процессоров. Тем не менее, такой подход делает код более компактным, что упрощает его поддержку.

Также мы по возможности осуществили замену процедур деления на умножение, где это было разумно. Проанализировав код, мы завершили начатый разработчиками пакета вынос некоторых частей вычислений в отдельные переменные, которые с большей вероятностью остаются в процессорном кэше и способствуют ускорению вычислений.

В оригинальном коде используется отсечка малых значений электронной плотности, что является стандартной практикой, т.к. ее околонулевое значение часто приводит к

численным неустойчивостям. Исходно это значение различалось для различных функционалов.

Мы выбрали единое значение 10^{-14} для всех случаев, как это принято в большинстве случаев. Заметим, что из-за исходной рассогласованности в функционале LSDA-PW92 была допущена ошибка при обработке плотности ниже порога отсечки. Это могло несколько искажать получаемые значения энергии для систем с большими участками вакуума. В случае спин-поляризованных систем разработчики пакета по-разному проводили их обработку. Помимо опечаток было обнаружено, что выбор LSDA-PW92 и GGA-PBE мог привести к аварийному завершению работы пакета при неудачном выборе спиновой конфигурации при инициализации расчета.

Вероятность такого события мы оцениванием как крайне малую, поскольку это потребовало бы использования нефизичных входных данных. Однако, соответствующие исправления внесены.

В результате выполненной работы файл xclib.c содержит исправленную и оптимизированную библиотеку обменно-корреляционных функционалов, ранее содержавшихся в пяти отдельных файлов. В результате рефакторинга объем исходного кода уменьшился примерно в 2,5 раза. В свою очередь, это привело к уменьшению объектного файла в 1,8 раза, что свидетельствует о большей компактности кода в машинном представлении. Проверена работоспособность пакета OpenMX с новой версией внутренней библиотеки обменно-корреляционных функционалов.

Для тестирования была написан враппер для библиотеки, и подготовлены наборы тестов для всех функционалов. Анализ показал, что наблюдется небольшое расхождение в тестовых наборах данных, поскольку исходные численные константы были заменены на более точные, а также была изменена последовательность и тип операций. Усредненное отклонение составило менее 10^{-8} по всему тестовому набору.

При анализе производительности было обнаружено, что в операционной системе Linux с использованием компилятора GCC с уровнем оптимизации -O2 новая библиотека в 2–4 раза быстрее исходной, в то время как в операционной системе Windows с использованием компилятора MSVC с таким же уровнем оптимизации новая библиотека медленнее оригинальной в среднем на 30%.

Анализ показал, что реализация функции cbrt в библиотеке LibM значительно быстрее функции pow, использованной в оригинале, в то время как в библиотеке MSVCRT ситуация обратная – cbrt значительно медленнее pow, что подтверждено отдельными тестами. Детальное тестирование воспроизводимости и анализ производительности будет представлен в отдельной публикации.

Заключение. Проанализирован исходный код пакета OpenMX, отвечающий за обменно-корреляционные функционалы. Установлено, что данный код получен портированием с языка Fortran проекта Siesta. Обнаружено, что в процессе переноса были созданы неиспользуемые переменные, допущены ошибки в обработке сильно спин-поляризованных систем, а также при обработке низких значений распределения электронной плотности, а низкоуровневая оптимизация проведена не полностью.

В результате проведения рефакторинга и оптимизации исходного кода предложена новая внутренняя библиотека обменно-корреляционных функционалов, в текущий момент полностью сохраняющая функциональность оригинала, но содержащая необходимые исправления.

Библиотека является кроссплатформенной и может быть скомпилирована на любой платформе, поддерживающей ANSI C.

Рефакторинг позволил сократить размер исходного кода в примерно 2,5 раза, а объем объектных данных уменьшить примерно в 1,8 раза при повышении читаемости и упрощении сопровождения.

Оптимизация позволила добиться ускорения от двух до четырех раз в операционной системе Linux, в то время как использование компилятора MSVC в операционной системе Windows приводит к замедлению примерно на 30%, что обусловлено особенностями реализации функции взятия кубического корня cbrt в стандартных математических библиотеках.

Разработанная библиотека послужит фундаментом для интеграции современных обменно-корреляционных функционалов (PBEsol, RPBE, AM05 и др.), востребованных в теоретической физике твердого тела и наноматериаловедении.

Расширенная версия библиотеки будет особенно полезна для подготовки набора референтных данных, пригодных для обучения больших языковых моделей, применяемых в прикладном материаловедении, например, для генерации высокоточных датасетов электронных структур.

Уникальность данной работы для Республики Беларусь заключается в создании высокотехнологичного программного решения в области квантово-механического моделирования, не имеющего отечественных аналогов.

При этом открытый характер лицензии GPLv3 обеспечивает интеграцию библиотеки в международное научное сообщество, подтверждая высокий уровень и конкурентоспособность белорусской научной школы.

При этом самодостаточность библиотеки позволяет использовать её в текущем виде в учебном процессе для студентов теоретических специальностей, исключая необходимость полного развертывания пакета OpenMX.

Благодарности. Работа частично поддержана БРФФИ (проект № Т25МЭ-007), а также в рамках задания № 1.2.2 «Тонкопленочные и двумерные полупроводниковые материалы на основе углерода, халькогенидов, в том числе тугоплавких металлов, и неорганических перовскитов для приложений сенсорики, фотовольтаики, нано- и оптоэлектроники» подпрограммы «Физика конденсированного состояния, разработка технологий создания функциональных инструментальных материалов и устройств», НИР «Низкотемпературный синтез и свойства наноструктурированных композитов на основе перовскитных соединений с кластерами редкоземельных элементов и углерода для конвертеров оптического излучения и фотовольтаики» подпрограммы «Наноструктуры, наноматериалы и технологии» ГПНИ «Современное материаловедение, перспективные материалы и новые технологии»; НИР «Разработка научных основ и методов сенсибилизации, детектирования и фотокаталитической деструкции органических загрязнителей, включая микро- и нанопластики, в водных средах с применением гибридных мультифункциональных нанокомпозитов на основе оксидных полупроводников» подпрограммы «Междисциплинарные исследования и перспективные технологии» ГПНИ «Конвергенция 2030».

Список литературы

[1] Hohenberg, P. Inhomogeneous Electron Gas / P. Hohenberg, W. Kohn // Physical Review. – 1964. – Vol. 136. – PP. B864–B871.

[2] Kawazoe, Y. How well can physical, chemical, and mechanical properties of materials be predicted by ab initio techniques? / Y. Kawazoe // Materials & Design. – 2001. – Vol. 22, № 1. – PP. 61–67.

[3] Shevlin, S. Computational materials design / S. Shevlin, A. H. Castro Neto // Nature Materials. – 2021. – Vol. 20. – PP. 727–728.

[4] Kohn, W. Nobel Lecture: Electronic structure of matter–wave functions and density functionals / W. Kohn // Reviews of Modern Physics. – 1999. – Vol. 71, № 5. – PP. 1253–1266.

[5] Kohn, W. Self-Consistent Equations Including Exchange and Correlation Effects / W. Kohn, L. J. Sham // Physical Review. – 1965. – Vol. 140, № 4A. – PP. A1133–A1138.

[6] Perdew, J. P. Generalized Gradient Approximation Made Simple / J. P. Perdew, K. Burke, M. Ernzerhof // Physical Review Letters. – 1996. – Vol. 77, № 18. – PP. 3865–3868.

[7] Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures / T. Ozaki // Physical Review B. – 2003. – Vol. 67, № 15. – P. 155108.

[8] Ozaki, T. Numerical atomic basis orbitals from H to Kr / T. Ozaki, H. Kino // Physical Review B. – 2004. – Vol. 69, № 19. – P. 195113.

[9] Ozaki, T. Efficient projector expansion for the ab initio LCAO method / T. Ozaki, H. Kino // Physical Review B. – 2005. – Vol. 72, № 4. – P. 045121.

[10] Baglov, A. V. Atom Species Energy Dependence on Magnetic Configurations in the Perovskite Yttrium Orthoferrite / A. V. Baglov, L. S. Khoroshko // Doklady BGUIR. – 2021. – Vol. 19, № 8. – P. 63–67.

[11] Baglov, A. V. Crystal Structure and Electronic Properties of Rhenium Disulfide / A. V. Baglov, L. S. Khoroshko // Journal of Applied Spectroscopy. – 2022. – Vol. 89. – P. 860–864.

[12] Baglov, A. Evolution of Structural and Electronic Properties Standardized Description in Rhenium Disulfide at the Bulk-Monolayer Transition / A. Baglov, L. Khoroshko, A. Zhoidzik [et al.] // Heliyon. – 2024. – Vol. 10. – P. e28646.

[13] Baglov, A. V. Electronic Structure of Bismuth Ferromanganite $BiFe_{0.5}Mn_{0.5}O_3$ / A. V. Baglov, L. S. Khoroshko, M. V. Silibin, D. V. Karpinsky // Semiconductors. – 2024. – Vol. 58, № 13. – P. 1054–1059.

[14] Baglov, A. V. Structural and Electronic Properties of SmGaGe2O7 Studied by First Principles Methods / A. V. Baglov, L. S. Khoroshko // Inorganic Materials. – 2023. – Vol. 59, № 1. – P. 1–7.

[15] Ozaki, T. Functional // User's manual of OpenMX Ver. 3.9 / T. Ozaki, H. Kino, J. Yu [et al.]. – 2018. – URL: https://openmx-square.org/openmx_man3.9/node23.html (дата обращения: 27.03.2026).

[16] Ceperley, D. M. Ground State of the Electron Gas by a Stochastic Method / D. M. Ceperley, B. J. Alder // Physical Review Letters. – 1980. – Vol. 45, № 7. – P. 566.

[17] Perdew, J. P. Accurate and Simple Analytic Representation of the Electron-Gas Correlation Energy / J. P. Perdew, Y. Wang // Physical Review B. – 1992. – Vol. 45. – P. 13244.

[18] Perdew, J. P. Generalized Gradient Approximation Made Simple / J. P. Perdew, K. Burke, M. Ernzerhof // Physical Review Letters. – 1996. – Vol. 77, № 18. – P. 3865.[19] Perdew, J. P. Self-Interaction Correction to Density-Functional Approximations for Many-Electron Systems / J. P. Perdew, A. Zunger // Physical Review B. – 1981. – Vol. 23, № 10. – P. 5048.

[20] García, A. Siesta: Recent developments and applications / A. García, N. Papior, A. Akhtar [et al.] // Journal of Chemical Physics. – 2020. – Vol. 152, № 20. – P. 204108.

[21] Рейтинг 50 наиболее мощных суперкомпьютеров России // TOP50. – URL: https://top50.supercomputers.ru/list (дата обращения: 27.03.2026).

[22] Tate, J. PH 575 Introduction to Solid State Physics / J. Tate // Oregon State University. – 2019. – URL: http://physics.oregonstate.edu/~tatej/COURSES/ph575 (дата обращения: 27.03.2026).

Авторский вклад

Баглов Алексей Викторович – постановка задачи исследования, анализ, написание, тестирование и сопровождение кода, написание статьи.

Хорошко Людмила Сергеевна – постановка задачи исследования, обсуждение, написание статьи.

ANALYSIS AND REFACTORING OF THE LIBRARY OF EXCHANGE-CORRELATION FUNCTIONS OF THE OPENMX PACKAGE

A.V. Baglov

Senior researcher in the R&D Lab of Energy-effective materials and technologies, Faculty of Physics, BSU; Researcher in the Center 4.11 of R&D Department, BSUIR

L.S. Khoroshko

Lead researcher in the R&D Lab of Energy-effective materials and technologies, Faculty of Physics, BSU; Lead Researcher in the Center 4.11 of R&D Department, BSUIR

Abstract. The exchange-correlation functionality of the OpenMX package was refactored to eliminate the original's shortcomings and optimize performance. The resulting cross-platform library, written in ANSI C, delivers a 4x speed boost on Linux systems and reduces source code size by 2.5x. This work paves the way for the integration of additional modern functionality required for theoretical solid-state physics, nanomaterials science, and the development of materials science AI models using the OpenMX package. The open license of OpenMX and the self-sufficiency of the developed library facilitate its integration into the international scientific community and its use in education.

Keywords: modeling, ab initio, OpenMX, density functional theory, programming, ANSI, C.

УДК 004.75:004.62

АРХИТЕКТУРА ПОТОКОВОЙ ОБРАБОТКИ И ХРАНЕНИЯ ДАННЫХ БИОГАЗОВЫХ УСТАНОВОК НА ОСНОВЕ ТЕХНОЛОГИЙ BIG DATA



А.А. Исмаилова
И.о. профессора ГОП «Информационные системы», Академик Международной академии информатизации, Доктор философии (PhD), ассоциированный профессор (доцент)
a.a.ismailova@kazatu.kz



Р.Н. Тажиева
Докторант 1 курса образовательной программы 8D06101 «Аналитика больших данных»,
Магистр Информационных систем
024.risti.1978@gmail.com

А.А.Исмаилова

Окончила Кызылординский педагогический институт им. Қоркыт Ата, квалификация: учитель физики, информатики, вычислительной техники и программного обеспечения. Научное направление: Анализ, моделирование и проектирование информационной системы (по отраслям), интеллектуальный анализ данных

Р.Н.Тажиева

Окончила Таразский государственный университет имени М.Х.Дулати, квалификация информатик, преподаватель информатики. Научное направление: Анализ, моделирование и проектирование информационной системы (по отраслям), интеллектуальный анализ данных

Аннотация: Построена многоуровневая архитектура системы сбора, потоковой обработки и хранения данных, поступающих от сенсорной сети биогазовой установки. Обоснован выбор между «длинным» и «широким» форматами хранения данных в зависимости от требуемой скорости аналитических данных. Цель работы разработка архитектуры сбора и потока данных на основе Big Data для оптимизации процесса анаэробного сбраживания и повышения выхода метана. Научная новизна заключается в интеграции Big Data к специфике биогазовых процессов, который позволяет интегрировать новые типы датчиков без изменения схемы базы данных.

Ключевые слова: биогазовая установка, Big Data, потоковая обработка данных, IoT, анаэробное сбраживание, архитектура данных, хранилище данных.

Введение. Производство альтернативного источника энергии на основе возобновляемого сырья рассматривается Правительством Республики Казахстана как прорывное направление, позволяющее вывести сельскохозяйственное производство на новый технологический уровень, обеспечивающий рост его эффективности, увеличить экспорт продукции с высокой добавленной стоимостью. Во всех странах в основном содержат скот на молочно-товарных фермах в загонах, отходы, которых загрязняют окружающую среду болезнетворными бактериями, яиц гельминтов и сорными травами земли населенных пунктов, что требуют их утилизации и переработки в биоудобрение и источник энергии [1].

Исследования зарубежных ученых и практиков, в основном направлены на создание биогазовых установок, которых отличает большое количество ежедневного потребления отходов и растительной биомассы, что увеличивает транспортные и иные расходы на сбор, хранение и подготовку, доступное не каждому фермеру [2]. В России, в Казахстане и стран

СНГ из-за высокой начальной стоимости биогазовых установок их общее количество не превышает нескольких сотен. Анализ результатов обследования работы этих установок показывает, что требуются серьезные конструктивные доработки по всей цепочке технологического процесса, обеспечение простоты контроля, доступа к обслуживанию [3].

Функционирование биогазовых установок связано с рядом специфических трудностей [4].

Оптимальный выход биогаза ограничивается нестабильностью процесса, неопределённостью и сложностью контроля параметров анаэробного сбраживания, а также необходимостью мониторинга в режиме реального времени - всё перечисленное относится к числу основных проблем, приводящих к снижению эффективности производства [5].

Традиционные стратегии управления, основанные на ограниченном наборе сенсоров, не обладают достаточной прогностической способностью для своевременного выявления подобных возмущений [6].

В связи с этим оптимизация процессов анаэробного сбраживания требует применения новых подходов, сочетающих мониторинг с высоким пространственно-временным разрешением и передовые аналитические методы, и инструменты поддержки принятия решений.

Современные биогазовые установки представляют собой биогазовые реакторы, эффективность которых в полной мере зависит от поддержания параметров среды (температура, pH, состав субстрата и т.д.).

Интеграция технологий Big Data в управление биогазовыми установками позволяет не только отслеживать текущее состояние биореактора, но предсказывать падение выхода биогаза.

Первым шагом для внедрения аналитики Big Data является построение корректной архитектуры потоков данных.

Архитектура сбора данных с датчиков биогазовой установки.

Архитектура сбора данных – это нижний уровень, который отвечает за физическое получение сигнала от датчиков и его первичную оцифровку. Прежде чем данные окажутся в базе, им предстоит пройти путь от датчика до программного хранилища. Имеется два варианта пути: классический и ориентированный на IoT-инфраструктуру.

Классический вариант подходит для производственного объекта. Датчики подключаются к ПЛК через стандартные промышленные интерфейсы, ПЛК принимает сигналы, обрабатывает их и хранит значения в регистрах, откуда их забирает SCADA или шлюз по протоколам OPC UA или Modbus TCP. По второму IoT-ориентированному варианту умные датчики сами отправляют измерения по проводным или беспроводным протоколам (LoRaWAN, MQTT) прямо на облачный сервер или шлюз (рисунок 1).

Протоколы передачи данных. Работа всей системы зависит от протокола передачи данных. Для мониторинга биогазовой установки можно использовать три основных варианта:

- Modbus TCP – промышленный протокол регистров ПЛК, который применяется для считывания показаний датчиков температуры, давления, оборотов мешалки непосредственно из регистров контроллера:

- OPC UA – современный стандарт промышленной коммуникации, который обеспечивает семантическое описание данных, кросс-платформенное взаимодействие и информационную безопасность:

- MQTT – легкий протокол отимизированный для IoT – устройств с ограниченными ресурсами.



Рисунок 1. Основная схема архитектуры сбора данных

На начале схемы расположены биореактор с установленными сенсорами (датчиками) данных, которые измеряют ключевые технологические параметры, такие как температура реактора, pH среды, уровень сырья, давление газа, концентрация метана и т.д. Датчики формируют поток данных с заданной оператором периодичностью и передают его на устройства сбора. Данные с датчиков поступают в программируемые логические контроллеры. Контроллеры выполняют функции: опрос датчиков с заданной частотой, первичная фильтрация и усреднение данных, преобразование аналоговых сигналов на цифровые сигналы. Для подключения сенсоров к контроллеру в зависимости от типа датчиков.

Архитектура потока данных

Для обеспечения возможности и масштабируемости обработки больших объемов данных предлагается трехуровневая архитектура: уровень восприятия, сетевой уровень и прикладной уровень. Уровень восприятия включает сеть первичных преобразователей (датчиков), установленных непосредственно на узлах подготовки сырья, в ферментаторе и газовых линиях. Локальные контроллеры PLC, осуществляющую первичную оцифровку тоже находятся в этих местах. Сетевой уровень служит связующим звеном между протоколами передачи и шлюзами. Использование протоколов передачи (LoRaWAN, MQTT) позволяет передавать данные на шлюзы (gateways). На этом же этапе реализуется предварительная фильтрация шумов и агрегация данных перед отправкой в центральное хранилище или облако. Данная концепция называется Edge Computing. Прикладной уровень состоит из среды хранения, как MongoDB для NoSQL данных и аналитических инструментов Apache Kafka, Spark. Здесь выполняются алгоритмы Advanced Analytics для оптимизации процессов (рисунок 2).

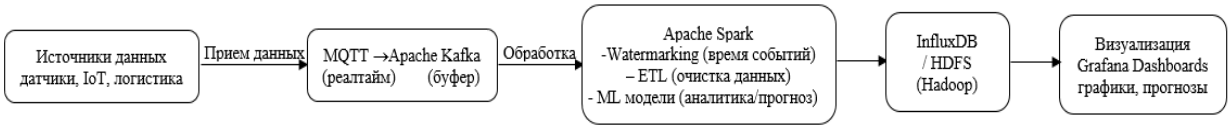


Рисунок 2. Архитектура потока данных

Сенсорные данные с датчиков биогазовой установки в реальном времени передаются через MQTT-брокер, который обеспечивает легкую и энергоэффективную передачу сообщений от датчиков. Центральный элемент принимает данные от всех устройств, маршрутизирует сообщения и обеспечивает асинхронную передачу. Для повышения масштабируемости и надежности данные поступают в Apache Kafka, где формируется очередь сообщений, сохраняется необработанный поток данных и обеспечивается отказоустойчивость системы. Обработка данных выполняется с использованием Apache

Spark в режиме потоковой обработки. На данном этапе реализуется синхронизация событий (watermarking) для корректной работы с задержанными данными и ETL процессы, как очистка, фильтрация и нормализация данных [7]. Для прогнозирования и выявления аномалий применяются модели машинного обучения. Для хранения данных используется комбинированный подход: InfluxDB - для хранения оперативных временных рядов и HDFS (Hadoop) - для долговременного хранения исторических данных. InfluxDB обеспечивает быстрый доступ к актуальным данным, а HDFS (Hadoop) используется для анализа, построения моделей и отчетности. Для визуализации и мониторинга данных используется система визуализации Grafana, которая обеспечивает построение графиков и панелей мониторинга, отобржение ключевых показателей установки, настройку оповещений и поддержку принятия решений оператором.

Источники, частота и формат данных

Технология метанового сбраживания позволяет оптимизировать основные параметры технологического режима (водородный показатель, степень измельчения, температура, давление, перемешивание и подача сырья в биореактор (BR)) на всех стадиях процесса анаэробного сбраживания, что обеспечивает максимизацию выхода биогаза и сокращение времени ферментации субстрата в реакторе [8]. К сожалению, в настоящее время непосредственно в режиме реального времени поддается измерению ограниченное число факторов: температура, pH, ОВП, перемешивание в биореакторе, расход и состав получаемого биогаза и расход подаваемого раствора. Некоторые переменные определяются с помощью аналитического оборудования (например, концентрация VFA с помощью газового хроматографа), а остальные с помощью ручного анализа (например, концентрация органических веществ) [9].

На основе анализа литературы [8,9] выделены основные группы параметров: газовый состав, физические параметры и химические показатели.

Газовый состав: содержание метана (CH_4), углекислого газа (CO_2) и сероводорода (H_2S). Примеры анализаторов: MQ-4, MG-811 и промышленные ИК-анализаторы.

Физические параметры: температура субстрата, давление в газгольдере, уровень заполнения биореактора [10].

Химические показатели: pH-уровень (критичен для предотвращения гибели метаногенов), щелочность (TAC), содержание летучих жирных кислот (FOS/VFA), соотношение C/N, ОВП (окислительно-восстановительный потенциал).

Частота и формат данных

Для задач Big Data важен баланс между детализацией и избыточностью данных. Рекомендуемые параметры представлены в таблице 1.

Таблица 1. Категория и частота данных

Категория данных	Источник	Рекомендуемая частота	Формат передачи
Динамические данные (газ/давление)	Газоанализаторы	1 раз в 1-5 минут	MQTT (JSON payload)
Инерционные (температура)	Термопары, термистры	1 раз в 10 - 15 минут	MQTT (Modbus TCP)
Химические (pH/ОВП)	pH-электроды	1 раз в 15 - 30 минут	JSON/CSV (пакетно)
Энергетические (выработка)	Электросчетчики	1 раз в час	Modbus/ JSON

Формат данных: Наиболее предпочтительным для BiG Data систем является JSON, благодаря его гибкости при хранении в документированных БД. От организации документа

в коллекции зависит и скорость запросов, и то, насколько легко будет добавить новый датчик через месяц. В практике IoT-систем выделяют два принципиально различных подхода: длинный формат (один замер на документ) и широкий формат (снимок всех датчиков на документ). Выбор между подходами является критически важным архитектурным решением для Big Data в IoT [11].

Суть проста: один документ - одно показание одного датчика. Неважно, сколько датчиков в системе - структура записи остаётся одинаковой:

```
{
  "ts": ISODate("2026-02-17T10:15:30Z"),
  "deviceId": "biogas-reactor-01",
  "sensor": "temperature",
  "value": 37.4,
  "unit": "C",
  "quality": "ok"
}
```

У такого подхода есть очевидные плюсы: добавить новый датчик можно без каких-либо изменений в схеме; агрегировать данные по времени или по типу датчика удобно через единый pipeline; масштабирование при росте числа источников данных не требует миграции схемы.

Альтернатива - собирать в один документ показания сразу всех датчиков за один момент времени. Удобно, когда нужно быстро посмотреть, что происходит прямо сейчас:

```
{
  "ts": ISODate("2026-02-17T10:15:30Z"),
  "deviceId": "biogas-reactor-01",
  "temperature_c": 37.4,
  "pressure_kpa": 102.3,
  "mixer_rpm": 45,
  "gas_flow_m3h": 2.1
}
```

Проблема возникает, если датчики опрашиваются с разной частотой - часть полей будет просто пустой. А если понадобится добавить новый параметр, придётся переписывать все старые документы. Для систем с разнородными датчиками длинный формат надёжнее. Для ясности и сравнения в Таблице 2 приведены основные характеристики двух подходов.

Таблица 2. Сравнение двух подходов формата данных

Характистика	Длинный формат	Широкий формат
Структура	{sensor: "temp", value: 37, time: T}	{time: T, temp: 37, pH: 7.2, CH4: 60}
Гибкость	Высокая	Ниже
Скорость записи	Быстрее при редких замерах	Эффективнее при одновременном снятии всех показаний
Аналитика	Удобно для поиска по конкретному датчику	Удобно для анализа корреляций

Хранилище данных

Создание стандартизированных протоколов тестирования и открытых баз данных для оценки датчиков способствовало бы более объективной оценке и интеграции с моделями [12]. Вопрос о том, когда целесообразно применять Big Data-платформы - Apache Hadoop, Apache Spark, Flink - вместо традиционного хранилища данных, не имеет однозначного ответа. Он требует анализа сразу четырёх измерений: объёма данных, скорости их поступления, структурной неоднородности и требований к аналитике.

Тем не менее, отвечая на исходный вопрос напрямую: когда именно Big Data-платформы оправданы вместо классического DW - вот четыре сигнала [13]:

– Данных больше 50–100 ТБ и они продолжают расти - вертикальное масштабирование обходится слишком дорого.

– Нужна реакция на данные в режиме реального времени - задержка в часы или даже минуты недопустима.

– Большая часть данных неструктурирована: тексты, логи, JSON, изображения - и их схема заранее неизвестна.

– Аналитика выходит за рамки SQL: машинное обучение, обработка текстов, работа с изображениями - классический SQL тут не поможет. (MLSQL)

Если ни одно из этих четырёх условий не выполняется - классический DW остаётся правильным выбором [14]. Он проще, предсказуемее, лучше интегрируется с BI-инструментами и дешевле в поддержке. Поэтапное название процессов от датчиков до базы данных приведен в Таблице 3.

Таблица 3. Поэтапный план внедрения базы данных

Этап	Наименование	Состав работ	Инструментарий
1	Инвентаризация тегов	Составление реестра датчиков: имя тега, ед. изм., диапазон, частота	Excel / YAML-файл
2	Выбор протокола	Настройка Modbus TCP / OPC UA для чтения регистров ПЛК	Node-RED Modbus-узел
3	Настройка сборщика	Поток: «опрос ПЛК → формирование документа → публикация в MQTT»	Node-RED, Mosquitto
4	Создание коллекции	Создание Time Series коллекции, индексов, TTL-политики	MongoDB Shell / Compass
5	Запись данных	Подписка на MQTT, запись в sensor_readings через драйвер	Node-RED MongoDB-узел
6	Валидация запросов	График температуры за сутки; расчёт средних за 10 мин	MongoDB Compass, Python
7	Визуализация	Подключение Grafana к MongoDB, настройка дашборда	Grafana + MongoDB плагин

Мониторинг процессов анаэробного сбраживания зависит от точности сенсоров, однако, используемые в настоящее время устройства, имеют ряд эксплуатационных недостатков. Электрохимические сенсоры, такие как pH-электроды, широко применяются, но требуют частой калибровки и очистки поверхности для поддержания точности. На базе Института информационных и вычислительных технологий, построена биогазовая установка, включающая 8 полноценных биореакторов объемом 30 кубометров, объединенных в единую технологическую цепочку. Разработанная в нем интеллектуальная система управления качеством биогаза представляет собой интегрированное решение для автоматизированного мониторинга и контроля параметров производства биогаза. Система включает в себя следующие ключевые функции: сбор данных о параметрах качества биогаза; передача данных через Интернет с использованием беспроводных сетей; обработка поступающих данных с датчиков в режиме реального времени; запись и хранение информации в облачной базе данных; отображение параметров системы через веб-интерфейс для удаленного пользователя. Такая система обеспечивает не только мониторинг, но и интеллектуальное управление технологическим процессом, повышая стабильность, безопасность и эффективность работы биогазовой установки [15].

Описанные примеры демонстрируют потенциал использования IoT в системах мониторинга, а также указывают на важность дальнейшего развития в направлении интеллектуализации таких решений. Эти разработки могут быть адаптированы для создания системы управления качеством биогаза, в которой датчики будут отслеживать параметры.

Заключение. Формирование четкой архитектуры и стандартизация сенсорных данных – необходимое условие для перехода к «Индустрии 4.0» в биоэнергетике. Использование протокола MQTT и формата обеспечивает легкую интеграцию с современными аналитическими платформами, что позволяет оптимизировать IT-решения для управления биогазовыми установками в режиме реального времени.

Виртуальные сенсоры, интегрирующие электрохимические, микробиологические, оптические и гибридные системы, обеспечивают адаптивное управление процессом в реальном времени и мониторинг стабильности анаэробного сбраживания. Дальнейшее развитие интеллектуальных систем анаэробного сбраживания требует создания более надёжных сенсоров, стандартизации открытых наборов данных и повышения интерпретируемости моделей искусственного интеллекта.

В сочетании с новыми сенсорными технологиями и концепциями мягких сенсоров ИИ имеет потенциал сделать АД более предсказуемым, устойчивым и ресурсосберегающим, что соответствует принципам циркулярной экономики и текущему энергетическому переходу.

Список литературы

[1] Karayeva, Yu. V. Review of biogas technologies and intensification methods processes of an anaerobic fermentation. *Akademenergo Works*; **2010**, *3*, 109-127.

[2] Tlebayev, M.B., Taszhurekova, Z.K., Biibosunov, B.I., Baizharikova, M.A., Aitbayeva, Z.K. (2020). Creation of a computer-assisted mathematical model for the raw materials biological processing. Periódico Tchê Química, 17(35): 640-654A. https://doi.org/10.1007/s13399-023-03812-x

[3] Тлебаев М.Б., Тажиева Р.Н., Айтбаева З.К., Нуржигитова Ж.Н. Разработка и проектирование трехступенчатой высокоскоростной биогазовой установки. ИЗВЕСТИЯ ВУЗОВ КЫРГЫЗСТАНА№ 5, 2016

[4] Ziganshin, A. M., Schmidt, T., Lv, Z., Liebetrau, J., Richnow, H. H., Kleinsteuber, S., & Nikolausz, M. (2016). Reduction of the hydraulic retention time at constant high organic loading rate to reach the microbial limits of anaerobic digestion in various reactor systems. Bioresource Technology, 217, 62–71.

[5] Ukpai, P. A., & Nnabuchi, M. N. (2012). Comparative study of biogas production from cow dung, cow peel and cassava peeling using 45 litres biogas digester. Advances in Applied Science Research, 3, 1864-1869. http://edn.link/cm2nmf

[6] Cinar, S.; Cinar, S.O.; Wieczorek, N.; Sohoo, I.; Kuchta, K. Integration of Artificial Intelligence into Biogas Plant Operation. Processes 2021, 9, 85 https://doi.org/10.3390/pr9010085

[7] Zaharia M. et al. Apache Spark: A Unified Engine for Big Data Processing // Communications of the ACM. –2016. –Vol. 59, No. 11. –P. 56–65.

[8] Baizharikova, M., Tlebayeva, G., Tlebayev, M., Beksultanov, Z., & Tazhiyeva, R. (2024). Development of a new innovative technological process for continuous methane fermentation in a three stage biogas plant. Biomass Conversion and Biorefinery, 14, 29219. https://doi.org/10.1007/s13399-023-03812-x

[9] Simeonov, I., & Hubenov, V. (2025). Application of Artificial Intelligence for Prediction, Monitoring, Optimization and Control of Anaerobic Digestion Processes–A Review. Processes, 13(12), 3812. https://doi.org/10.3390/pr13123812

[10] Deublein, D., & Steinhauser, A. (2011). Biogas from Waste and Renewable Resources. Publisher: WILEY VCHISBN: 978-3-527-32798-0 https://doi.org/10.1002/9783527632794

[11] Gaida, D., Wolf, Ch., State estimation for anaerobic digesters using the ADM1 July 201266(5):1088-95 https://doi.org/10.2166/wst.2012.286

[12] Marycz, M., Turowska, I., Glazik, S., & Jasiński, P. (2024). Artificial intelligence in anaerobic digestion: A review of sensors, modeling approaches, and optimization strategies. Applied Sciences, 14(4), 1612. https://doi.org/10.3390/app14041612

[13] Armbrust M. et al. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics // CIDR 2021: 11th Annual Conference on Innovative Data Systems Research. - 2021.

[14] Kimball R., Ross M. The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. 3rd ed. - Indianapolis: Wiley, 2013.

[15] Development of an IoT System for Monitoring the Operation of a Biogas Plant https://doi.org/10.20944/preprints202507.0368.v1

Авторский вклад

Исмаилова Айсулу Абжаппаровна – постановка задачи исследования, разработка архитектуры потока и хранения данных, руководство исследованием.

Тажиева Рысты Нарбаевна – постановка задачи исследования, разработка архитектуры потока и хранения данных, руководство исследованием.

STREAM PROCESSING AND DATA STORAGE ARCHITECTURE FOR BIOGAS PLANTS BASED ON BIG DATA TECHNOLOGIES

Ismailova Aisulu
Acting professor of the GEP «Information Systems», Academician of the International academy of informatization PhD, Associate Professor (docent), a.a.ismailova@kazatu.kz

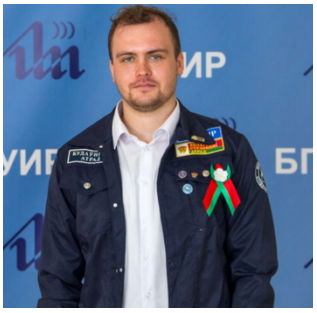
Tazhiyeva Rysty
First-year PhD student in the educational program 8D06101 «Big Data Analytics», Master's in Information Systems, 024.risti.1978@gmail.com.

Abstract: A multi-layered architecture for the collection, stream processing, and storage of data received from a biogas plant sensor network has been constructed. The choice between "long" and "wide" data storage formats is justified depending on the required speed of analytical data processing. The objective of the work is to develop a data collection and streaming architecture based on Big Data to optimize the anaerobic digestion process and increase methane yield. The scientific novelty lies in the integration of Big Data with the specific requirements of biogas processes, which allows for the integration of new sensor types without modifying the database schema.

Keywords: biogas plant, Big Data, stream processing, IoT, anaerobic digestion, data architecture, data warehouse.

УДК 004.75:004.8

ИНТЕЛЛЕКТУАЛЬНОЕ ПРОГНОЗИРОВАНИЕ ЗАГРУЗКИ ОБЛАЧНОЙ ИНФРАСТРУКТУРЫ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ CHRONOS-2 И МЕТОДОВ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ





К.А. Афанасенко
Магистрант факультета информационных технологий и управления специальности «Системы управления информацией» karnei.afanasenka@gmail.com

Н.А. Жиляк
Доцент кафедры информационных технологий автоматизированных систем, кандидат технических наук gznadya@gmail.com

К.А. Афанасенко
Магистрант Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой алгоритмов поддержки принятия решений, разработкой масштабируемых информационно-компьютерных систем и сервисов.

Н.А. Жиляк
Доцент кафедры информационных технологий автоматизированных систем, доцент, кандидат технических наук. Область профессиональных и научных интересов связана с информационными системами и технологиями, веб-программированием, веб-проектированием, веб-дизайном, front-end'ом, операционными системами.

Аннотация. В статье рассматривается возможность применения модели Chronos-2 для прогнозирования загрузки облачной инфраструктуры по временным рядам эксплуатационных метрик, включая загрузку центрального процессора, использование оперативной памяти и сетевую активность. Выполнен обзор существующих подходов к прогнозированию временных рядов в облачных системах, включая статистические, нейросетевые и предобученные модели. Показаны особенности модели Chronos-2 как инструмента одномерного, многомерного и covariate-informed прогнозирования. Рассматриваются возможные направления использования такой модели в задачах управления ресурсами, масштабирования сервисов и планирования нагрузки облачной платформы.

Ключевые слова: Big Data, облачные вычисления, временные ряды, прогнозирование нагрузки, Chronos-2, оптимизация ресурсов, машинное обучение

Введение. Эффективное управление облачной инфраструктурой связано с необходимостью своевременного перераспределения вычислительных ресурсов при изменяющейся нагрузке. Для облачных приложений задача прогнозирования нагрузки рассматривается как важный элемент упреждающего управления ресурсами, поскольку позволяет заранее оценивать будущие изменения спроса на вычислительные мощности, память и сетевые ресурсы [1]. При этом телеметрия облачной платформы формируется в виде временных рядов и включает метрики загрузки CPU, памяти, дисковой подсистемы и сетевого ввода-вывода, что делает задачу прогнозирования непосредственно связанной с методами анализа временных рядов и инструментами Big Data [1].

В последние годы наряду со специализированными моделями для временных рядов получили развитие foundation-модели, обучаемые на больших массивах временных данных. Семейство Chronos основано на языковых архитектурах, в которых временной ряд преобразуется в последовательность токенов, после чего модель обучается по принципу, близкому к обучению языковых моделей [2]. Модель Chronos-2 развивает этот подход и, согласно описанию авторов, поддерживает одномерное, многомерное и covariate-informed прогнозирование в zero-shot режиме, что делает её применимой для задач совместного прогнозирования нескольких взаимосвязанных метрик облачной инфраструктуры [3].

В связи с этим актуальной является задача исследования возможностей модели Chronos-2 для интеллектуального прогнозирования загрузки облачной инфраструктуры. Целью работы является анализ применимости Chronos-2 для предсказания изменения эксплуатационных метрик облачной платформы и рассмотрение возможностей использования такого прогноза в задачах управления ресурсами, автоскейлинга и планирования нагрузки.

Анализ задачи прогнозирования загрузки облачной инфраструктуры. Прогнозирование загрузки облачной инфраструктуры рассматривается как одна из ключевых задач упреждающего управления ресурсами. В обзоре по cloud workload prediction отмечается, что точный прогноз нагрузки важен для proactive resource management облачных приложений. На практике телеметрия инфраструктуры формируется в виде временных рядов и включает показатели загрузки процессора, памяти, дисковой подсистемы и сетевого трафика [1], [4].

Особенность этой задачи состоит в том, что нагрузка облачной платформы изменяется неравномерно. Во временных рядах могут одновременно присутствовать периодические колебания, кратковременные всплески, долговременные тренды и нерегулярные изменения, связанные с поведением пользователей, расписанием вычислительных задач и характером развернутых сервисов. Поэтому задача прогнозирования не сводится к простой экстраполяции предыдущих значений и требует применения моделей, способных учитывать сложную структуру зависимостей во времени.

Практическая значимость прогноза определяется тем, что его результаты могут использоваться при масштабировании сервисов, планировании вычислительных мощностей и перераспределении ресурсов между виртуальными машинами и приложениями. При наличии прогноза система управления инфраструктурой может

заранее реагировать на ожидаемый рост или спад нагрузки, а не только фиксировать уже наступившее изменение состояния.

С точки зрения методов анализа данных задача относится к прогнозированию временных рядов. Для нее применяются статистические, машинные и нейросетевые подходы, однако в последние годы появились универсальные предобученные модели для временных рядов. В работе Chronos временной ряд преобразуется в последовательность токенов, после чего используются подходы, близкие к языковому моделированию [5]. В работе Chronos-2 модель описана как zero-shot предобученная модель для временных рядов для одномерного, многомерного и covariate-informed прогнозирования, что представляет интерес для облачной инфраструктуры, где разные метрики часто взаимосвязаны и должны анализироваться совместно.

Таким образом, задача прогнозирования загрузки облачной инфраструктуры является практически значимой задачей анализа временных рядов, связанной с обработкой потоков эксплуатационной телеметрии и поддержкой решений по управлению ресурсами. В рамках данной работы целесообразно рассматривать применение Chronos-2 как современного инструмента прогнозирования, пригодного для анализа нескольких связанных метрик облачной платформы.

Обзор методов прогнозирования временных рядов. Для прогнозирования временных рядов традиционно применяются статистические методы, среди которых наиболее распространены экспоненциальное сглаживание и ARIMA. В учебнике Forecasting: Principles and Practice отмечается, что именно эти два подхода относятся к числу наиболее широко используемых в задачах прогнозирования. При этом модели экспоненциального сглаживания описывают тренд и сезонность, а ARIMA ориентирована на моделирование автокорреляционной структуры ряда [6]. Для задач с относительно устойчивой динамикой такие методы остаются полезными, однако при сложных нелинейных зависимостях и большом числе взаимосвязанных признаков их возможности оказываются ограниченными.

С развитием методов глубокого обучения в задачах прогнозирования временных рядов широкое распространение получили рекуррентные нейросетевые архитектуры. Модель LSTM была предложена как способ обучения на последовательностях с длинными временными зависимостями за счет механизма поддержания постоянного потока ошибки во времени [7]. В дальнейшем появились и более специализированные архитектуры. Например, в работе N-BEATS представлена глубокая нейросетевая модель для точечного прогнозирования одномерных временных рядов, ориентированная на универсальность применения и интерпретируемость [8]. Эти методы расширили возможности прогнозирования по сравнению с классическими статистическими моделями, однако в большинстве случаев оставались ориентированными либо на конкретный тип временного ряда, либо на отдельный класс задач.

Следующий этап развития связан с использованием Transformer-подходов. В работе Autoformer предложена архитектура, сочетающая декомпозицию временного ряда на составляющие и механизм Auto-Correlation для выявления зависимостей на длинных интервалах [9]. В модели PatchTST временной ряд разбивается на фрагменты, используемые как входные токены Transformer, что позволяет увеличить длину анализируемого контекста и повысить качество долгосрочного прогнозирования многомерных рядов [10]. Эти модели показали, что механизмы внимания могут быть эффективно адаптированы к временным рядам, особенно в задачах долгосрочного прогноза.

В последние годы внимание сместилось к foundation-моделям для временных рядов. В работе Chronos предложен подход, в котором временной ряд преобразуется в последовательность токенов, после чего используется архитектура, близкая к языковым моделям. Модель Chronos-2 развивает этот подход и описывается авторами как zero-shot

универсальная предобученная модель для временных рядов для одномерного, многомерного и covariate-informed прогнозирования. Для задачи прогнозирования загрузки облачной инфраструктуры это особенно важно, поскольку эксплуатационные метрики, такие как загрузка CPU, использование памяти и сетевой трафик, часто взаимосвязаны и должны анализироваться совместно. Поэтому в рамках данной работы именно модели класса Chronos представляют наибольший интерес как современные универсальные средства прогнозирования временных рядов .

Применение модели Chronos-2 для прогнозирования нагрузки облачной инфраструктуры. Модель Chronos была предложена как foundation-подход к прогнозированию временных рядов, в котором значения ряда предварительно масштабируются, квантуются и преобразуются в последовательность токенов, после чего для обучения используются архитектуры, основанные на Transformer-моделях языкового типа. Такой подход позволяет рассматривать прогнозирование временных рядов как задачу, близкую к последовательностному моделированию, и использовать преимущества предобученных моделей при работе с новыми рядами без построения отдельной специализированной архитектуры под каждую прикладную задачу [2].

В работе Chronos-2 этот подход был расширен от одномерного прогнозирования к более универсальному сценарию. Авторы описывают Chronos-2 как предобученную zero-shot модель, поддерживающую одномерное, многомерное и covariate-informed прогнозирование. Это особенно важно для задач облачной инфраструктуры, поскольку эксплуатационная нагрузка обычно не сводится к одному показателю: загрузка CPU, использование памяти, сетевой трафик и другие метрики изменяются совместно и могут оказывать влияние друг на друга. Поэтому для данной предметной области более уместно рассматривать не изолированный прогноз одной метрики, а совместный анализ нескольких связанных временных рядов.

Chronos-2 использует механизм group attention, который поддерживает обмен информацией между временными рядами внутри одной группы, включая связанные ряды, разные переменные многомерного ряда, а также целевые признаки и ковариаты. Для облачной платформы это означает возможность учитывать не только историю самой прогнозируемой метрики, но и дополнительные признаки, связанные с поведением системы. В прикладной постановке это может соответствовать совместному использованию рядов CPU, RAM, Network I/O и, при необходимости, внешних факторов, если они доступны в данных мониторинга.

Еще одной важной особенностью Chronos-2 является ориентация на zero-shot использование, то есть получение прогноза без обязательного дообучения под каждую новую задачу. Для облачной аналитики это представляет практический интерес, поскольку позволяет применять модель в условиях, где инфраструктурные конфигурации, профили сервисов и состав наблюдаемых метрик могут меняться. В таком случае модель может рассматриваться как универсальный инструмент предварительного прогноза, который затем используется в контуре принятия решений по масштабированию и распределению ресурсов.

Подход к формированию входных данных и прогнозируемых метрик. В задаче прогнозирования загрузки облачной инфраструктуры входные данные целесообразно формировать на основе эксплуатационной телеметрии, поступающей от виртуальных машин, контейнеров и сервисов мониторинга. К числу основных метрик обычно относят загрузку центрального процессора, использование оперативной памяти, показатели дисковой подсистемы и сетевой трафик. Именно такие данные рассматриваются как базовые показатели производительности облачной среды и используются при анализе состояния инфраструктуры.

В рамках предлагаемого подхода в качестве прогнозируемых величин можно рассматривать как отдельную метрику, например загрузку CPU, так и совокупность

взаимосвязанных показателей. Второй вариант представляется более содержательным для облачной инфраструктуры, поскольку изменение нагрузки на процессор, память и сеть в реальных системах часто происходит не изолированно, а совместно. В этом случае модель получает не один временной ряд, а набор синхронизированных рядов, описывающих состояние платформы на каждом временном шаге. Это соответствует заявленным возможностям Chronos-2 для многомерного и covariate-informed прогнозирования.

Для использования таких данных необходимо привести их к единой временной сетке. Это означает, что значения всех метрик должны быть согласованы по шагу наблюдения, например по минутам, пяти минутам или часам. После этого данные могут быть агрегированы, очищены от пропусков и явных выбросов, а затем представлены в виде последовательностей фиксированной длины.

Входной фрагмент должен содержать историю изменения метрик за некоторый предшествующий интервал, а выходом модели должен становиться прогноз на ближайший горизонт времени. Такой способ представления позволяет использовать временной контекст и подготавливает данные к подаче в модель временных рядов.

Практическое применение прогнозирования в задачах управления облачной инфраструктурой. Прогнозирование нагрузки представляет практический интерес для систем управления облачной инфраструктурой, поскольку телеметрические данные используются при оценке производительности и планировании ресурсов, а в обзоре по cloud workload prediction сама задача рассматривается как часть proactive resource management. Для облачной платформы это означает, что прогнозируемые значения эксплуатационных метрик могут применяться не только для анализа текущего состояния, но и для подготовки инфраструктуры к ожидаемому изменению нагрузки.

В прикладной постановке прогноз может использоваться в нескольких направлениях. Во-первых, он может служить основой для предварительного масштабирования вычислительных ресурсов при ожидаемом росте нагрузки. Во-вторых, он может использоваться при перераспределении ресурсов между виртуальными машинами, контейнерами или сервисами, если по ряду метрик наблюдается устойчивое приближение к перегрузке. В-третьих, прогноз может быть полезен при планировании снижения резервирования в периоды ожидаемого спада активности.

Такая логика соответствует общему подходу к упреждающему управлению ресурсами, описываемому в работах по прогнозированию нагрузки облачных систем.

Для данной задачи особый интерес представляет возможность совместного анализа нескольких метрик. Chronos-2 описывается как zero-shot модель для одномерного, многомерного и covariate-informed прогнозирования, а также использует механизм group attention для обмена информацией между связанными временными рядами внутри группы. Поэтому в облачной инфраструктуре модель может рассматриваться как средство совместного прогноза CPU, памяти, сетевой активности и других взаимосвязанных показателей, а не только одного изолированного ряда. Это делает подход более содержательным для задач управления платформой, где решение обычно принимается не по одной метрике, а по совокупности признаков состояния системы.

Дополнительное прикладное значение связано с zero-shot характером модели Chronos-2. Авторы позиционируют ее как универсальную предобученную модель, пригодную для использования без обязательного дообучения под каждую новую задачу прогнозирования [3]. Для облачной среды это может быть полезно в случаях, когда инфраструктурная конфигурация изменяется, появляются новые сервисы или набор доступных метрик отличается между площадками.

В такой постановке модель может использоваться как общий инструмент прогностической аналитики, который встраивается в контур мониторинга и поддержки решений по управлению ресурсами.

Заключение. В работе рассмотрена задача интеллектуального прогнозирования загрузки облачной инфраструктуры на основе анализа временных рядов эксплуатационных метрик.

Рассмотрено, что телеметрия облачной платформы формируется в виде последовательностей значений загрузки процессора, памяти, дисковой подсистемы и сетевой активности, а потому задача управления ресурсами тесно связана с методами прогнозирования временных рядов.

Проведённый обзор показывает, что наряду с традиционными статистическими и нейросетевыми подходами существенный интерес представляют предобученные модели для временных рядов.

Модели семейства Chronos основаны на представлении временного ряда в виде последовательности токенов, а Chronos-2 расширяет этот подход до одномерного, многомерного и covariate-informed прогнозирования в zero-shot постановке.

Для задач облачной инфраструктуры это особенно важно, поскольку эксплуатационные метрики обычно взаимосвязаны и должны рассматриваться совместно.

На основании рассмотренных материалов можно отметить, что модель Chronos-2 представляет методический интерес для задач упреждающего управления облачными ресурсами.

Такая модель может рассматриваться как перспективная основа для прогнозирования изменения нагрузки и последующего применения прогноза в задачах масштабирования, перераспределения ресурсов и планирования эксплуатации облачной платформы.

Перспективы дальнейшей работы связаны с уточнением состава входных метрик, формированием согласованных многомерных рядов облачной телеметрии и последующей оценкой применимости модели Chronos-2 в конкретных сценариях управления инфраструктурой.

Список литературы

[1] Feng B., et al. Application-Oriented Cloud Workload Prediction: A Survey. 2025. DOI: 10.26599/TST.2024.9010024.

[2] Ansari A.F., et al. Chronos: Learning the Language of Time Series. Amazon Science. 2024. DOI: 10.48550/arXiv.2510.15821.

[3] Ansari A.F., et al. Chronos-2: From Univariate to Universal Forecasting. arXiv. 2025. DOI: 10.48550/arXiv.2510.15821.

[4] Architecture strategies for collecting performance data. Microsoft Learn.URL: https://learn.microsoft.com/en-us/azure/well-architected/performance-efficiency/collect-performance-data.

[5] Adapting language model architectures for time series forecasting. Amazon Science Blog. URL: https://www.amazon.science/blog/adapting-language-model-architectures-for-time-series-forecasting.

[6] Hyndman R.J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne: OTexts; 2018.

[7] Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9(8):1735–1780. DOI: 10.1162/neco.1997.9.8.1735.

[8] Oreshkin B.N., Carpov D., Chapados N., Bengio Y. N-BEATS: Neural Basis Expansion Analysis for Interpretable Time Series Forecasting. ICLR. 2020. DOI: 10.48550/arXiv.1905.10437.

[9] Wu H., Xu J., Wang J., Long M. Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting. NeurIPS. 2021. DOI: 10.48550/arXiv.2106.13008.

[10] Nie Y., Nguyen N.H., Sinthong P., Kalagnanam J. A Time Series is Worth 64 Words: Long-term Forecasting with Transformers. ICLR. 2023. DOI: https://doi.org/10.48550/arXiv.2211.14730.

Авторский вклад

Афанасенко Корней Александрович – постановка задачи исследования, руководство исследованием, анализ существующих решений, изучение возможных проектных решений.

Жиляк Надежда Александровна – научное руководство исследованием, формирование направления работы и постановки задач, критический анализ содержания работы и редактирование научного текста

INTELLIGENT FORECASTING OF CLOUD INFRASTRUCTURE LOAD USING THE CHRONOS-2 MODEL AND TIME SERIES ANALYSIS METHODS

K.A. Afanasenko
Master's student in the Faculty of Information Technologies and Control, majoring in "Information Management Systems"

N.A. Zhilyak
Associate Professor, Department of Information Technologies for Automated Systems, Candidate of Technical Sciences

Abstract. This article discusses the application of the Chronos-2 model for predicting the dynamics of computing resource utilization, including CPU load, RAM usage, and network activity. The use of the Chronos-2 model is proposed for predicting the dynamics of computing resource utilization, including CPU load, RAM usage, and network activity. An analysis of existing approaches to time series forecasting in cloud systems, including statistical and neural network methods, is conducted. It is described that the use of the Chronos-2 model allows for improved forecasting accuracy compared to classical models and can also be used to optimize resource allocation and enhance the operational efficiency of cloud platforms.

Keywords: Big Data, cloud computing, time series, load forecasting, Chronos-2, resource optimization, machine learning.

УДК 330.47

МЕТОДИКА ВЫЯВЛЕНИЯ СТРУКТУРНЫХ НЕДОСТАТКОВ ЛОГИСТИЧЕСКИХ СЕТЕЙ НА ОСНОВЕ АНАЛИЗА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ





А.И. Лёвина
Профессор Высшей школы бизнес-инжиниринга Института промышленного менеджмента, экономики и торговли ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого», доктор экономических наук, доцент
alyovina@gmail.com

А.С. Дубгорн
Доцент Высшей школы бизнес-инжиниринга Института промышленного менеджмента, экономики и торговли ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого», кандидат экономических наук
alissa.dubgorn@gmail.com

А.И. Лёвина

Окончила Санкт-Петербургский политехнический университет Петра Великого. В 2008 г. защитила диссертацию на соискание степени к.э.н., в 2020 г. – на соискание степени д.э.н. в том же университете. Область научных интересов связана с проектированием комплексных архитектурных решений по автоматизации, цифровизации, роботизации бизнеса, включающим системы управления на основе данных, комплексов математических моделей.

А.С. Дубгорн

Окончила Санкт-Петербургский государственный университет в 2009. В 2020 г. защитила диссертацию на соискание степени к.э.н. в Санкт-Петербургском политехническом университете Петра Великого. Область научных интересов связана с управлением ИТ-сервисами, проектированием архитектуры предприятий различных отраслей, разработкой систем поддержки принятия решений на основе данных.

Аннотация. Современные региональные логистические системы характеризуются высокой сложностью и динамичностью, при этом доступные данные об их функционировании используются недостаточно эффективно для обоснования решений по развитию инфраструктуры. Традиционные методы проектирования опираются на заданные параметры потоков, что ограничивает их применимость в условиях эндогенной динамики. Целью работы является разработка data-driven подхода к анализу и развитию логистических сетей на основе интеграции онтологического моделирования, системной динамики и агентно-ориентированного моделирования. Предложена онтология логистической системы, обеспечивающая интеграцию разнородных данных, и гибридная имитационная модель, позволяющая анализировать как макродинамику системы, так и поведение её участников. В качестве результата сформулирован подход к выявлению аномалий функционирования сети (перегрузки, задержки, дисбалансы потоков) на основе анализа данных и их интерпретации как индикаторов структурной недостаточности. Показано, что интеграция методов анализа больших данных и имитационного моделирования позволяет обосновывать направления развития логистической инфраструктуры.

Ключевые слова: логистические системы, развитие логистической инфраструктуры, системная динамика, агентное моделирование, управление на основе данных

Исследование выполнено за счет гранта Российского научного фонда № 23-78-10190, https://rscf.ru/project/23-78-10190/

Введение. Современные региональные логистические системы представляют собой сложные распределённые сети, функционирующие в условиях высокой неопределённости, сезонности и ограниченности инфраструктурных ресурсов. В процессе функционирования региональных логистических сетей генерируются и накапливаются большие массивы разнородных данных: информации о перемещении грузов, загрузке транспортных маршрутов, работе портовой и складской инфраструктуры, а также параметрах спроса со стороны хозяйствующих субъектов и населения. Несмотря на наличие таких данных, их использование для анализа и развития логистических систем остаётся фрагментарным и недостаточно формализованным.

Существующие логистические сети не всегда эффективны: при их проектировании и строительстве пользуются только расчётными данными. В условиях реальной эксплуатации могут возникать такие узкие места, как перегрузка отдельных узлов, формирование очередей транспортных средств, превышение складских мощностей, увеличение времени доставки и пространственная неравномерность потоков. Подобные эффекты отражают несоответствие структуры сети фактическим потокам. Таким образом, параметры функционирования системы являются эндогенными и могут быть выявлены только на основе анализа данных о её реальном поведении.

Традиционные подходы к развитию логистической инфраструктуры опираются на нормативные модели, экспертные оценки или задачи оптимального размещения объектов при заданных параметрах спроса. В условиях доступности больших данных представляется целесообразным переход к data-driven подходам, в рамках которых структура логистической сети выводится из анализа её наблюдаемой динамики. В данной работе предлагается методологический подход, основанный на выявлении аномалий в функционировании логистической системы и интерпретации их как индикаторов структурной недостаточности сети.

Целью исследования является разработка методологии анализа и развития логистических сетей региона на основе интеграции онтологического представления данных и имитационного моделирования. Для достижения поставленной цели решаются следующие задачи: формализация предметной области логистической сети в виде онтологии; интеграция разнородных источников данных; построение имитационной модели, воспроизводящей функционирование системы; выявление аномалий поведения на основе анализа данных; формирование и проверка гипотез о развитии структуры сети.

Методология исследования. Предлагаемая методология основывается на интеграции методов управления данными, онтологического моделирования и

имитационного анализа, что позволяет реализовать data-driven подход к исследованию и развитию логистических систем. Ключевой особенностью подхода является использование больших данных о функционировании сети не только для описания текущего состояния системы, но и для выявления закономерностей её динамики и структурных ограничений.

На 1-ом этапе осуществляется формализация предметной области в виде онтологической модели. Онтология задаёт согласованную структуру данных, определяя классы объектов (транспортные узлы, маршруты, грузовые потоки, инфраструктурные элементы), их атрибуты и взаимосвязи. Это позволяет интегрировать разнородные источники данных – статистические, операционные и геопространственные – в единую семантическую модель. Таким образом, онтология выступает не только средством описания предметной области, но и инструментом обеспечения качества, сопоставимости и интероперабельности данных, необходимых для последующего моделирования.

На 2-ом этапе на основе интегрированных данных строится имитационная модель существующей логистической сети. В работе предлагается комбинированное использование методов системной динамики и агентно-ориентированного моделирования. Системная динамика применяется для анализа макроуровневых закономерностей функционирования системы, включая накопление запасов, формирование задержек и влияние ограничений пропускной способности. Агентное моделирование, в свою очередь, позволяет воспроизводить поведение отдельных участников логистического процесса (транспортных средств, инфраструктурных объектов, грузов) и локализовать структурные проблемы на уровне конкретных взаимодействий.

Важным аспектом является то, что модель калибруется и верифицируется на основе реальных данных, что обеспечивает её адекватность и практическую применимость. В процессе моделирования формируются временные ряды показателей функционирования системы: загрузка маршрутов, длина очередей, уровень использования складских мощностей, время доставки грузов и другие характеристики. Анализ этих данных позволяет выявлять устойчивые отклонения поведения системы, интерпретируемые как аномалии.

На 3-ем этапе осуществляется выявление аномалий (структурных несовершенств логистической сети) с использованием методов анализа данных, включая статистические подходы, анализ временных рядов и методы выявления выбросов. При анализе данных для выявления аномалий уделяется внимание устойчивым, систематически воспроизводимым отклонениям, отражающим структурные ограничения сети. Пространственно-временная локализация таких аномалий позволяет определить места системы, в которых возникают перегрузки или дефицит инфраструктурных мощностей.

На 4-ом этапе на основе выявленных аномалий формируются гипотезы о развитии логистической сети, включая добавление новых узлов, изменение топологии маршрутов или перераспределение потоков. Эти гипотезы проверяются с использованием имитационной модели, что позволяет количественно оценить их влияние на показатели функционирования системы. Таким образом, предлагаемый подход обеспечивает переход от анализа данных к обоснованию управленческих решений в области развития региональной логистической инфраструктуры.

Предлагаемая методология сочетает методы больших данных, онтологического моделирования и имитационного моделирования. Такое сочетание позволяет сформировать новый класс инструментов для исследования сложных экономических систем (к которым относятся и логистические системы), в которых структура определяется динамикой взаимодействия элементов, выявляемой на основе данных.

Онтология логистической системы. Формализация предметной области в виде онтологий становится всё более востребованным инструментом формализации и систематизации предметной области – её основных сущностей и их взаимосвязей. Онтология представляет собой формальное, явное описание концептуализации предметной области [1], Классики этой области отмечают роль онтологий как средства обеспечения

семантической согласованности моделей и данных [2]. Онтологии позволяют перейти от неструктурированного описания системы к формальной модели, задающей составляющие её сущности, их атрибуты и отношения [3]. В прикладном аспекте онтологии выполняют функцию интеграционного слоя, объединяющего разнородные источники данных (в т.ч. статистические, операционные и геопространственные данные), что особенно важно в условиях работы с большими данными.

При управлении на основе данных онтология выступает в качестве каркаса, обеспечивающего интерпретируемость и согласованность информации.

В отличие от традиционных схем хранения данных, ориентированных на конкретные приложения, онтологическая модель задаёт семантическую структуру, в рамках которой могут быть реализованы различные методы анализа – от статистических моделей до имитационного моделирования и методов машинного обучения.

Таким образом, онтология является связующим элементом между данными и математическим аппаратом, позволяя «запускать» различные аналитические и моделирующие процедуры на единой концептуальной основе.

Особую значимость онтологическое моделирование приобретает при анализе логистических систем, характеризующихся высокой степенью гетерогенности данных и сложной структурой взаимодействий.

В таких системах данные о потоках, инфраструктуре, ограничениях и внешней среде поступают из различных источников и имеют различную структуру и уровень детализации. Онтология позволяет унифицировать эти данные, обеспечивая их совместное использование в рамках аналитических и имитационных моделей.

В рамках настоящего исследования разработана онтология логистической однго из приморских регионов Арктической зоны Российской Федерации (АЗРФ), отражающая основные элементы и взаимодействия, определяющие её функционирование (рисунок 1). В онтологии выделены следующие классы верхнего уровня:

- Инфраструктура: Класс включает стационарные элементы системы (порты, транспортные узлы, складские и перевалочные базы). Данные объекты определяют пространственную структуру сети и выступают в качестве точек концентрации и перераспределения потоков.
- Мобильные агенты: Класс включает объекты, непосредственно участвующие в перемещении грузов: суда, речные караваны (баржи) и грузовые партии. Эти элементы являются носителями динамики системы, формируя потоки и определяя загрузку инфраструктуры.
- Связи: класс включает маршруты (морские и речные) и логистические операции, обеспечивающими взаимодействие между элементами системы. Сущности этого класса задают топологию сети и определяют возможные траектории движения потоков.
- Контекст: Класс включает внешние по отношению к системе факторы, влияющие на её функционирование – спрос, навигационные окна и природно-климатические условия. Эти факторы задают ограничения и условия функционирования системы, определяя допустимые режимы её работы

Разработанная онтология формирует основу для последующего построения имитационной модели и анализа данных.

Она обеспечивает согласованное представление всех элементов системы и их взаимосвязей, что позволяет интегрировать разнородные данные и использовать их для анализа динамики логистической сети, выявления аномалий и обоснования направлений её развития.

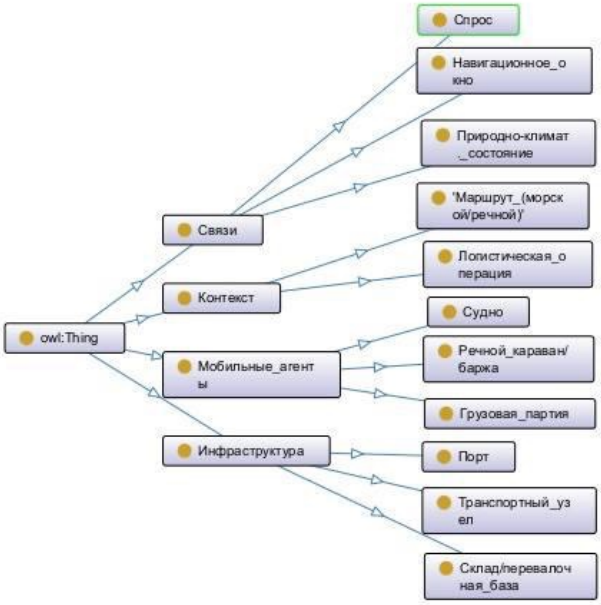


Рисунок 1. Онтология региональной логистической сети АЗРФ

Имитационная модель и работа с данными. Имитационная часть исследования опирается на комбинирование системной динамики (SD) и агентного моделирования (ABM) как взаимодополняющих методов анализа сложных экономических систем. Системная динамика, сформулированная в работах Дж. Форрестера [4] и развитая в классическом учебнике Дж. Стермана [5], предназначена для описания накоплений и потоков, задержек, обратных связей и неустойчивых режимов функционирования системы. Агентное моделирование, как показано в обзорной работе C. Macal и M. North [6], используется для моделирования систем, состоящих из автономных взаимодействующих агентов, и позволяет наблюдать коллективные эффекты, возникающие из локальных правил поведения и микровзаимодействий. Современная литература по гибридному моделированию подчёркивает, что SD и ABM могут и должны сочетаться, а способы их соединения обычно описываются как интерфейсные, интегрированные и последовательные гибридные схемы [7-11].

В рамках рассматриваемой логистической задачи системная динамика используется для анализа макроуровневой эволюции сети. В этой постановке моделируются такие агрегированные переменные, как объём грузов в системе, уровень запасов на складах, очереди в портах, загрузка транспортных маршрутов, время оборота груза и длительность задержек. Уравнения баланса потоков позволяют описывать накопление и перераспределение ресурсов, а также выявлять условия, при которых система переходит в режим устойчивой перегрузки. Иначе говоря, SD отвечает на вопрос, почему логистическая сеть становится неустойчивой: из-за превышения пропускной способности, нарастания очередей, эффекта накопления незавершённого потока или запаздывания реакции на рост спроса. Для арктического узла это особенно важно, поскольку сезонность навигации и ограниченность инфраструктуры усиливают чувствительность сети к небольшим изменениям параметров.

Агентное моделирование, напротив, позволяет детализировать микрорешения участников логистического процесса и локализовать структурные проблемы сети. В модели агентами выступают суда, грузовые партии, портовые и складские объекты, а также, при необходимости, пункты спроса. Их поведение задаётся локальными правилами: выбор маршрута с учётом стоимости и риска, ожидание обслуживания, приоритет разгрузки,

принятие решений о накоплении или выпуске груза, реакция на погодные ограничения и изменение загрузки соседних узлов. Именно ABM позволяет учесть неоднородность участников, конкуренцию за ограниченные ресурсы и формирование нелинейных эффектов в цепочке взаимодействий. В логистике и управлении цепями поставок это особенно ценно, поскольку агентный подход хорошо подходит для анализа динамичных и сложных сетей, где традиционные аналитические схемы требуют слишком жёстких допущений и плохо отражают реальное поведение участников.

Для конкретного арктического логистического узла предлагается связать оба метода в единую гибридную схему. На уровне данных онтология задаёт семантический каркас предметной области и обеспечивает сопоставимость разнородных источников информации; на уровне моделей SD обрабатывает агрегированные ряды и оценивает системные эффекты, а ABM воспроизводит поведение отдельных агентов и выявляет узкие места в потоках. Такая архитектура позволяет использовать реальные данные не только для параметризации модели, но и для её калибровки и верификации. Важнейшими наблюдаемыми характеристиками становятся устойчивые отклонения между фактическим и модельным поведением: длительное превышение очередей, недозагрузка или, напротив, переполнение складов, рост времени доставки, частые перенаправления потоков и появление зон хронической перегрузки. Если такие аномалии воспроизводятся при допустимых значениях параметров и не устраняются локальной перенастройкой правил агентов, они интерпретируются как признаки структурной недостаточности сети и становятся основанием для выдвижения гипотез о её изменении. Современные работы по гибридной симуляции как раз подчёркивают, что именно на стадии концептуального моделирования необходимо формально определить, какие переменные и какие уровни системы должны описываться системной динамикой, а какие – агентным моделированием.

Для исследуемого арктического узла это означает следующее распределение задач моделирования. Системная динамика должна описывать сезонный приток грузов, накопление запасов, загрузку причалов, динамику очередей, дефицит складских ёмкостей и влияние навигационных окон на пропускную способность сети. Агентное моделирование должно воспроизводить индивидуальные решения судов и операторов инфраструктуры, маршрутизацию грузов, поведение складов как буферов, реакцию на погодные и ледовые ограничения, а также перераспределение потоков между узлами. Такое соединение методов позволяет перейти от статического описания сети к data-driven анализу её поведения и, следовательно, к обоснованному выявлению аномалий, которые указывают на необходимость появления новых логистических элементов или изменения характеристик существующих.

Математическая постановка гибридной модели. Рассматривается региональная логистическая сеть как ориентированный граф G=(V,E), где множество вершин V включает инфраструктурные узлы и пункты спроса, а множество рёбер E описывает транспортные связи между ними. Вершины делятся на типы: порты, склады (перевалочные базы), узлы спроса. Рёбра соответствуют морским и речным маршрутам. Модель строится в дискретном времени t=0,1,…,T и объединяет два уровня описания: макроуровень системной динамики и микроуровень агентного поведения.

Вершины делятся на типы: порты, склады (перевалочные базы), узлы спроса. Рёбра соответствуют морским и речным маршрутам. Модель строится в дискретном времени $t = 0,1,\dots,T$ и объединяет два уровня описания: макроуровень системной динамики и микроуровень агентного поведения.

Переменные модели:

Для каждого узла $i \in V$ и маршрута $e \in E$ вводятся следующие переменные:

$S_i(t)$ – запас груза в узле i,

$Q_i(t)$ – очередь на обслуживание в узле i,

$L_e(t)$ – загрузка маршрута e,

$F_e(t)$ – фактический поток по маршруту e,

C_i – пропускная способность узла i,

C_e – пропускная способность маршрута e,

$D_i(t)$ – спрос в точке i,

$\Delta_i(t)$ – дефицит удовлетворения спроса.

Для каждого агента $a \in A$ определяются состояние $x_a(t)$ и правило принятия решения $u_a(t)$. В состав агентов входят суда, грузовые партии, порты, склады и, при необходимости, пункты спроса.

Блок системной динамики

Системная динамика используется для описания агрегированного движения потоков, накопления запасов и образования очередей. Для каждого узла баланс запасов задаётся уравнением:

$$S_i(t+1) = S_i(t) + In_i(t) - Out_i(t),$$

где $In_i(t)$ – входящий поток груза, а $Out_i(t)$ – выходящий поток.

Очередь в узле изменяется по правилу:

$$Q(t+1) = Q_i(t) + Arr_i(t) - Serv_i(t),$$

где $Arr_i(t)$ – число прибывших единиц груза или транспортных средств, а $Serv_i(t)$ – обслуженный объём за шаг времени.

Загрузка маршрута определяется как:

$$L_e(t) = \frac{F_e(t)}{C_e}$$

Если $L_e(t) > 1$, маршрут находится в состоянии перегрузки. Для узлов аналогично вводится коэффициент загрузки:

$$U_i(t) = \frac{S_i(t) + Q_i(t)}{C_i}$$

Для узла спроса дефицит описывается выражением:

$$\Delta_i(t) = D(t) - Out_i(t), \qquad \Delta_i(t) \geq 0$$

Тем самым системная динамика фиксирует устойчивые макроэффекты: накопление запасов, рост очередей, перегрузку маршрутов и дефицит удовлетворения спроса. Эти зависимости отвечают на вопрос, почему система теряет устойчивость.

Блок агентного моделирования

Агентное моделирование используется для описания локального поведения участников логистического процесса и объяснения того, где именно возникает структурная недостаточность сети. Ниже описаны ключевые агенты, выделенные в ABM-модели:

1. Судно:

Каждое судно $v \in A_v$ выбирает маршрут и режим движения, минимизируя индивидуальные издержки:

$$J_v(t) = \alpha_1 T_v(t) + \alpha_2 Cost_v(t) + \alpha_3 Delay_v(t) + \alpha_4 Risk_v(t),$$

где:

$T_v(t)$ – время рейса,

$Cost_v(t)$ – транспортные затраты,

$Delay_v(t)$ – задержки в портах и на маршруте,
$Risk_v(t)$ – риск нарушения навигационных условий.
Ограничения судна:

$$draft_v \le draft_{route}$$
$$iceclass_v \le iceclass_{route}$$
$$t \in [t_{start}^{nav}, t_{end}^{nav}]$$

То есть судно может двигаться только при допустимой осадке, ледовом классе и в пределах навигационного окна.

2. Порт:

Каждый порт $p \in A_p$ обслуживает входящий поток при ограниченных ресурсах. Целевая функция порта задаётся как максимизация обработанного объёма при минимизации очередей:

$$J_p(t) = -\beta_1 Q_p(t) - \beta_2 W_p(t) + \beta_3 H_p(t),$$

где:
$Q_p(t)$ – очередь,
$W_p(t)$ – среднее время ожидания,
$H_p(t)$ – объём обработанного груза.
Ограничения функционирования порта:

$$H_p(t) \le C_p \ , \quad N_p^{sh}(t) \le B_p \ ,$$

где B_p – число одновременно обслуживаемых судов (количество причалов/операционных линий).

3. Склад (перевалочная база):

Каждый склад $w \in A_w$ выполняет буферную функцию. Его поведение описывается целевой функцией минимизации издержек хранения и потерь:

$$J_w(t) = \gamma_1 Loss_w(t) + \gamma_2 OverFl_w(t) + \gamma_3 HC_w(t),$$

где:
$Loss_w(t)$ – потери груза,
$OverFl_w(t)$ – переполнение склада, $OverFl_w(t) = \max(0, S_w(t) - C_w)$,
$HC_w(t)$ – стоимость хранения.
Ограничение:

$$S_w(t) \le C_w$$

4. Грузовая партия как агент

Грузовая партия $c \in A_c$ рассматривается как объект, обладающий требованиями к срокам и маршруту доставки. Её целевая функция:

$$J_c(t) = \delta_1 Del_c(t) + \delta_2 Pen_c(t) + \delta_3 Dam_c(t).$$

Груз не «принимает решения» в классическом смысле, но задаёт давление на систему через дедлайны, приоритеты и штрафы за просрочку.

Связь двух частей модели (системная динамика и агентное моделирование) реализуется через переход от локальных решений агентов к агрегированным потокам и обратно.

Фактический поток по маршруту определяется суммой решений судов:

$$F_e(t) = \sum_{v \in A_v} f_{ve}(t)$$

где $f_{ve}(t) \in \{0,1\}$ показывает, использует ли судно v маршрут e на шаге t.

Совокупный входящий поток в узел:

$$In_i(t) = \sum_{e \in \delta^-(i)} F_e(t)$$

а исходящий поток:

$$Out_i(t) = \sum_{e \in \delta^+(i)} F_e(t)$$

где $\delta^-(i)$ и $\delta^+(i)$ – множества входящих и выходящих рёбер узла i.

Состояния очередей и загрузки используются в правилах выбора агентов. Например, судно выбирает маршрут с минимальным ожидаемым временем:

$$e_v^*(t) = \arg\min\left\{\hat{T}_e(t) + \mu_1 \widehat{W}_{dest(e)}(t) + \mu_2 \hat{R}_e(t)\right\}$$

где:

$\hat{T}_e(t)$ – прогнозное время прохождения,

$\widehat{W}_{dest(e)}(t)$ – ожидаемое ожидание в пункте назначения,

$\hat{R}_e(t)$ – риск нарушения навигации.

Таким образом, системная динамика задаёт контекст состояния сети, а агенты реагируют на него своими локальными решениями. В результате возникает замкнутый контур:

решения агентов – потоки – накопления и очереди – новые решения агентов.

Именно в этом контуре формируются устойчивые аномалии – перегрузки, задержки, переполнение складов и хронические узкие места. Сеть считается структурно недостаточной, если при допустимых параметрах агентов и существующей топологии сохраняется устойчивое нарушение ограничений:

$$\exists\ i \in V,\ e \in E: \quad U_i(t) > 1 \ \text{ или } \ L_e(t) > 1 \quad \forall\, t \in [T_1, T_2],$$

а локальная адаптация агентов не устраняет перегрузку.

Тогда возникает гипотеза о необходимости модификации сети:

$$G' = (V \cup \{v_{new}\}, E')$$

где

v_{new} – новый логистический узел, добавление которого проверяется повторным моделированием.

Важным результатом исследования является формирование требований к структуре данных, обеспечивающей реализацию предложенной гибридной модели. В соответствии с разработанной онтологией данные должны быть организованы по нескольким взаимосвязанным контурам: инфраструктурному (характеристики портов, складов, транспортных узлов), транспортному (маршруты, их пропускная способность, сезонность, ограничения), операционному (перемещение грузов, загрузка мощностей, очереди, время обработки), а также контекстному (спрос, навигационные окна, природно-климатические условия). Такая структура данных позволяет перейти от разрозненных источников информации к единой семантически согласованной модели, в рамках которой обеспечивается корректная параметризация как системно-динамических зависимостей, так и правил поведения агентов. В частности, она даёт возможность формировать временные ряды потоков и состояний системы, необходимые для калибровки модели, а также реконструировать траектории движения грузов и транспортных средств.

Предложенная структура данных ориентирована на использование как ретроспективной, так и оперативной информации. Для целей анализа устойчивости логистической системы и выявления аномалий целесообразно использование исторических данных за период не менее 1-3 навигационных циклов, что позволяет учитывать сезонные эффекты и вариативность внешних условий. В качестве источников могут выступать данные автоматических систем идентификации судов (AIS), статистика портовых операций, информация о складских остатках и перемещениях грузов, а также метеорологические и ледовые данные. Интеграция указанных источников в рамках онтологической модели обеспечивает возможность построения согласованных временных и пространственных представлений функционирования системы, что является необходимым условием для применения методов имитационного моделирования и анализа аномалий.

Заключение. В работе предложен методологический подход к анализу и развитию региональных логистических систем, основанный на интеграции онтологического моделирования, методов системной динамики и агентно-ориентированного моделирования. В отличие от традиционных подходов, ориентированных на прямое проектирование инфраструктуры, предложенное решение реализует обратную постановку задачи, при которой структура логистической сети определяется на основе анализа динамики её функционирования и выявления устойчивых аномалий поведения.

В рамках исследования разработана онтологическая модель логистической системы, обеспечивающая формализацию предметной области и интеграцию разнородных источников данных. Показано, что онтология выступает ключевым элементом data-driven подхода, задавая семантический каркас, в рамках которого могут быть реализованы методы обработки, анализа и моделирования данных. На основе онтологического представления предложена гибридная имитационная модель, объединяющая системную динамику и агентное моделирование. Такое сочетание позволяет одновременно учитывать макроуровневые закономерности функционирования системы и микроповедение её участников, что обеспечивает более полное и реалистичное описание логистических процессов.

Ключевым результатом работы является формализация подхода к выявлению структурной недостаточности логистической сети через анализ наблюдаемых аномалий, таких как перегрузка узлов, переполнение складских мощностей, устойчивые задержки и дисбалансы потоков. Показано, что данные явления могут быть интерпретированы как индикаторы необходимости изменения структуры сети, включая добавление новых инфраструктурных элементов или перераспределение потоков. Тем самым предложенный подход обеспечивает переход от анализа данных к обоснованию управленческих решений в области развития логистической инфраструктуры.

Подход может быть использован при планировании развития транспортно-логистической инфраструктуры, оптимизации цепей поставок, а также в системах поддержки принятия решений на основе больших данных. В частности, подход может быть полезен при анализе эффективности структуры региональных логистических сетей (в статье приведён пример анализа логистической сети одного из приморских районов Арктической зоны Российской Федерации).

Вместе с тем, реализация предложенного подхода связана с рядом ограничений и направлений дальнейших исследований. К числу потенциальных узких мест относится неполнота и неоднородность исходных данных, различия в уровнях агрегации информации, а также необходимость обеспечения их семантической согласованности. Дополнительной задачей является разработка методов автоматической идентификации параметров модели и формализация критериев устойчивых аномалий на основе наблюдаемых данных. Перспективным направлением является также интеграция методов машинного обучения для повышения точности прогнозирования поведения агентов и выявления скрытых

закономерностей в потоках. Таким образом, предложенная структура данных и модель создают основу для дальнейшего развития инструментов data-driven анализа и проектирования логистических систем, однако требуют последующей эмпирической апробации и уточнения на реальных данных.

Вместе с тем, реализация предложенного подхода связана с рядом ограничений и направлений дальнейших исследований.

К числу потенциальных узких мест относится неполнота и неоднородность исходных данных, различия в уровнях агрегации информации, а также необходимость обеспечения их семантической согласованности.

Дополнительной задачей является разработка методов автоматической идентификации параметров модели и формализация критериев устойчивых аномалий на основе наблюдаемых данных. Перспективным направлением является также интеграция методов машинного обучения для повышения точности прогнозирования поведения агентов и выявления скрытых закономерностей в потоках.

Таким образом, предложенная структура данных и модель создают основу для дальнейшего развития инструментов data-driven анализа и проектирования логистических систем, однако требуют последующей эмпирической апробации и уточнения на реальных данных.

Список литературы

[1] Gruber T.R. A translation approach to portable ontology specifications // Knowledge Acquisition. – 1993. – Vol. 5, No. 2. – P. 199–220.

[2] Guarino N. Formal Ontology and Information Systems // Proceedings of FOIS. – 1998. – P. 3–15.

Studer R., Benjamins V.R., Fensel D. Knowledge engineering: Principles and methods // Data & Knowledge Engineering. – 1998. – Vol. 25. – P. 161–197.

[3] Noy N.F., McGuinness D.L. Ontology Development 101: A Guide to Creating Your First Ontology. – Stanford Knowledge Systems Laboratory, 2001.

[4] Forrester J. W. Industrial Dynamics. Cambridge, MA: MIT Press, 1961.

[5] Sterman J. D. Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston: Irwin/McGraw-Hill, 2000.

[6] Macal C. M., North M. J. Tutorial on agent-based modelling and simulation // Journal of Simulation. 2010. Vol. 4, No. 3. P. 151–162. DOI: 10.1057/jos.2010.3.

[7] Swinerd C., McNaught K. R. Design classes for hybrid simulations involving agent-based and system dynamics models // Simulation Modelling Practice and Theory. 2012. Vol. 25. P. 118–133. DOI: 10.1016/j.simpat.2011.09.002.

[8] Nguyen L. K. N., Howick S., Megiddo I. A framework for conceptualising hybrid system dynamics and agent-based simulation models // European Journal of Operational Research. 2024. Vol. 315, No. 3. P. 1153–1166. DOI: 10.1016/j.ejor.2024.01.027.

[9] Bozdogan A., et al. An agent-based modeling framework for the design of a dynamic closed-loop supply chain network // PLOS ONE. 2022.

Борисов, И. А. Генеративные методы в проектном управлении / А. И. Борисов // Управление проектами. – 2020. – № 12(3). – С. 45-59.

[10] Лёвина А. Агентная модель поддержки принятия решений по управлению портфелем проектов / Лёвина А., Калязина С., Трифонова Н., Антонов А., Абрамов В. // Устойчивое развитие и инженерная экономика. – 2025. – № 1 (15). – С. 58-79.

[11] Smith, J. Machine Learning Applications for Generative Decision-Making / J. Smith, T. Johnson // Journal of Business Intelligence. – 2021. – № 15(2). – PP. 88-99.

Авторский вклад

Лёвина Анастасия Ивановна – постановка задачи исследования, разработка методологии интеграции методов имитационного моделирования.

Дубгорн Алиса Сергеевна – анализ литературы, разработка математических моделей.

EVALUATION OF THE EFFECTIVENESS OF GENERATIVE POLICY IN OPTIMIZING DESIGN DECISIONS

A.I. Levina
Professor, Graduate School of Business Engineering, Peter the Great Polytechnic University,
DrSc (Econ.), Associate Professor

A.S. Dubgorn
Associate Professor, Graduate School of Business Engineering, Peter the Great Polytechnic University,
CandSc (Econ.)

Abstract. Modern regional logistics systems are characterized by high complexity and dynamism, while the available data on their functioning are not used effectively enough to substantiate infrastructure development decisions. Traditional design methods rely on predefined flow parameters, which limits their applicability under conditions of endogenous dynamics. The aim of this work is to develop a data-driven approach to the analysis and development of logistics networks based on the integration of ontological modeling, system dynamics, and agent-based modeling. An ontology of the logistics system is proposed to ensure the integration of heterogeneous data, along with a hybrid simulation model that enables the analysis of both the system's macrodynamics and the behavior of its participants. As a result, an approach is formulated for identifying network functioning anomalies (congestion, delays, flow imbalances) based on data analysis and their interpretation as indicators of structural insufficiency. It is shown that integrating big data analysis methods with simulation modeling makes it possible to substantiate directions for the development of logistics infrastructure.

Keywords: logistics systems, logistics infrastructure development, system dynamics, agent-based modeling, data-driven management.

УДК 004.89:629.7.064.5

ПРИМЕНЕНИЕ МЕТОДОВ BIG DATA В РАЗРАБОТКЕ АДАПТИВНЫХ СИСТЕМ УПРАВЛЕНИЯ ЭЛЕРОНАМИ ЛЕТАТЕЛЬНЫХ АППАРАТОВ



Е.И. Бавбель
Аспирант кафедры
проектирования информационно-компьютерных систем БГУИР
egorigorevichw9@gmail.com

Е.И. Бавбель
Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации.

Аннотация. В статье рассмотрены особенности проектирования и применения элеронов как органов поперечного управления летательного аппарата, а также связанных с ними систем управления. Проанализированы принципы формирования управляющих моментов, конструктивные решения и функциональные особенности исполнительных механизмов. Показана взаимосвязь аэродинамических и механических характеристик элеронов с параметрами системы управления. Обосновано, что выбор рациональных конструктивных параметров и алгоритмов управления позволяет повысить устойчивость и управляемость летательного аппарата.

Особое внимание уделено возможностям использования методов анализа больших данных (Big Data) при разработке адаптивных систем управления элеронами. Предложен подход к обработке массивов полетной

информации, включающей данные о параметрах движения, характеристиках аэродинамики и работе исполнительных механизмов. На основе анализа этих данных формируются адаптивные алгоритмы, позволяющие корректировать параметры управления в зависимости от текущего режима полета и состояния конструкции. Использование методов Big Data дает возможность выявлять скрытые закономерности, прогнозировать изменения аэродинамической эффективности элеронов и своевременно адаптировать управляющие воздействия.

Результаты работы могут быть применены при создании систем управления беспилотных летательных аппаратов с повышенными требованиями к точности, надежности и адаптивности в изменяющихся условиях эксплуатации.

Ключевые слова: элероны, системы управления, летательный аппарат, управляющий момент, аэродинамика, устойчивость, управляемость.

Введение. Управляемость и устойчивость летательного аппарата в значительной степени определяются эффективностью его органов управления. Элероны, обеспечивающие создание управляющего момента относительно продольной оси за счет перераспределения подъемной силы по консолям крыла, занимают среди них центральное место. От характеристик элеронов напрямую зависят маневренность, точность траекторного управления и устойчивость полета [1, с. 2].

Современные летательные аппараты эксплуатируются в широком диапазоне режимов, что предъявляет повышенные требования к точности, надежности и адаптивности систем управления. Это влечет усложнение как конструктивного исполнения элеронов, так и алгоритмического обеспечения их управления. В числе ключевых факторов, подлежащих учету, – аэродинамические нелинейности, зависимость характеристик от скорости полета и взаимодействие с другими органами управления [1, с. 3]. Разработка элеронов и систем их управления представляет собой многокомпонентную задачу, объединяющую аэродинамический анализ, механическое проектирование и синтез алгоритмов управления. Согласование этих компонентов становится критическим: недостаточная эффективность любого из них может привести к ухудшению общей управляемости аппарата [2, с. 47].

В последние годы активно развиваются подходы, основанные на анализе больших данных (Big Data). В авиационной отрасли такие методы применяются для обработки телеметрической информации, идентификации параметров летательных аппаратов, прогнозирования отказов и оптимизации управления [1, с. 1; 3, с. 8]. В работе [4] предложена методика автоматизированного анализа данных бортовых систем ракеты-носителя с использованием гибридной архитектуры на основе сверточных нейронных сетей (CNN) и долгой краткосрочной памяти (LSTM), интегрированной с платформами Apache Hadoop и SparkML. Авторами показано, что применение методов машинного обучения в сочетании с инфраструктурой Big Data позволяет сократить время анализа и повысить точность выявления аномалий [4, с. 2]. В исследовании [3] представлена усовершенствованная стратегия управления летательным аппаратом на основе данных, включающая идентификацию, управление и валидацию. Предложенный подход позволяет синтезировать регулятор без явного моделирования объекта, используя только вход-выходные данные. Авторами также показано, что при параметризованном регуляторе задача управления сводится к идентификации параметров, что опирается на развитый аппарат теории оценивания [3, с. 2–3].

Вместе с тем вопросы применения методов Big Data и аналитики для проектирования и управления элеронами как исполнительными органами летательного аппарата остаются проработанными недостаточно. Требуют уточнения особенности учета аэродинамических характеристик элеронов в условиях больших массивов полетных данных, а также возможности использования интеллектуальных алгоритмов для адаптации параметров управления в реальном времени.

Таким образом, цель настоящих исследований – анализ особенностей проектирования элеронов и систем их управления с позиции интеграции методов анализа больших данных,

направленный на повышение эффективности формирования управляющих моментов и устойчивости летательного аппарата.

Аэродинамические и конструктивные особенности работы элеронов.

Элероны являются основным органом поперечного управления летательного аппарата и обеспечивают формирование кренящего момента за счет перераспределения подъемной силы по размаху крыла [5, с. 2]. При отклонении элерона на одной консоли происходит увеличение подъемной силы, на противоположной – ее уменьшение, что приводит к возникновению вращательного движения вокруг продольной оси [7]. Эффективность данного процесса определяется совокупностью аэродинамических и геометрических факторов, а также условиями полета [13].

Характер обтекания крыла с отклоненными элеронами отличается от базового случая наличием локального изменения угла атаки [5, с. 3]. На участке крыла, где элерон отклонен вниз, эффективный угол атаки возрастает, что сопровождается ростом подъемной силы и, как правило, увеличением сопротивления. На противоположной консоли наблюдается обратный эффект. В результате возникает не только полезный управляющий момент, но и побочный рыскательный момент, обусловленный асимметрией сопротивления. Это явление, известное как неблагоприятный рыскательный момент, требует учета при проектировании систем управления [7].

Величина формируемого управляющего момента зависит от плеча приложения аэродинамической силы, которое определяется положением элерона по размаху крыла [5, с. 4]. Наиболее эффективно размещение элеронов в удаленных от корня зонах, где плечо максимальное. Однако такая компоновка приводит к увеличению изгибающих нагрузок на консоль крыла и требует усиления конструкции. Таким образом, при проектировании возникает необходимость поиска компромисса между аэродинамической эффективностью и прочностными ограничениями [2, с. 47].

Существенное влияние на характеристики элеронов оказывает их относительная площадь. Увеличение площади приводит к росту управляющего воздействия, однако одновременно возрастает сопротивление и нагрузка на приводы [11]. При этом избыточное увеличение площади может привести к ухудшению общей аэродинамической эффективности крыла. Практика проектирования летательных аппаратов свидетельствует о том, что рациональная площадь элеронов определяется в пределах 20–30 % от площади консоли крыла. Данное соотношение позволяет достичь компромисса между требуемым уровнем управляемости и аэродинамическим сопротивлением [9].

Значительное влияние на эффективность поперечного управления оказывает величина угла отклонения элеронов. Экспериментальные исследования показывают, что при небольших углах отклонения наблюдается практически пропорциональная зависимость между управляющим моментом и углом отклонения, что облегчает процесс управления. По мере увеличения угла отклонения начинают проявляться нелинейные аэродинамические эффекты, обусловленные развитием отрывных явлений в пограничном слое. Указанные процессы приводят к снижению инкремента подъёмной силы и деградации аэродинамической эффективности элеронов. В связи с этим в штатных режимах эксплуатации углы отклонения ограничиваются значениями, при которых обеспечивается прикреплённое обтекание профиля [13].

С конструктивной точки зрения элероны выполняются как подвижные поверхности, шарнирно соединённые с несущими элементами крыла. Функционирование элеронов сопряжено с передачей значительных аэродинамических нагрузок на элементы привода, что предъявляет повышенные требования к жёсткостным и прочностным характеристикам механизмов управления [2, с. 48]. В современной авиационной технике находят применение как традиционные механические системы привода, так и электромеханические актуаторы, отличающиеся высокой точностью позиционирования и быстродействием.

Система управления элеронами должна обеспечивать формирование требуемого управляющего воздействия и его координацию с функционированием иных органов управления. Для устранения нежелательного рыскательного момента, возникающего при отклонении элеронов, применяется принцип дифференциального управления, при котором величины углов отклонения элеронов вверх и вниз различны. Указанный подход способствует снижению асимметрии аэродинамического сопротивления и повышению устойчивости самолёта в режиме поперечного управления [5, с. 5].

Количественная оценка управляющего момента элеронов производится по зависимости:

$$Mx=12V2SbCl, \quad (1)$$

где Cl – коэффициент крена, определяемый параметрами крыла и углом отклонения элеронов.

Из (1) видно, что эффективность управления возрастает с увеличением скорости полета и геометрических размеров крыла.

Анализ выражения показывает, что эффективность работы элеронов определяется совокупностью аэродинамических, геометрических и конструктивных факторов. Необходимо комплексно подходить к проектированию БПЛА, учитывая характеристики обтекания, особенности силовой схемы крыла, а также системы управления.

Динамика управления элеронами и особенности их взаимодействия с системой управления. Функционирование элеронов не сводится исключительно к формированию статического управляющего момента. В условиях реального полёта необходим учёт динамических характеристик их отклика. Процесс изменения угла отклонения управляющих поверхностей сопровождается переходными режимами, в ходе которых возникают не только целевой крен, но и сопутствующие аэродинамические и инерционные эффекты [14]. Система автоматического управления обязана обеспечивать устойчивый и детерминированный отклик на управляющее воздействие.

При отклонении элеронов фиксируется временной лаг между формированием управляющей команды и фактическим изменением пространственного положения летательного аппарата. Указанная задержка обусловлена инерционными свойствами привода и аэродинамической инерцией конструкции [1, с. 5]. Система управления корректирует динамические свойства объекта с целью устранения перерегулирования и колебательных режимов. При недостаточном демпфировании возможно возбуждение автоколебаний по крену, что детерминирует снижение устойчивости и точности управления [14].

Изменение крена неизбежно сопровождается изменением рыскания из-за асимметрии сопротивления при отклонении элеронов [1, с. 3]. При отсутствии компенсации аппарат отклоняется от заданного курса. Для устранения этого эффекта применяется согласованное управление: элероны работают совместно с рулем направления. Такое взаимодействие снижает паразитные моменты и повышает точность управления [15, с. 112].

Тип привода оказывает существенное влияние на динамические свойства системы управления. В механических системах доминирующими ограничивающими факторами выступают люфты и упругие деформации тяг. В электромеханических приводах приоритетными характеристиками являются быстродействие и точность позиционирования [16]. Современные системы управления базируются на цифровых алгоритмах, обеспечивающих адаптацию динамических характеристик отклика в зависимости от скоростного режима полёта. Указанный подход гарантирует устойчивое функционирование системы в диапазоне от низких до высоких скоростей [16].

По мере нарастания скоростного режима полёта аэродинамическая эффективность элеронов возрастает, что может инициировать избыточную резкость отклика. Для нейтрализации данного эффекта применяются ограничения на максимальный угол отклонения либо реализуются алгоритмы управления с адаптацией к скорости полёта [15,

с. 113]. Перечисленные меры обеспечивают равномерность управляемости на всём диапазоне эксплуатационных режимов.

Устойчивость конструкции элеронов при воздействии динамических нагрузок представляет самостоятельный аспект проектирования. Вибрационные воздействия и переменные аэродинамические нагрузки инициируют процессы усталостного накопления повреждений, преимущественно в зонах узлов крепления и приводных механизмов [2, с. 48]. Это обусловливает необходимость выполнения расчётов на долговечность и выбора конструктивно-технологических решений с требуемым запасом прочности.

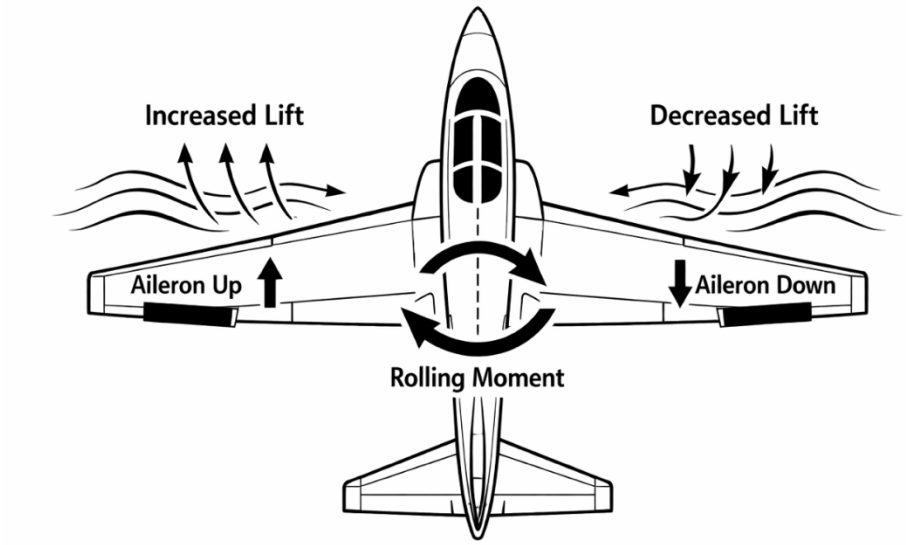


Рисунок 1. Схема работы элеронов и формирования управляющего момента

Представленная на рисунке схема иллюстрирует принципиальный механизм формирования управляющего момента. В условиях реального полёта обтекание несущей поверхности приобретает более сложный пространственный характер. На распределение подъёмной силы оказывают влияние индуктивные эффекты, концевые вихревые структуры и взаимодействие с элементами фюзеляжа [1, с. 3]. Фактическая аэродинамическая эффективность элеронов может девиировать от расчётных значений, преимущественно в режимах низких скоростей и при значительных углах отклонения. Учёт перечисленных факторов осуществляется на этапе аэродинамического моделирования и натурных экспериментальных исследований [2, с. 47].

Особенности применения элеронов в различных режимах полета. Аэродинамическая эффективность элеронов варьируется в зависимости от скоростного режима, высоты полёта и угла атаки. Модификация условий обтекания крыла детерминирует характеристики формируемого управляющего момента [15, с. 108].

На этапах взлёта и посадки, характеризующихся низкими скоростями, эффективность элеронов деградирует вследствие снижения динамического напора. Для обеспечения требуемой управляемости приходится увеличивать углы отклонения, что усиливает нелинейные эффекты, включая локальный отрыв потока. Это снижает прирост подъемной силы и ухудшает эффективность управления. В ряде случаев для компенсации недостатка управляющего момента применяются закрылки или интерцепторы [17].

С увеличением скорости до крейсерских значений эффективность элеронов возрастает. Зависимость управляющего момента от угла отклонения остается близкой к линейной, что упрощает реализацию алгоритмов управления [17]. При этом рост скорости

увеличивает нагрузки на конструкцию и приводы, что требует ограничения максимальных углов отклонения [15, с. 109].

Высокие скорости. При больших скоростях проявляются эффекты, связанные с изменением распределения давления по профилю крыла. Возможно уменьшение эффективности элеронов или даже изменение знака создаваемого управляющего момента из-за перераспределения аэродинамических сил и изменения структуры потока. В таких условиях традиционные схемы управления могут оказаться недостаточными, используются дополнительные органы управления или специальные законы управления [16, с. 215].

Влияние угла атаки и пространственного обтекания. С ростом угла атаки возрастает вероятность отрыва потока, что снижает эффективность органов управления. Особенно это заметно в околокритических режимах, где управление становится менее устойчивым. В указанных условиях возникает необходимость пристального внимания к характеристикам системы управления, обеспечивающей сохранение контролируемости летательного аппарата [15, с. 110].

Пространственный характер обтекания несущей поверхности оказывает существенное влияние на аэродинамическую эффективность элеронов. В условиях реального полёта поток обладает трёхмерной структурой, детерминированной взаимодействием с концевыми вихревыми образованиями и элементами фюзеляжа [18]. В приконцевых зонах крыла индуктивные эффекты усиливаются, что может детерминировать снижение инкремента подъёмной силы при отклонении элеронов. Данное обстоятельство приобретает особое значение при размещении элеронов в периферийных зонах: с одной стороны, достигается максимальное плечо силы, с другой –возрастает влияние вихревых структур [18].

Отклонение элеронов сопровождается существенной модификацией распределения давления по поверхности крыла. Перераспределение аэродинамических нагрузок инициирует формирование локальных зон повышенных напряжений. При неблагоприятных аэродинамических условиях это способствует развитию отрывных явлений, преимущественно вблизи задней кромки. В результате наблюдается деградация эффективности управления и инкремент аэродинамического сопротивления [18].

Упругие свойства конструкции крыла оказывают влияние на функционирование элеронов. Под воздействием аэродинамических нагрузок возникают деформации, модифицирующие геометрию профиля и угол установки управляющих поверхностей. Это детерминирует изменение характеристик управления и может инициировать аэроупругие явления: деградацию управляющего момента или запаздывание реакции летательного аппарата на управляющее воздействие [2, с. 49]. Взаимодействие элеронов с системой автоматического управления представляет собой дополнительный фактор, подлежащий учёту. При анализе эффективности системы поперечного управления необходимо рассматривать совокупное влияние аэродинамических и кинематических параметров. В частности, вариация угла атаки, скоростного режима полёта и характеристик набегающего потока обусловливает изменение коэффициента крена, которое может быть формализовано следующей функциональной зависимостью:

$$Cl=f(a,,V), \quad (2)$$

где: a– угол отклонения элерона, – угол атаки, V– скорость полета.

Указанное соотношение отражает многофакторный характер формирования управляющего воздействия: аэродинамическая эффективность элеронов определяется не изолированными параметрами, а их взаимосвязанным влиянием [19].

Конструктивные и эксплуатационные ограничения при проектировании элеронов. При разработке элеронов существенное значение приобретают ограничения, обусловленные как прочностными характеристиками конструкции, так и условиями эксплуатации летательного аппарата. В отличие от идеализированных аэродинамических

моделей, реальные управляющие поверхности функционируют в условиях переменных аэродинамических нагрузок, вибрационных воздействий и температурных факторов, что требует комплексного учёта перечисленных явлений на этапе проектирования.

Одним из определяющих ограничений является прочность и жёсткость конструкции крыла в зоне размещения элеронов. При отклонении управляющих поверхностей возникают дополнительные аэродинамические силы, порождающие изгибающие и крутящие моменты. Перечисленные нагрузки транслируются на силовой набор крыла, что может инициировать локальные концентрации напряжений и деформации. При недостаточной жёсткости конструктивных элементов возможны нежелательные явления, в частности упругая деформация профиля несущей поверхности, приводящая к деградации эффективности управления. В ряде случаев указанный эффект может трансформироваться в аэродинамическую неустойчивость, обусловленную взаимодействием упругих и аэродинамических сил [20].

Ограничение по усилиям в системе управления. По мере нарастания скоростного режима полёта возрастает аэродинамическая нагрузка на элероны, что детерминирует увеличение требуемого усилия на приводах. В механических системах управления указанное обстоятельство проявляется в росте усилий на органах управления, снижая эргономические характеристики пилотирования. В автоматизированных системах возникает необходимость в наращивании мощности приводов и повышении точности их позиционирования. Для оценки нагрузок на привод применяется следующее выражение:

$$F=12V2SaCh, \quad (3)$$

где: F – сила, действующая на элерон; ρ – плотность воздуха; V – скорость полета; Sa – площадь элерона; Ch– коэффициент шарнирного момента.

Нагрузка на привод возрастает пропорционально квадрату скорости полёта, что детерминирует критичность режимов высоких скоростей. Это обусловливает необходимость введения ограничений на максимальные углы отклонения либо применения усиленных приводов с отрицательной обратной связью.

Эксплуатационные условия. Вибрационные воздействия, возникающие в процессе полёта, способны инициировать процессы усталостного накопления повреждений в элементах крепления и приводных механизмах. Особую актуальность данная проблематика приобретает для летательных аппаратов с продолжительным ресурсом эксплуатации. Для повышения надёжности функционирования применяются конструктивно-технологические решения, обеспечивающие снижение концентрации напряжений и равномерное распределение нагрузок.

Влияние на аэродинамическое качество. Отклонение элеронов сопровождается инкрементом лобового сопротивления, что может детерминировать снижение энергетической эффективности полёта, преимущественно при длительном выполнении маневров. При проектировании реализуются подходы, направленные на минимизацию указанного влияния путём оптимизации геометрических параметров и диапазона углов отклонения.

Заключение. Формирование управляющего момента детерминировано перераспределением подъёмной силы по размаху крыла и обусловлено геометрическими параметрами элеронов, величиной их отклонения и параметрами полёта. Аэродинамическая эффективность элеронов существенно зависит от скоростного режима. При низких скоростях управляющее воздействие деградирует вследствие снижения динамического напора; при высоких скоростях возрастают нагрузки на конструкцию и систему управления. По мере увеличения угла атаки повышается вероятность развития отрывных явлений и деградации эффективности управляющих поверхностей.

На характеристики поперечного управления оказывают влияние пространственная структура потока, аэроупругие явления и взаимодействие с системой автоматического

управления. Это подтверждает необходимость системного подхода при проектировании, учитывающего взаимосвязь аэродинамических и конструктивных параметров. При разработке элеронов требуется учёт ограничений, связанных с прочностью конструкции, нагрузками на приводы и условиями эксплуатации. Полученные результаты могут быть использованы при проектировании перспективных летательных аппаратов и модернизации систем управления.

Список литературы

[1] Improved data driven strategy for aircraft controller design / Scientific Reports. – 2025. – Vol. 15, Article number 40438.

[2] Васильев, В. В. Механика конструкций из композиционных материалов / В. В. Васильев. – Москва : Машиностроение, 1988. – 272 с.

[3] Ван, С. Управление летательным аппаратом на основе данных: идентификация, управление и валидация / С. Ван, Х. Чжан, Л. Лю // Авиакосмические системы. – 2024. – Т. 12, № 3. – С. 1–15.

[4] Saju, S. Bigdata Analytical Framework For Launch Vehicle Avionics System Automated Clearance Using Advanced ML Algorithms / S. Saju, A. Varghese, S. Sujith [et al.] // Proceedings of the International Astronautical Congress (GLEX-2025). – 2025. – Paper GLEX-2025,13,IP,6,x93078.

[5] Adamopoulou, E. Applications and Technologies of Big Data in the Aerospace Domain / E. Adamopoulou, E. Daskalakis // Electronics. – 2023. – Vol. 12, No. 10. – P. 2225.

[6] Badea, G. Big Data in the Aerospace Industry / G. Badea // Informatica Economica. – 2018. – Vol. 22, No. 1. – P. 17–24.

[7] Приводимый в действие поворотом элерон, установленный с зазором и создающий большую подъемную силу : пат. 2625384 Российская Федерация, МПК B64C 9/02 / заявитель и патентообладатель The Boeing Company. – № 2015149638 ; заявл. 17.11.2015 ; опубл. 12.07.2017, Бюл. № 20. – 16 с.

[8] Chinchanikar, S. A Review on Machine Learning, Big Data Analytics, and Design for Additive Manufacturing for Aerospace Applications / S. Chinchanikar, S.S. Ohol // Journal of Materials Engineering and Performance. – 2022. – Vol. 31, No. 8. – P. 6112–6130.

[9] Левитский, Н.Ф. Видоизменение элеронов самолета : пат. 18073 СССР, МПК B64C 9/02 / Н.Ф. Левитский. – № 29301 ; заявл. 23.06.1928 ; опубл. 30.09.1930, Бюл. № 1. – 3 с.

[10] Wang, S. A New Data Processing Architecture for Multi-Scenario Applications in Aviation Manufacturing / S. Wang [et al.] // IEEE Access. – 2019. – Vol. 7. – P. 83637–83648.

[11] Левитский, Н. Ф. Элероны самолета : пат. 18072 СССР, МПК B64C 9/12 / Н.Ф. Левитский. – № 29300 ; заявл. 22.09.1928 ; опубл. 30.09.1930, Бюл. № 1. – 2 с.

[12] Mikalef, P. Big Data Analytics and Firm Performance: Findings from a Mixed-Method Approach / P. Mikalef [et al.] // Journal of Business Research. – 2019. – Vol. 98. – P. 261–276.

[13] Орган управления летательного аппарата : пат. 2028251 Российская Федерация, МПК B64C 9/00 / И. М. Харитонов, В. А. Зинченко ; заявитель и патентообладатель Авиационный научно-технический комплекс им. А.Н. Туполева. – № 5034788 ; заявл. 14.02.1992 ; опубл. 09.02.1995, Бюл. № 4. – 4 с.

[14] Козлов, В. В. Управляемость и устойчивость летательных аппаратов / В. В. Козлов. – Москва : Машиностроение, 1985. – 304 с.

[15] Петров, А. В. Аэродинамика летательных аппаратов / А. В. Петров. – Санкт-Петербург : Политехника, 2010. – 512 с.

[16] Воскресенский, А. А. Алгоритмы управления и распознавания, применяемые на БПЛА / А. А. Воскресенский, Е. И. Бавбель // Электронные системы и технологии. – 2014. – № 2. – С. 45–49.

[17] Алексеев, В. Ф. Автономная посадка БПЛА с использованием визуального сервоуправления / В. Ф. Алексеев, Е. И. Бавбель // Электронные системы и технологии. – 2024. – № 1. – С. 32–38.

[18] Жэнь, Цзиньху [и др.] Разработка и применение системы управления элеронами на основе моделей [на кит. яз.] // Проектирование и исследование гражданских самолетов. – 2022. – № 1. – С. 50–59.

[19] Мерзляков, В. Н. Системы управления летательными аппаратами / В. Н. Мерзляков. – Москва : Машиностроение, 1990. – 368 с.

[20] Nelson, R. C. Flight Stability and Automatic Control / R. C. Nelson. – New York : McGraw-Hill, 1998. – 512 p.

Авторский вклад

Бавбель Егор Игоревич – постановка задачи исследования, выполнен анализ конструктивных и аэродинамических особенностей элеронов и систем их управления летательных аппаратов, сформулированы основные зависимости, описывающие формирование управляющих моментов. Проведено обобщение современных подходов к проектированию систем управления, а также выполнена интерпретация полученных результатов с позиции их практического применения.

APPLICATION OF BIG DATA METHODS IN THE DEVELOPMENT OF ADAPTIVE AILERON CONTROL SYSTEMS FOR AIRCRAFT

E.I. Bavbel
Postgraduate student, Department of Information
Computer Systems Design BSUIR

Abstract. The paper discusses the design and application features of ailerons as aircraft lateral control surfaces, as well as the associated control systems. The principles of control moment generation, design solutions, and functional characteristics of actuation mechanisms are analyzed. The relationship between the aerodynamic and mechanical properties of ailerons and the parameters of the control system is demonstrated. It is substantiated that the selection of rational design parameters and control algorithms can improve aircraft stability and controllability.

Particular attention is paid to the potential of using Big Data analytics in the development of adaptive aileron control systems. An approach is proposed for processing flight data arrays, including motion parameters, aerodynamic characteristics, and actuation mechanism performance. Based on the analysis of these data, adaptive algorithms are formed to adjust control parameters depending on the current flight mode and structural condition. The use of Big Data methods enables the identification of hidden patterns, prediction of changes in aileron aerodynamic efficiency, and timely adaptation of control inputs.

The results of the work can be applied in the development of control systems for unmanned aerial vehicles with increased requirements for accuracy, reliability, and adaptability under varying operating conditions.

Keywords: ailerons, control systems, aircraft, roll moment, aerodynamics, stability, controllability

УДК 004.67

СЕМАНТИЧЕСКОЕ СЛИЯНИЕ ДАННЫХ НА ОСНОВЕ ГРАФОВ ЗНАНИЙ: ИНТЕГРАЦИЯ СТРУКТУРИРОВАННЫХ И НЕСТРУКТУРИРОВАННЫХ ДАННЫХ ДЛЯ РАСКРЫТИЯ КОНТЕКСТНОГО ИНТЕЛЛЕКТА



Е.А. Алуев
Инженер-исследователь ATEK,
Бакалавр технических наук
alooeff@atek.dev



М.А. Булычева
Инженер ATEK,
Магистр математических наук
mbulytcheva@gmail.com

Е.А. Алуев

Бакалавр компьютерных наук, факультет ЭВМиС Брестского государственного технического университета. Научные интересы: облачные технологии, машинное обучение и мультиагентное моделирование.

М.А. Булычева

Магистр математических, кафедра механики и математики Государственного университета им. Ломоносова, Москва. Научные интересы: машинное обучение, крупномасштабная оптимизация и прогнозирование.

Аннотация. Компании сектора профессиональных услуг работают с разнородными данными, где структурированные системы сосуществуют с большими массивами неструктурированных документов. Их раздельная обработка ограничивает получение контекстно-зависимых и межисточниковых инсайтов, необходимых для таких задач, как выявление конфликтов, поиск экспертизы и оценка рисков. В работе

предлагается архитектура слияния данных на основе графа знаний, объединяющая структурированные и неструктурированные корпоративные данные в единое семантическое представление с сохранением происхождения данных. Для извлечения сущностей и отношений из документов используются методы обработки естественного языка, а разрешение сущностей между источниками выполняется с помощью гибридной стратегии, сочетающей лексическое, семантическое и реляционное сопоставление. Полученный размеченный граф свойств поддерживает логический вывод и построение признаков для машинного обучения. Результаты показывают, что такой подход повышает качество сопоставления сущностей, улучшает эффективность последующей аналитики и обеспечивает более высокую интерпретируемость. Семантическая интеграция на основе графов знаний тем самым выступает масштабируемой и эффективной основой для корпоративной аналитики в гетерогенных средах данных.

Ключевые слова: граф знаний; интеграция данных; неструктурированные данные; сопоставление сущностей; семантический анализ; корпоративная аналитика.

Введение. Рост объемов структурированных и неструктурированных данных в корпоративных средах создает серьезные вызовы для аналитики и систем поддержки принятия решений. В компаниях сектора профессиональных услуг данные распределены по разрозненным источникам, включая CRM-системы, workflow-платформы и массивы текстовых документов, таких как договоры, отчеты и переписка. Их раздельная обработка ограничивает возможность получения межисточниковых инсайтов и устранения семантической неоднозначности.

Графы знаний представляют собой эффективный подход к интеграции гетерогенных данных в единое семантическое пространство, где сущности и их отношения могут анализироваться совместно. Такое представление позволяет учитывать контекст и поддерживать более сложные аналитические сценарии, включая выявление конфликтов, поиск экспертизы и профилирование рисков.

В данной статье предлагается масштабируемая архитектура построения корпоративного графа знаний, объединяющая структурированные и неструктурированные данные из нескольких источников для поддержки аналитики и задач машинного обучения. Показано, что семантическая интеграция повышает полноту данных, контекстную релевантность и объяснимость результатов.

Основной вклад работы состоит в следующем:

- предложен масштабируемый конвейер приема и семантического слияния данных из нескольких источников;
- разработаны методы извлечения и сопоставления сущностей в гетерогенных данных;
- продемонстрированы преимущества объединенных представлений по сравнению с раздельной обработкой источников.

Предпосылки и обзор литературы.

Графы знаний и интеграция данных. Граф знаний формально определяется как графовая репрезентация объектов реального мира и их семантических отношений, позволяющая интегрировать различные типы данных через узлы, представляющие сущности, и ребра, представляющие отношения. Предыдущие исследования показали, что графы знаний являются фундаментальной структурой интеграции гетерогенных данных, включая как структурированные, так и неструктурированные источники. Tamašauskaitė и соавт. определяют графы знаний как системы, аккумулирующие и передающие знания о реальном мире путем представления сущностей и их связей в семантически насыщенном формате, подчеркивая роль графов в преодолении разрозненности данных [1].

Cudré-Mauroux и соавт. описывают конвейер интеграции данных, демонстрирующий, каким образом полуструктурированный и неструктурированный контент может быть интегрирован с использованием графов знаний, включая предварительную обработку, преобразование и семантическое связывание элементов данных [2].

Сопоставление сущностей и слияние знаний. Сопоставление сущностей означает установление соответствия между семантически идентичными сущностями в разных наборах данных или графах знаний, что является критически важным для многоканального

слияния знаний. Современные обзоры подчеркивают, что сопоставление сущностей - это ключевая технология, обеспечивающая интеграцию разрозненных данных в единые графовые структуры. Методы обучения представлений для сопоставления сущностей демонстрируют более высокую эффективность по сравнению с традиционными признаковыми подходами, особенно в крупных и гетерогенных графовых средах [3].

Исследования в области многоканального слияния знаний подчеркивают, что эффективное объединение требует не только связывания сущностей, но и разрешения конфликтов, выравнивания атрибутов и вывода новых отношений. Именно эти процессы во многом определяют качество и практическую ценность итогового графа знаний для аналитики и логического вывода [4].

Построение графов знаний из структурированных и неструктурированных данных. Исследования, посвященные построению корпоративных графов знаний, подчеркивают сложность интеграции структурированных записей с сущностями и отношениями, извлеченными из текста. Yan и соавт. предлагают подход к объединению структурированных и неструктурированных данных в интегрированный граф знаний, рассматривая предварительную обработку текста, извлечение информации и стратегии хранения [5]. Более ранняя работа Masoud и соавт. представляет обзор методов автоматического построения графов знаний как из структурированных, так и из текстовых источников данных, выделяя точки интеграции до, во время и после построения графа [6]. В совокупности эти исследования формируют основу для понимания слияния данных на основе графов знаний как самостоятельного исследовательского и инженерного направления, демонстрируя необходимость семантического выравнивания, разрешения сущностей и мультимодальной интеграции для корпоративных аналитических приложений.

Постановка задачи и сценарии использования.

Гетерогенные корпоративные данные. Компании обычно хранят структурированные данные в реляционных базах данных, CRM-системах и системах учета финансовых транзакций, тогда как неструктурированные данные, такие как электронные письма, отчеты, юридические заключения и договоры, размещаются в документных хранилищах. Такое разнообразие усложняет аналитические задачи, требующие межисточникового контекста, например выявление всех упоминаний клиентской сущности в документах разных типов или сопоставление структурированных транзакционных историй с их описанием в повествовательной форме.

Аналитические сценарии использования. Объединенные представления знаний поддерживают ряд высокоценных сценариев применения:

– выявление конфликтов: обнаружение пересечений или несовместимостей в клиентских взаимодействиях между подразделениями;

– поиск экспертизы: профилирование внутренней предметной экспертизы путем связывания структурированных кадровых баз с вкладом сотрудников в документы;

– оценка рисков: агрегирование факторов риска из структурированных журналов комплаенса и текстовых оценок рисков для формирования более полной риск-оценки.

Иллюстративный сценарий из сферы коммерческой недвижимости. Чтобы показать архитектуру на конкретном примере, рассмотрим процесс инвестирования в коммерческую недвижимость. Структурированные CRM-системы отслеживают инвестиционные возможности, брокеров, показатели андеррайтинга и стадии воронки. Параллельно неструктурированные информационные меморандумы о продаже объекта (Offering Memoranda, OM) содержат описания характеристик объектов, финансовых допущений, состава арендаторов, долговых ковенант и раскрытия рисков.

Например, запись в CRM может содержать:

- идентификатор возможности #7421;
- целевой объект: “Orchard Lofts”;

- ожидаемый NOI;
-назначенного брокера.

В то же время документ OM, например, содержит текстовые утверждения, такие как:

- заявленная ставка капитализации (5,2%);
- долговой ковенант (DSCR ≥ 1,25);
- экологическая проверка Phase I находится в процессе;
- сведения о концентрации арендаторов.

Изолированные системы не могут ответить на такие запросы, как:

- «Показать 10 лучших возможностей по ожидаемому NOI, исключив сделки, где в OM указана заполняемость ниже 90%»;
- «Выявить возможности, где одна и та же организация выступает и брокером, и управляющей компанией объекта».

Для ответа на такие вопросы требуется объединение структурированных CRM-данных с фактами, извлеченными из неструктурированных OM-документов (рисунок 1).

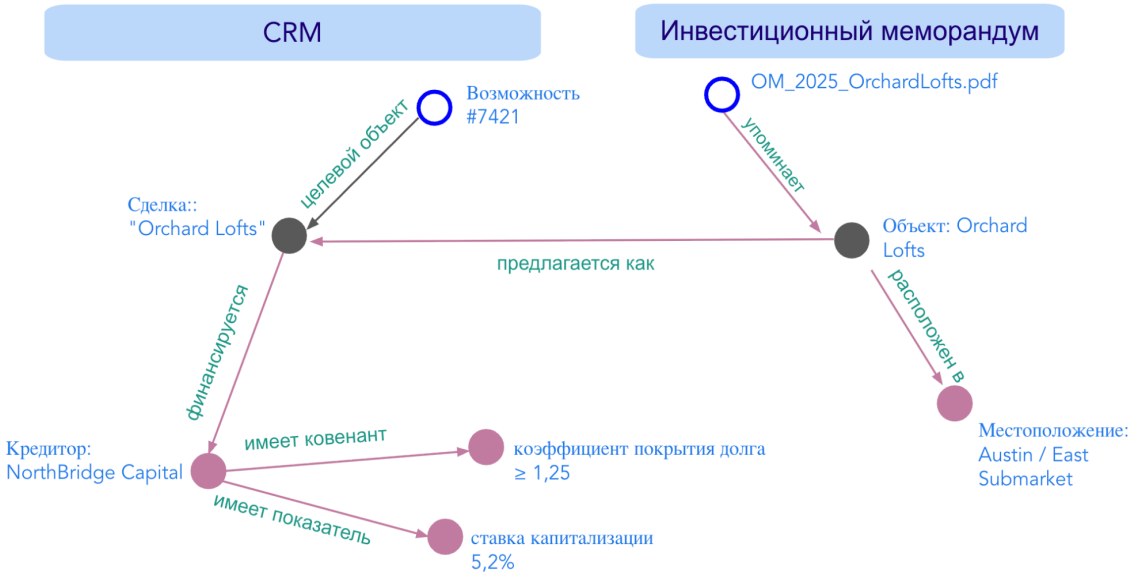


Рисунок 1. Пример

Помимо простого извлечения данных, такие запросы требуют интеграции разнородных свидетельств из разных модальностей и логического вывода по отношениям, охватывающим несколько систем. Например, чтобы определить, превышает ли концентрация арендаторов заданный порог, может потребоваться объединение структурированных данных андеррайтинга со списками арендаторов, извлеченными из нарративных разделов OM. Аналогично, выявление организационных конфликтов может требовать разрешения разных вариантов имен сущностей в CRM-записях и упоминаниях в документах, а затем анализа ролевых связей, например брокера, управляющего объектом или кредитора, в рамках одной и той же сделки. Эти задачи по своей природе реляционны и не могут быть сведены к изолированному поиску по ключевым словам или простым соединениям таблиц.

Архитектура системы. Предлагаемая архитектура представляет собой модульный многоэтапный конвейер, в рамках которого гетерогенные корпоративные данные последовательно преобразуются в единый граф знаний с сохранением происхождения данных. Вместо того чтобы рассматривать структурированные и неструктурированные данные как параллельные аналитические потоки, архитектура изначально проектируется

так, чтобы сводить их в единый семантический слой, поддерживающий логический вывод, аналитику и машинное обучение.

На высоком уровне система разделяет задачи приема данных, семантической интерпретации, выравнивания идентичности, построения графа и последующей аналитики. Каждый этап может развиваться и масштабироваться независимо, что позволяет архитектуре адаптироваться к новым источникам данных, обновленным моделям извлечения и меняющимся требованиям к схеме без нарушения работы всей системы. Такая модульность особенно важна для корпоративных сред, где форматы данных, бизнес-процессы и регуляторные ограничения меняются со временем.

Центральным принципом проектирования является рассмотрение отношений как объектов первого класса. Отношения не редуцируются к неявным соединениям по внешним ключам и не уплощаются на этапе построения признаков; напротив, они явно моделируются, типизируются, версионируются и доступны для запросов внутри графа. Такой подход сохраняет структурные зависимости, придающие корпоративным данным смысл, включая договорные связи, консультационные отношения, финансовые зависимости и ссылки на документы, что делает возможными многоскачковый логический вывод и контекстную агрегацию.

Не менее важно и то, что происхождение данных и их временная семантика сохраняются на всех этапах конвейера. Каждая сущность и каждое отношение сопровождаются метаданными об источнике, времени извлечения, интервале действия и уровне уверенности. Это позволяет рассматривать результирующий граф знаний не просто как артефакт интеграции, а как управляемое и аудируемое представление, пригодное для аналитических процессов с высокими требованиями к надежности.



Рисунок 2. Архитектура системы

Прием и предварительная обработка данных. Корпоративные данные поступают из гетерогенных источников, различающихся по схемам, форматам и частоте обновления. Структурированные данные обычно включают записи из CRM-систем, баз workflow и систем учета финансовых транзакций, тогда как неструктурированные данные состоят из текстовых документов, таких как договоры, отчеты, электронные письма и внутренние меморандумы.

На этапе приема структурированные данные отображаются в каноническую схему с помощью процедур выравнивания схем и нормализации. Этот шаг устраняет

синтаксические несоответствия, например различия в соглашениях об именовании и типах данных, и обеспечивает согласованное представление основных идентификаторов, таких как клиенты, дела, организации и физические лица, во всех источниках.

Неструктурированные документы проходят предварительную обработку, включающую извлечение текста, определение языка, сегментацию на предложения и токенизацию. Метаданные, такие как тип документа, время создания, авторство и права доступа, сохраняются и связываются с представлением документа, чтобы позднее использоваться для контекстной фильтрации.

Семантическое извлечение из неструктурированных данных. Чтобы интегрировать неструктурированный текст в граф знаний, необходимо извлечь из него семантическую информацию в структурированной форме. Это достигается с помощью конвейера обработки естественного языка, который выявляет сущности, отношения и релевантные атрибуты в документах.

Модели распознавания именованных сущностей (Named Entity Recognition, NER) используются для обнаружения специфичных для предметной области типов сущностей, таких как физические лица, организации, клиенты, юрисдикции и договорные понятия. Методы извлечения отношений определяют семантические связи между сущностями, например связи между людьми и организациями или отсылки между документами и бизнес-сущностями.

Извлеченные сущности не рассматриваются как изолированные упоминания; каждая сущность обогащается контекстными сигналами, полученными из окружающего текста, метаданных документа и лингвистических признаков. Эти сигналы впоследствии используются для дизамбигуации и сопоставления сущностей между источниками.

Разрешение и сопоставление сущностей между источниками.

Постановка задачи. Разрешение сущностей между источниками направлено на выявление случаев, когда сущности, происходящие из структурированных систем E_s, и сущности, извлеченные из неструктурированного текста E_u, обозначают один и тот же объект реального мира.

Пусть: $E_S = \{e_1^s, ..., e_n^s\},\ E_u = \{e_1^u, ..., e_m^u\}$

Тогда целью является построение отображения: $\Phi : E_S \bigcup E_u \to E_c$

где E_c обозначает канонические сущности в объединенном графе.

Генерация кандидатов. Для обеспечения масштабируемости сопоставление выполняется в два этапа. Сначала формируются пары-кандидаты с высоким recall с использованием эвристик блокировки:

- нормализованного сравнения строк;
- общих идентификаторов, таких как домены электронной почты и адреса;
- совместной встречаемости в одном документе или сделке;
- пересекающихся структурированных атрибутов.

Это позволяет сократить квадратичное пространство сравнений до вычислительно приемлемого подмножества.

Построение гибридных признаков. Для каждой пары-кандидата (ei,ej) вычисляется вектор признаков x_ij, основанный на трех независимых группах сигналов.

Лексическое сходство включает:

- нормализованное расстояние редактирования;
- пересечение токенов;
- раскрытие аббревиатур.

Семантическое сходство включает:

- контекстные эмбеддинги, полученные из упоминаний сущностей;
- косинусное сходство между векторами эмбеддингов;
- агрегированные контекстные представления, полученные из разделов документов.

Реляционная согласованность включает:
- общих соседей в частичной структуре графа;
- совместимые типы отношений;
- метрики сходства окрестностей.

Реляционный сигнал формализует интуицию о том, что сущности с похожими реляционными окрестностями с высокой вероятностью являются идентичными.

Сопоставление на основе уверенности. Пары-кандидаты оцениваются с помощью обучаемой или эвристической функции:

$s(e_i,e_j)=f(x_{ij})$

Далее применяются пороговые правила принятия решений:
- высокая уверенность - автоматическое слияние;
- средняя уверенность - мягкая связь, при которой сущности сохраняются раздельно, но соединяются;
- низкая уверенность - отсутствие слияния.

Такой механизм позволяет сбалансировать цену ложных объединений и пропущенных совпадений.

Сохранение происхождения данных. При слиянии сущностей исходные узлы источников сохраняются как ссылки на происхождение. Каждая каноническая сущность хранит:
- исходные записи;
- уровень уверенности сопоставления;
- временные метки извлечения.

Это обеспечивает аудируемость и обратимость.

Схема графа знаний и его построение. Граф знаний строится как размеченный граф свойств, в котором узлы представляют сущности, а ребра – типизированные отношения. Схема задается через легковесную онтологию, охватывающую ключевые предметные понятия и их связи, оставаясь при этом достаточно гибкой для учета развивающихся источников данных.

Узлы сущностей хранят атрибуты, происходящие как из структурированных систем, так и из текстовой информации, извлеченной из документов. Ребра отношений кодируют как явные связи, например отношения между клиентом и проектом, так и выведенные отношения, полученные из текста или графового вывода.

Построение графа выполняется инкрементально, что позволяет добавлять новые данные без необходимости полного пересчета. Механизмы версионирования обеспечивают отслеживание изменений сущностей и отношений во времени, поддерживая аудируемость и воспроизводимость аналитических результатов.

Граф следует модели размеченного графа свойств.

Типы узлов (иллюстративный сценарий для рынка недвижимости). Граф знаний реализован как размеченный граф свойств, в котором узлы представляют канонизированные сущности предметной области, охватывающие транзакционные записи, организационных участников, документы, финансовые атрибуты и индикаторы риска. Ребра представляют типизированные семантические отношения, отражающие структурные, договорные и контекстные зависимости между сущностями.

В единой схеме представлены как отношения из структурированных систем, так и ассоциации, извлеченные из текста.

Каждое ребро хранит метаданные о происхождении, временной валидности и уровне уверенности. Это позволяет графу поддерживать версионирование, анализ с учетом

неопределенности и воспроизводимую аналитику, одновременно сохраняя связь с исходными структурированными записями и документными источниками.

Рисунок 3 показывает, каким образом гетерогенные элементы, происходящие из структурированных CRM-систем и неструктурированных информационных меморандумов, объединяются в рамках единой реляционной модели.

Транзакционные записи, финансовые метрики, ссылки на документы и раскрытия рисков представлены как взаимосвязанные сущности, а не как изолированные артефакты данных. Такая конструкция позволяет явно выражать и единообразно анализировать такие отношения, как финансовые зависимости, консультативные роли, связи с арендаторами или ковенантные ограничения.

Важно отметить, что граф не просто воспроизводит структуры исходных данных; он реифицирует семантические отношения как объекты первого класса с собственными метаданными. За счет присоединения к каждому отношению атрибутов происхождения, временной валидности и уверенности граф поддерживает аналитику с учетом неопределенности и принятие решений с возможностью трассировки. В результате последующие запросы и модели машинного обучения работают с представлением, которое сохраняет как структурные зависимости, так и доказательную основу.

Типы узлов:

- объект недвижимости
- сделка / листинг
- организация (продавец, покупатель, кредитор, брокер)
- физическое лицо (команда сделки, подписанты)
- местоположение (адрес, город, субрынок)
- договор аренды / арендное соглашение
- арендатор (может быть подтипом)
- финансовый показатель (NOI, ставка капитализации, DSCR, IRR)
- документ (инвестиционный меморандум, договор аренды, отчет об оценке)
- фактор риска (экологический риск, риск вакантности, долговые ковенанты)

Связи:

- расположен в
- предлагается как
- ринадлежит / продается
- представлен
- управляется
- сдан в аренду
- имеет договор аренды
- имеет показатель
- финансируется
- подпадает под ковенант
- имеет фактор риска
- ссылается на *(для подтверждения источника)*

Рисунок 3. Структура графа знаний

Слой графового вывода и аналитики. Граф знаний поддерживает два взаимодополняющих режима анализа:

- детерминированные многоскачковые графовые запросы;
- статистическое обучение с использованием представлений, производных от графа.

Обход графа и запросы на сопоставление шаблонов позволяют выявлять косвенные отношения, агрегировать свидетельства из документов и отслеживать взаимодействия сущностей во времени.

Параллельно структура графа преобразуется в численные представления посредством:

- локальных структурных статистик, таких как степень вершины и количество соседей разных типов;
- признаков, основанных на путях и реляционном контексте;

- представлений на основе эмбеддингов, обученных методами графовых эмбеддингов или графовых нейронных сетей.

Эти признаки используются в последующих задачах классификации, ранжирования и предсказания связей.

Графовые запросы и символический вывод. Граф поддерживает многоскачковые запросы и структурное сопоставление шаблонов, что позволяет решать задачи логического вывода, такие как обнаружение косвенных связей, агрегирование свидетельств из документов или отслеживание взаимодействий сущностей во времени. Эти запросы выполняются непосредственно над топологией графа и семантикой ребер, обеспечивая детерминированные и объяснимые результаты.

Генерация признаков на основе графа. Для поддержки моделей машинного обучения структура графа преобразуется в численные представления. Для каждого узла-сущности его признаки могут включать:

- локальные структурные статистики, такие как степень и состав окрестности;
- признаки на основе путей, отражающие реляционный контекст;
- векторы эмбеддингов, обученные на структуре графа.

Графовые эмбеддинги обучаются методами, сохраняющими близость и реляционное сходство, что позволяет сущностям с похожими ролями или контекстами располагаться близко друг к другу в пространстве эмбеддингов. Такие эмбеддинги служат компактными представлениями сложной реляционной информации.

Последующие задачи машинного обучения. Графовые признаки интегрируются в модели машинного обучения с учителем и без учителя для решения таких задач, как классификация, ранжирование и обнаружение аномалий. Важно, что использование графовых признаков часто повышает обобщающую способность моделей за счет учета реляционного контекста, отсутствующего в плоских признаковых представлениях.

Предсказания, формируемые этими моделями, могут быть соотнесены с лежащими в их основе структурами графа, что позволяет проводить post hoc объяснение путем анализа влиятельных узлов, ребер или путей. Это особенно важно для компаний сектора профессиональных услуг, где требуются прозрачность и аудируемость.

Оценка. Мы оцениваем предлагаемый подход, сравнивая эффективность аналитики при использовании и без использования слияния на основе графа знаний. В качестве метрик рассматриваются precision и recall для разрешения сущностей, точность классификации в задаче поиска экспертизы, а также контекстная релевантность, оцениваемая экспертами-аннотаторами. Предварительные результаты показывают заметные улучшения при использовании объединенных графов, подтверждая ценность семантической интеграции.

Цели оценки. Оценка направлена на анализ эффективности слияния данных на основе графа знаний по трем направлениям:

- качество разрешения сущностей;
- влияние на последующие аналитические задачи;
- качественные улучшения в контекстном логическом выводе.

Вместо концентрации на абсолютных значениях метрик акцент делается на сравнительных улучшениях относительно базовых подходов, обрабатывающих структурированные и неструктурированные данные независимо.

Оценка разрешения сущностей. Качество сопоставления сущностей оценивается с использованием вручную подготовленных эталонных соответствий для репрезентативного подмножества сущностей. Используются стандартные метрики:

- precision - для измерения числа ошибочных слияний;
- recall - для измерения числа пропущенных соответствий;
- F1-мера - для сбалансированной оценки обоих аспектов.

Результаты показывают, что включение семантических и реляционных признаков существенно улучшает качество сопоставления по сравнению с базовыми подходами,

использующими только лексические признаки, особенно в случаях неоднозначных или сокращенных упоминаний сущностей.

Оценка последующей аналитики. Для оценки влияния слияния данных на аналитические задачи графовые признаки используются в моделях машинного обучения с учителем для репрезентативных задач, таких как классификация экспертизы и предсказание отношений.

Качество измеряется с помощью метрик, соответствующих конкретной задаче, включая точность классификации и качество ранжирования. Модели, использующие граф-обогащенные представления, последовательно превосходят базовые модели, обученные только на структурированных или только на неструктурированных данных, что демонстрирует ценность интегрированной контекстной информации.

Качественный анализ. Помимо количественных метрик, проводится качественный анализ путем изучения путей логического вывода внутри графа знаний.

Эти анализы показывают, что объединенные представления позволяют осуществлять многоскачковый вывод через документы и структурированные записи, поддерживая сложные запросы, которые невозможно реализовать в изолированных конвейерах обработки данных. Такая интерпретируемость особенно важна в среде профессиональных услуг, где аналитики и предметные эксперты должны иметь возможность проверять и обоснованно доверять результатам системы.

Ограничения. Проведенная оценка имеет ряд ограничений.

Разметка ground truth для сопоставления доступна только для части сущностей, а показатели эффективности могут варьироваться в предметных областях с иными соглашениями об именовании или другой структурой документов. Тем не менее наблюдаемые тенденции последовательно свидетельствуют в пользу графового слияния по сравнению с базовыми подходами.

Заключение и направления дальнейшей работы. Мы представили архитектуру слияния данных на основе графа знаний, интегрирующую структурированные и неструктурированные корпоративные данные для обеспечения более глубокой аналитики и контекстного интеллекта.

В дальнейшем планируется исследовать интеграцию данных в режиме реального времени, федеративные графы знаний, охватывающие несколько организаций, а также более тесную интеграцию с большими языковыми моделями для интерактивного логического вывода.

Список литературы

[1] G. Tamašauskaitė, P. Groth. Defining a Knowledge Graph Development Process Through a Systematic Review (February 2023).

[2] Philippe Cudré-Mauroux. Leveraging Knowledge Graphs for Big Data Integration: the XI Pipeline (January 2020).

[3] Beibei Zhu, Ruolin Wang, Junyi Wang, Fei Shao, Kerun Wang. A survey: knowledge graph entity alignment research based on graph embedding (August 2024).

[4] Xiaojuan Zhao, Yan Jia, Aiping Li, Rong Jiang, Yichen Song. Multi-source knowledge fusion: a survey (April 2020).

[5] Chenwei Yan, Xinyue Fang, Xiaotong Huang, Chenyi Guo, Ji Wu. A solution and practice for combining multi-source heterogeneous data to construct enterprise knowledge graph (September 2023).

[6] Maraim Masoud, Bianca Pereira, John McCrae, Paul Buitelaar. Automatic Construction of Knowledge Graphs.

Авторский вклад

Е.А. Алуев – анализ существующих решений и общее руководство проектом.

М.А. Булычева – архитектурный анализ и разработка.

SEMANTIC DATA FUSION BASED ON KNOWLEDGE GRAPHS: INTEGRATION OF STRUCTURED AND UNSTRUCTURED DATA TO UNLEASH CONTEXTUAL INTELLIGENCE

Eugene Alooeff
R&D Engineer ATEK,
Bachelor of Computer Science
alooeff@atek.dev

Mariia Bulycheva
Engineer ATEK,
Master of Mathematics
mbulytcheva@gmail.com

Abstract. Professional services companies work with heterogeneous data, where structured systems coexist with large volumes of unstructured documents. Processing them separately limits the acquisition of context-sensitive and cross-source insights needed for tasks such as conflict detection, expertise search, and risk assessment. This paper proposes a knowledge graph-based data fusion architecture that combines structured and unstructured enterprise data into a single semantic representation while preserving data lineage. Natural language processing methods are used to extract entities and relationships from documents, and entity resolution across sources is accomplished using a hybrid strategy combining lexical, semantic, and relational matching. The resulting annotated feature graph supports logical inference and feature generation for machine learning. Results demonstrate that this approach improves the quality of entity matching, enhances the efficiency of subsequent analytics, and ensures greater interpretability. Knowledge graph-based semantic integration thus provides a scalable and efficient foundation for enterprise analytics in heterogeneous data environments.

Keywords: knowledge graph; data integration; unstructured data; entity mapping; semantic analysis; enterprise analytics.

УДК 004.622:004.654

МОДИФИКАЦИЯ АЛГОРИТМОВ НОРМАЛИЗАЦИИ ТАБЛИЦ БАЗ ДАННЫХ ДЛЯ ПРИМЕНЕНИЯ В СИСТЕМАХ BIG DATA





А.А. Карпук
Профессор кафедры программного обеспечения сетей телекоммуникаций Белорусской государственной академии связи, кандидат технических наук, доцент
a_karpuk@mail.ru

Л.С. Лазута
Аспирант кафедры телекоммуникационных систем Белорусской государственной академии связи, магистр
lenya.lazuta@mail.ru

А.А. Карпук

Окончил Белорусский государственный университет, автор более 260 опубликованных научных трудов, включая 3 монографии. Область научных интересов связана с моделированием и оптимизацией сложных систем, проектированием баз данных и хранилищ данных, оценкой качества радиосвязи и оптимизацией присвоения радиочастот в радиосетях и радиолиниях.

Л.С. Лазута

Окончил Белорусскую государственную академию связи, имеет 8 опубликованных научных трудов. Область научных интересов связана с моделированием и оптимизацией сложных систем, проектированием баз данных и хранилищ данных.

Аннотация. Рассмотрены классические алгоритмы нормализации таблиц в реляционных базах данных. Показаны недостатки алгоритмов, ограничивающие их применение для нормализации таблиц в системах Big Data. Предложены модификации алгоритмов Делобеля-Кейси, Бернштейна и Ислура, которые имеют полиномиальную сложность и могут применяться в системах Big Data для автоматического приведения таблиц данных к третьей нормальной форме.

Ключевые слова: большие данные, функциональная зависимость, третья нормальная форма, алгоритмы нормализации таблиц, минимальное покрытие, замыкание атрибутов, ключ таблицы.

Введение. Современные системы Big Data могут иметь архитектуру реляционного хранилища данных (Relational Data Warehouse, RDW), озера данных (Lake Data), современного хранилища данных (Modern Data Warehouse, MDW), фабрики данных (Data Fabric), озерного хранилища данных (Data Lakehouse), сетки данных (Data Mesh) [1]. Во всех перечисленных архитектурах, кроме архитектуры Lake Data, на одном или нескольких этапах обработки данных решаются задачи очистки данных, предварительной обработки данных и нормализации данных в виде приведения данных к третьей нормальной форме (3НФ). Возможные подходы к решению задач очистки и предварительной обработки данных рассмотрены в работах [2, 3]. Для решения задачи нормализации данных требуется знание функциональных зависимостей (ФЗ), существующих между данными. В работе [4] рассматривались методы выделения и алгоритмы поиска ФЗ между данными в системах Big Data. В настоящей работе рассмотрены классические алгоритмы приведения таблиц реляционной базы данных к 3НФ Делобеля-Кейси, Бернштейна, Ислура и Неклюдовой-Цаленко, показаны недостатки алгоритмов, ограничивающие их применение для нормализации таблиц данных в системах Big Data, и предложены модификации алгоритмов Делобеля-Кейси, Бернштейна и Ислура, которые могут применяться в системах Big Data для автоматического приведения таблиц данных к 3НФ.

Классические алгоритмы нормализации таблиц баз данных. Пусть $X \subset A$ – подмножество атрибутов (признаков) системы Big Data, ZA – некоторый атрибут. Говорят, что существует ФЗ XZ, если любой комбинации значений атрибутов из X всегда соответствует единственное значение атрибута Z. Очевидно, что из $Z \in X$ следует, что XZ. Такая ФЗ, в которой зависимый атрибут входит в состав левой части ФЗ, называется тривиальной. В дальнейшем будем рассматривать только нетривиальные ФЗ между атрибутами. Структура ФЗ на множестве атрибутов удовлетворяет аксиомам Армстронга [5]. Для задания структуры ФЗ, отличающейся от тривиальной, требуется постулировать конечное множество ФЗ $F=\{Fj=XjYj|Xj,YjA,j=1,m\}$, которое называется системой образующих структуры ФЗ на A. Структуру ФЗ, заданную системой образующих F, будем обозначать $S(F)$. Замыканием множества XA относительно структуры ФЗ $S(F)$ называется множество $X+(F)A$, такое, что для любого YA из XY следует $YX+(F)$.

Структуры ФЗ $S(F1)$ и $S(F2)$ на множестве A с системами образующих $F1=\{Xi1Yi1|Xi1 \subset A,Yi1 \subseteq A,i=1,m1\}$ и $F2=\{Xj2Yj2|Xj2 \subset A,Yj2 \subseteq A,j=1,m2\}$ соответственно называются эквивалентными, если для любого $X \subset A$ имеет место равенство $X+(F1)=X+(F2)$. В работе [6] было доказано, что структуры ФЗ $S(F1)$ и $S(F2)$ на множестве A эквивалентны тогда и только тогда, когда для всех $i=1,m1$ выполняются условия $Xi1+(F1)=Xi1+(F2)$ и для всех $j=1,m2$ выполняются условия $Xj2+(F1)=Xj2+(F2)$.

Система образующих $E=\{HjTj|Hj,TjA,j=1,m\}$ структуры ФЗ $S(E)$ на множестве A называется минимальным покрытием (minimal coverage), если выполняются следующие условия:

для любого $j1,m$ и любого XTj система образующих E', полученная из E путем замены ФЗ $HjTj$ на ФЗ $HjTj\backslash X$, задает структуру ФЗ $S(E')$, не эквивалентную структуре ФЗ $S(E)$;

для любого $j1,m$ и любого YHj система образующих E', полученная из E путем замены ФЗ $HjTj$ на ФЗ $Hj\backslash YTj$, задает структуру ФЗ $S(E')$, не эквивалентную структуре ФЗ $S(E)$.

В русскоязычной литературе минимальное покрытие иногда называют элементарным базисом. Классический алгоритм Делобеля-Кейси [7] состоит из следующих шагов.

1. Разложить ФЗ так, чтобы в правой части каждой ФЗ был только один атрибут. Каждая ФЗ XjYj из F, где Yj={Aj1, Aj2, ...}, заменяется на несколько ФЗ с одним атрибутом в правой части XjAj1, XjAj2, ….

2. Удалить избыточные атрибуты в левых частях полученных ФЗ. Атрибут BXj является избыточным в левой части ФЗ XjAj1, если Aj1(Xj\B)+. При построении замыкания рассматриваемая ФЗ не учитывается.

3. Удалить избыточные ФЗ из полученного множества ФЗ. ФЗ XjAj1 является избыточной, если после ее удаления выполняется условие Aj1(Xj)+.

4. Все полученные ФЗ с одинаковыми левыми частями собираются в одну ФЗ, в правой части которой находится объединение атрибутов правых частей собранных ФЗ.

5. На основе каждой полученной ФЗ формируется таблица, содержащая все атрибуты, находящиеся в левой и правой части этой ФЗ. Первичным ключом таблицы объявляется множество атрибутов левой части ФЗ.

6. Если среди полученных таблиц найдутся две таблицы, такие что все атрибуты первой таблицы входят в множество атрибутов второй таблицы, то первая таблица удаляется. При этом проверка ФЗ всех неключевых атрибутов первой таблицы от первичного ключа этой таблицы должна выполняться при обработке строк второй таблицы.

7. Построить замыкания первичных ключей полученных таблиц относительно полученной структуры ФЗ. Если ни одно из замыканий не совпадает с множеством атрибутов A, то найти любой ключ этого множества атрибутов и добавить к полученным таблицам в 3НФ таблицу, состоящую из атрибутов найденного ключа. Для поиска ключа применяется алгоритм последовательного удаления атрибутов из списка всех атрибутов и построения замыкания оставшихся атрибутов. Если удаленный атрибут не входит в полученное замыкание, то он входит в состав ключа и возвращается в список атрибутов.

Делобель и Кейси доказали, что полученные таблицы находятся в 3НФ, и исходная таблица восстанавливается из полученных таблиц с помощью операций естественного соединения. Легко убедиться, что шаги 1-4 алгоритма Делобеля-Кейси строят минимальное покрытие исходной структуры ФЗ на множестве атрибутов A.

Классический алгоритм Бернштейна [8] состоит из следующих шагов.

1-4. Построить минимальное покрытие структуры ФЗ аналогично шагам 1-4 алгоритма Делобеля-Кейси.

5. Построить замыкание левой части каждой из полученных ФЗ и разбить ФЗ на группы таким образом, чтобы в одной группе были все ФЗ с совпадающими замыканиями левых частей.

6. На основе каждой полученной группы ФЗ формируется таблица, содержащая все атрибуты, находящиеся в левых и правых частях всех ФЗ группы. Левая часть одной из ФЗ группы (любой) объявляется первичным ключом таблицы, левые части остальных ФЗ объявляются уникальными потенциальными ключами таблицы.

7. При формировании таблиц возможен случай, когда в таблице, содержащей более одного ключа, неключевой атрибут, попавший в таблицу из правой части одной ФЗ (ключа), окажется в частичной ФЗ от правой части другой ФЗ, попавшей в таблицу.(другого ключа). Такие атрибуты удаляются из таблицы. Для поиска таких атрибутов последовательно рассматриваются неключевые атрибуты таблицы, каждый атрибут временно удаляется из таблицы и из той ФЗ, из которой он попал в таблицу, и строится замыкание оставшихся атрибутов таблицы. Если рассматриваемый атрибут принадлежит этому замыканию, то его действительно надо удалить из таблицы, в противном случае он возвращается в таблицу.

8. Этот шаг аналогичен шагу 6 алгоритма Делобеля-Кейси.

9. Этот шаг аналогичен шагу 7 алгоритма Делобеля-Кейси.

Бернштейн доказал, что полученные таблицы находятся в 3НФ, и исходная таблица восстанавливается из полученных таблиц в 3НФ с помощью операций естественного

соединения. Классический алгоритм Ислура [9] совпадает с алгоритмом Бернштейна, за исключением того, что на шаге 9 для поиска ключа всего множества атрибутов используется усовершенствованный алгоритм. Из исходного списка всех атрибутов сразу удаляются все атрибуты, не входящие в левые части исходных ФЗ, а затем оставшиеся атрибуты проверяются на вхождение в ключ таким же образом, как в алгоритмах Делобеля-Кейси и Бернштейна.

На первых шагах классического алгоритма Неклюдовой–Цаленко [10] также строится минимальное покрытие, затем находятся все ключи таблицы, и множество всех ключевых атрибутов таблицы рассматривается как полная подструктура структуры ФЗ. Биекции всех ключей включаются в систему образующих структуры ФЗ, в результате чего в минимальном покрытии могут появиться транзитивные ФЗ, которые удаляются. Производится расширение подструктуры путем добавления неключевых атрибутов, функционально полно зависящих от ключей. Если все атрибуты таблицы войдут в формируемую подструктуру, то она находится в 3НФ. В противном случае анализируются ФЗ, оставшиеся в минимальном покрытии. Если в них входят не все атрибуты таблицы, то множество атрибутов этих ФЗ рассматривается как полная подструктура структуры ФЗ, для которой все шаги алгоритма повторяются, иначе применяется алгоритм Бернштейна.

Модификация алгоритмов для применения в системах Big Data. Сразу отметим, что поиск всей ключей таблицы, используемый в алгоритме Неклюдовой-Цаленко, является NP-трудной задачей, поэтому применение алгоритма Неклюдовой-Цаленко для автоматической нормализации таблиц систем Big Data, содержащих сотни атрибутов и сотни исходных ФЗ между атрибутами, невозможно.

В алгоритмах Делобеля-Кейси, Бернштейна и Ислура имеется шаг поглощения меньшей таблицы большей таблицей, если все атрибуты меньшей таблицы входят в состав большей таблицы (шаг 6 в алгоритме Делобеля-Кейси). Это фактически ведет к потере ФЗ минимального покрытия, на которой построена меньшая таблица. В базах данных поддержка таких ФЗ производится средствами СУБД с помощью специально написанных триггеров. Большинство проектировщиков баз данных по этой причине не выполняют шаг 6 алгоритма Делобеля-Кейси и шаг 8 алгоритмов Бернштейна и Ислура. Лучше иметь в базе данных лишнюю таблицу, чем решать проблему с поддержкой ФЗ между атрибутами. Мы считаем, что при использовании алгоритмов Делобеля-Кейси, Бернштейна и Ислура для автоматической нормализации таблиц в системах Big Data этот шаг следует исключить из алгоритмов.

Во всех рассмотренных алгоритмах при построении минимального покрытия исходной структуры ФЗ сначала все ФЗ раскладываются на ФЗ, содержащие один атрибут в правой части, а после удаления избыточных атрибутов в левых частях и избыточных ФЗ снова собираются в ФЗ с одинаковыми левыми частями. Применяемый алгоритм удаления избыточных атрибутов в левых частях и избыточных ФЗ основан на том, что в правой части ФЗ находится один атрибут. В работе [11] Карпук предложил алгоритм построения минимального покрытия структуры ФЗ, в котором не требуется раскладывать ФЗ на ФЗ, содержащие один атрибут в правой части, а затем собирать ФЗ с одинаковыми левыми частями. Этот алгоритм состоит из следующих шагов.

1. Удалить избыточные атрибуты из левых частей ФЗ из F. Атрибут BXj называется избыточным в Xj, если B(Xj\B)+(F'), где через F' обозначена система образующих структуры ФЗ, полученная из F путем удаления ФЗ XjYj.

2. Если в результате получим в F две или более ФЗ с одинаковыми левыми частями, то объединить их в одну ФЗ.

3. Удалить избыточные атрибуты из правых частей ФЗ из F. Атрибут BYj называется избыточным в Yj, если B(Xj)+(F'), где через F' обозначена система образующих структуры ФЗ, полученная из F путем замены ФЗ XjYj на XjY\Bj.

4. Удалить из F ФЗ с пустыми правыми частями.

Карпук доказал, что описанный алгоритм строит минимальное покрытие структуры ФЗ, совпадающее с минимальным покрытием, построенным по шагам 1-4 алгоритма Делобеля-Кейси. Отказ от разложения ФЗ по одному атрибуту в правой части позволяет избежать увеличения мощности множества ФЗ и приводит к снижению вычислительной сложности построения минимального покрытия с $O(m2k2ln)$ до $O(m2(l+k)n)$, где n – количество атрибутов, m – количество ФЗ, k – среднее количество атрибутов в правой части ФЗ, l – среднее количество атрибутов в левой части ФЗ. Предложенный подход обеспечивает асимптотическое ускорение, особенно заметное при наличии ФЗ с большой мощностью правой части, что характерно для систем Big Data. Еще более заметно преимущество алгоритма Карпука для построения минимального покрытия структуры ФЗ при оценке объема оперативной памяти, необходимой для реализации алгоритма. Для реализации алгоритма Делобеля-Кейси объем требуемой оперативной памяти оценивается величиной $O(mkl)$, а для реализации алгоритма Карпука величиной $Om(k+l)$. Таким образом, при использовании алгоритмов Делобеля-Кейси, Бернштейна и Ислура для автоматической нормализации таблиц в системах Big Data вместо шагов 1-4 для построения минимального покрытия структуры ФЗ следует использовать алгоритм Карпука.

В таблицах систем Big Data количество атрибутов, не входящих в левые части ФЗ, может быть очень большим, поэтому при использовании алгоритмов Делобеля-Кейси, Бернштейна и Ислура для автоматической нормализации таблиц в системах Big Data для поиска ключа всего множества атрибутов следует использовать алгоритм Ислура.

Список литературы

[1] Serra James. Deciphering Data Architectures. Choosing Between a Modern Data Warehouse,Data Fabric, Data Lakehouse, and Data Mesh. O'Reilly 2024. – 252 p.

[2] Лазута Л.С. Методы и алгоритмы очистки и предварительной обработка данных в хранилищах данных // Новые информационные технологии в телекоммуникациях и почтовой связи: материалы XXV Междунар. науч.-техн. конф., 13-14 мая 2025 г. / редкол.: А.О. Зеневич [и др.]. – Минск: Белорусская государственная академия связи, 2025. – С. 153–154.

[3] Lazuta L.S., Karpuk A.A. Cleaning and Preprocessing of Data in Data Warehouses // International Journal of Engineering Research and Development. − 2025. − Vol. 21, Issue 9. − P. 263−269.

[4] Карпук А.А., Лазута Л.С. Выделение и поиск функциональных зависимостей между атрибутами в системах Big Data // 11-ая Международная научно-технической конференция «BIG DATA and Advanced Analytics», 23-24 апреля 2025 г. БГУИР, Минск, Беларусь. – С. 397–405.

[5] Armstrong W.W. Dependency Structure of Data Base Relationships // Proc. IFIP Congress. – Geneva, Switzerland, 1974. – P. 580–583.

[6] Карпук А.А. Выбор элементарного базиса структуры функциональных зависимостей при проектировании базы данных // Вопросы радиоэлектроники. Сер. ОВР. – 1983. – Вып. 6. – С. 38–41.

[7] Delobel C., Casey R.G. Decomposition of a data base and the theory Boolean switching functions // IBM J. Res. And Dev. – 1973. – Vol. 17, No 5. – P. 374–386.

[8] Bernstein P.A. Synthesizing third normal form relations from functional dependencies // ASM Transactions on Database Systems. – 1976. – Vol. l, No 4. – P. 277–298.

[9] Isloor S.S. An algorithm with logical simplicity for designing third normal form relational database schema from functional dependencies // Proc. of Int. Conf. on DBMSs (ICMOD 78). – Fast Milan, Italy, 1978. – P. 31–50.

[10] Неклюдова Е.А., Цаленко М.Ш. Синтез логической схемы реляционной базы данных // Программирование. – 1979. – № 6. – С. 58–68.

[11] Карпук А.А. О построении элементарного базиса системы функциональных зависимостей в базе данных // Информационные технологии и программные средства: проектирование, разработка и применение: Сб. научн. ст. – Гродно: ГрГУ, 2011. – С. 185–190.

Авторский вклад

Карпук Анатолий Алексеевич – руководство исследованием алгоритмов автоматической нормализации таблиц в таблицах систем Big Data, постановка задач исследования, оценка результатов исследования и определение направлений дальнейшей работы по модификации алгоритмов.

Лазута Леонид Сергеевич – анализ алгоритмов автоматической нормализации таблиц в таблицах систем Big Data, определение области применения алгоритмов, разработка предложений по модификации алгоритмов.

MODIFICATION OF DATABASE TABLE NORMALIZATION ALGORITHMS FOR USE IN BIG DATA SYSTEMS

A.A. Karpuk
Professor, Department of Telecommunication Network Software, Belarusian State Academy of Communications, PhD of Technical sciences, Associate Professor

L.S. Lazuta
Postgraduate student, Department of Telecommunication Systems, Belarusian State Academy of Communications, Master´s Degree

Abstract. Classic algorithms for normalizing tables in relational databases are examined. The algorithms' shortcomings that limit their use for normalizing tables in Big Data systems are highlighted. Modifications to the Delobel-Casey, Bernstein, and Isloor algorithms are proposed. These algorithms have polynomial complexity and can be used in Big Data systems to automatically convert data tables to third normal form.

Keywords: big data, functional dependency, third normal form, table normalization algorithms, minimum coverage, attribute closure, table key.

УДК 004.032.26:621.865.8.

ПРИНЦИПЫ И АЛГОРИТМЫ ПОСТРОЕНИЯ НЕЙРОСЕТЕВЫХ СИСТЕМ УПРАВЛЕНИЯ КООРДИНАТАМИ НИЖНЕГО УРОВНЯ В МОБИЛЬНОЙ РОБОТОТЕХНИКЕ



А. А. Радкевич
Младший научный сотрудник лаборатории робототехнических систем, ГНУ «Объединённый институт проблем информатики Национальной академии наук Беларуси»
a_radkevich@newman.bas-net.by

А.А. Радкевич
Младший научный сотрудник лаборатории робототехнических систем ГНУ «Объединённый институт проблем информатики Национальной академии наук Беларуси». Область научных интересов связана с разработкой методов и алгоритмов управления мобильными роботами колёсного типа и системами управления электроприводов в мобильной робототехнике.

Аннотация. В статье представлена систематизация нейросетевых подходов к управлению координатами нижнего уровня в мобильной робототехнике. На основе анализа литературных источников разработана иерархическая классификация, учитывающая тип колёсной базы (классические и всенаправленные колёса), тип двигателя (коллекторный ДПТ, вентильный БДПТ/СДПМ, асинхронный) и регулируемую координату. Для каждого сочетания факторов определены предпочтительные архитектуры нейросетей, методы обучения и типы контроллеров. Сформулированы рекомендации по выбору подходов в гибридных структурах, где нейросетевые регуляторы сочетаются с гибридными подходами.

Ключевые слова: мобильная робототехника, нейросетевые системы управления, нижний уровень управления, классические колёса, всенаправленные колёса, дифференциальный привод, меканум-колёса, омни-колёса, гибридные системы управления, нейро-ПИД-регулятор, скользящий режим, адаптивное управление, рекуррентные нейронные сети, радиально-базисные функции.

Введение. Точность и быстродействие мобильных роботов, их устойчивость к изменениям внешней среды и энергетическая эффективность в значительной степени определяются характеристиками системы управления его ходовой части. В иерархической структуре управления мобильными роботами выделяется нижний уровень, который представляет собой совокупность локальных регуляторов, непосредственно формирующих управляющие сигналы для электродвигателей и обеспечивающих поддержание регулируемых координат: угловой скорости, вращательного момента, угла положения вала. Качество отработки этих внутренних координат определяет возможность реализации задающих воздействий траектории, поступающих от верхних (навигационных) уровней [1, 2].

Конструктивные требования ходовой части мобильных платформ накладывают определенные требования на системы управления координатами нижнего уровня. В мобильных роботах выделяется два основных типа колёсных движителей: классические колёса и всенаправленные колёса (меканум-колёса и омни-колёса) [3, 4]. Оба типа колёс реализуются на основе дифференциального привода, однако кинематика всенаправленных платформ сложнее и требует учёта перекрёстных связей. Для классических колёс критически важна синхронизация ведущих колёс и компенсация нелинейностей (сухое и вязкое трение, упругие деформации трансмиссии) [5, 6]. Всенаправленные платформы, способные перемещаться в любом направлении без изменения ориентации, требуют согласованного управления всеми колёсами с учётом перекрёстных кинематических связей и высокой точности позиционирования каждого привода [7, 8]. Дополнительным фактором, определяющим сложность построения систем нижнего уровня, является тип используемого электродвигателя колёс: коллекторные электродвигатели постоянного тока (ДПТ) относительно просты в управлении, но имеют ограниченный ресурс эксплуатации, вентильные электродвигатели – бесколлекторные электродвигатели постоянного тока (БДПТ) и синхронные электродвигатели с постоянными магнитами (СДПМ) – обладают нелинейными электромагнитными характеристиками и перекрёстными связями между осями управления, асинхронные электродвигатели (АД) требуют управления потокосцеплением и также обладают выраженными нелинейностями [2, 9].

Классические методы управления координатами электродвигателей – ПИД-регулирование, управление скользящим режимом (SMC), модельное прогнозирующее управление (MPC) – хорошо изучены и находят широкое применение благодаря простоте реализации и возможности аналитического расчёта параметров [10, 11]. Однако их эффективность снижается в условиях, характерных для мобильных роботов: нелинейности (сухое трение, насыщение, упругие связи), параметрическая неопределённость (изменение момента инерции, проскальзывание колёс), многосвязность (взаимное влияние приводов). Регуляторы, настроенные на заданный режим, при отклонении от расчётных условий теряют качество или становятся неустойчивыми [1, 5, 12]. Линейное MPC требует точной модели, а нелинейное MPC – значительных вычислительных ресурсов, что ограничивает его применение на контроллерах систем управления [2].

Альтернативой классическим методам управления выступают алгоритмы, основанные на применении искусственных нейронных сетей. В отличие от классических подходов, они не требуют явного аналитического описания объекта, аппроксимируя сложные нелинейные зависимости на основе обучающих данных [13]. Среди их преимуществ – способность адаптироваться к изменяющимся условиям, компенсировать нелинейности и работать с многомерными системами без настройки взаимных связей [1, 2, 14]. Для платформ с колёсами всенаправленного типа нейросетевые регуляторы эффективно компенсируют перекрёстные кинематические связи [7, 8]. Вместе с тем нейросетевые методы имеют ограничения: потребность в больших объёмах данных, высокую вычислительную сложность, сложности с обеспечением устойчивости и интерпретацией результатов [14, 15]. Кроме того, большинство исследований

сосредоточено на регулировании частоты вращения вентильных и асинхронных электродвигателей, тогда как нейросетевое управление их током и потокосцеплением остаётся малоизученным [9].

Компромиссным подходом являются гибридные структуры, сочетающие сильные стороны классических и нейросетевых принципов управления. В таких системах нейросетевые компоненты активируются только при необходимости, а в стационарных режимах управление осуществляется традиционными регуляторами [1]. Это позволяет сохранить гарантии устойчивости классического контура и одновременно адаптироваться к неопределённостям. Примерами таких систем выступают: комбинация сети на радиально-базисных функциях (RBF) со скользящим режимом (SMC) [6, 10]; каскадное соединение модельного прогнозирующего управления (MPC) и скользящего режима с нейронаблюдателем на основе экстремального машинного обучения (ELM) [8]; нейро-нечёткий пропорционально-дифференциальный (ПД) регулятор [4]; нейро-ПИД с адаптивной настройкой коэффициентов [16, 17].

Исходя из изложенного, в данной статье решается задача систематизации существующих нейросетевых и гибридных подходов к управлению координатами нижнего уровня (угловая скорость, угловое положение, ток статора) в мобильной робототехнике с учётом трёх ключевых факторов: типа движителя (классические либо всенаправленные колёса), типа электродвигателя (коллекторный ДПТ, вентильный БДПТ/СДПМ, АД) и регулируемой координаты. В отличие от обзорных работ, рассматривающих нейросетевые методы управления электроприводами в общем виде [1, 2, 14], в данном исследовании акцентируется внимание на специфике мобильных роботов, а также на конкретных архитектурах нейросетей, методах обучения и их привязке к кинематическим схемам. На основе анализа литературных источников формулируются рекомендации по выбору подходов и определяются направления для дальнейших исследований.

Существующие подходы к проектированию систем регулирования координат нижнего уровня на основе методов искусственных нейронных сетей.

Базовые подходы к нейросетевому управлению электроприводами

Исследования в области нейросетевого управления электродвигателями условно можно разделить на две группы: работы, ориентированные на общие задачи регулирования без привязки к конкретной кинематике мобильных роботов, и работы, учитывающие особенности колёсных платформ. Фундаментальные разработки первой группы заложили основу для последующего применения в мобильной робототехнике.

Одним из ранних направлений стало использование нейросетевых регуляторов с предсказанием. В работе [10] предложен нейросетевой регулятор с предсказанием NARMA-L2 для двухмассной системы привода подачи токарного станка с асинхронным двигателем. Регулятор обучен офлайн по алгоритму Левенберга–Марквардта и обеспечивает стабилизацию угловой скорости с компенсацией упругих колебаний. В структуре управления используется одноконтурная схема, где нейросеть напрямую формирует управляющее напряжение, а классические регуляторы отсутствуют.

В работе [12] предложен комбинированный подход, в котором нейросеть обучается имитировать поведение синергетического регулятора. Обучение выполняется офлайн на гармонических сигналах, после чего обученная сеть (многослойный перцептрон, MLP) используется как прямой регулятор скорости двигателя постоянного тока (ДПТ). В этом случае нейросеть полностью заменяет аналитический закон управления, а классические регуляторы не применяются.

Гибридные структуры, сочетающие нейросети с классическими методами, представлены в работе [5], где разработан регулятор на основе сети радиально-базисных функций (RBF) и скользящего режима (SMC) для коллекторного двигателя постоянного тока. RBF-сеть аппроксимирует неизвестную нелинейную динамику, а закон скользящего режима обеспечивает робастность. Адаптация весов сети выполняется онлайн по закону,

выведенному из условия устойчивости по Ляпунову. Регулируемой координатой является угловое положение вала; классические регуляторы не используются, управление формируется гибридным контроллером. В работе [11] аналогичный гибридный подход применён для СДПМ. Здесь многослойный перцептрон (MLP) обучается онлайн по теории переменной структуры и выступает в качестве регулятора угловой скорости. Внутренние контуры тока остаются классическими (ПИ-регуляторы), что позволяет сочетать адаптивность нейросети с быстродействием традиционных регуляторов.

Альтернативой нейросетевым подходам служит адаптивный регулятор на рядах Тейлора, предложенный в работе [6] для двигателя постоянного тока в составе неголономного колёсного робота. Регулятор построен по двухконтурной схеме: внешний контур (скорость и положение) реализован на рядах Тейлора, а внутренний контур тока – классический ПИ-регулятор. Координаты – скорость и положение – отрабатываются адаптивным аналитическим регулятором без использования нейросетей.

В работе [8] предложен квазинейрорегулятор – структура выходного нейрона, веса которого вычисляются аналитически, без итерационного обучения. Объектом является двухмассная электромеханическая система (тип двигателя не конкретизирован, но используется модель с упругой связью). Регулируется скорость второй массы. Классические регуляторы не применяются, управление строится на основе аналитически рассчитанных весов.

Новый подход с использованием физико-информированных нейронных сетей (PINNs) представлен в работе [9] для обобщённой модели двухзвенного манипулятора и системы «масса–пружина–демпфер» (тип двигателя не указан). PINNs обучаются офлайн, а затем на их основе осуществляется адаптивная настройка коэффициентов ПИД-регулятора в реальном времени. Таким образом, нейросеть выполняет функцию идентификации, а управление реализуется классическим регулятором (гибридная структура).

Обзорные работы [1, 2, 14, 15] систематизируют эти подходы, подчёркивая, что для линейных систем предпочтительны традиционные регуляторы, а нейросетевые методы эффективны в условиях неопределённости. Особо выделяются гибридные структуры, в которых нейросети выполняют вспомогательные функции (идентификация, адаптивная настройка, компенсация), а базовое управление остаётся за классическими регуляторами.

Значительный вклад в развитие систем управления для мобильных роботов внесли работы Е. О. Мешковского [7, 8], в которых разработана математическая модель четырёхколёсного робота с двумя дифференциальными приводными блоками на основе метода Ньютона–Эйлера, синтезированы нечёткий регулятор стабилизации положения приводного блока и нелинейная система согласованного управления.

Систематизация подходов по типам движителей, двигателей и регулируемых координат

Анализ литературы позволяет выделить устойчивые закономерности в применении нейросетевых методов для нижнего уровня в зависимости от кинематической схемы робота, типа двигателя и регулируемой координаты. В рамках гибридных структур часто наблюдается разделение: нейросетевые регуляторы отвечают за одну координату (чаще скорость), а другие координаты (ток, потокосцепление) регулируются классическими методами.

Для платформ с классическими колёсами основное внимание уделяется регулированию угловой скорости ведущих колёс. В работах [13] и [16] для коллекторных двигателей постоянного тока предложены нейро-ПИД-регуляторы. В первом случае используется многослойный перцептрон (MLP), обученный офлайн, который полностью заменяет традиционный регулятор (одноконтурная структура). Во втором – однонейронная сеть с онлайн-обучением по правилу Хебба, также выполняющая функции прямого регулятора скорости. В обеих работах управление током не рассматривается.

Для компенсации упругих колебаний в двухмассных системах в работе [12] применены рекуррентные нейронные сети (Элман) в качестве наблюдателя состояния. Объектом служит асинхронный двигатель (в составе привода оптико-механического комплекса). Нейросетевой наблюдатель оценивает упругий момент и скорость нагрузки, которые затем используются в классическом ПИД-регуляторе положения. Здесь нейросеть выполняет вспомогательную функцию (оценка), а управление остаётся за традиционным регулятором (гибридная структура).

Специализированный прямой нейроконтроллер NETMORC [17] предназначен для коллекторных двигателей дифференциального привода. Регулятор обучается без учителя и непосредственно формирует управляющие сигналы скорости колёс. Классические регуляторы отсутствуют.

Гибридный подход RBF+SMC [5, 11] применительно к коллекторному ДПТ позволяет регулировать как угловую скорость, так и угловое положение. В этой структуре RBF-сеть аппроксимирует нелинейную динамику, а скользящий режим обеспечивает робастность. Классические регуляторы не используются.

Для вентильных двигателей (БДПТ/СДПМ) в составе классических колёс в работе [11] предложен адаптивный нейрорегулятор на основе MLP с онлайн-обучением по теории переменной структуры. Регулируется угловая скорость; внутренние контуры тока реализованы классическими ПИ-регуляторами (гибридная структура). В работе [18] для бесколлекторного двигателя постоянного тока (BLDC) разработан нейро-ПИД с BP-нейросетью, также ориентированный на регулирование скорости. Нейросеть полностью заменяет ПИД-регулятор (одноконтурная схема).

Для асинхронных двигателей в классической колёсной базе единственным примером служит нейросетевой регулятор с предсказанием NARMA-L2 [10], регулирующий угловую скорость. Внутренние контуры тока и потока остаются классическими, что позволяет отнести эту схему к гибридным.

Регулирование углового положения для классических колёс рассматривалось в работах [5, 6]. В первом случае используется адаптивный регулятор на рядах Тейлора (аналитический, без нейросетей), во втором – гибридный RBF+SMC, оба для коллекторных ДПТ.

Регулирование тока для классических колёс в проанализированных источниках отсутствует; токовые контуры, как правило, реализованы классическими ПИ-регуляторами, даже если нейросеть управляет скоростью (гибридные структуры). Это объясняется высокими требованиями к быстродействию контура тока и сложностью обеспечения устойчивости при нейросетевом управлении.

Для всенаправленных платформ в проанализированной литературе используются исключительно вентильные двигатели (БДПТ/СДПМ). Регулируемой координатой является угловая скорость колёс, причём регуляторы строятся по двухуровневой схеме: верхний кинематический уровень формирует желаемые скоростей колёс, нижний уровень – нейросетевые регуляторы отслеживают эти задания.

В работах [19, 20, 21] разработаны адаптивные нейросетевые регуляторы скорости с онлайн-обучением по методу фильтрованной ошибки. Архитектуры – MLP, RBF и ELM соответственно. Внутренние контуры тока реализованы классическими ПИ-регуляторами (гибридные структуры). В работе [21] предложена каскадная структура: верхний уровень – модельное прогнозирующее управление (MPC), нижний уровень – скользящий режим с нейронаблюдателем на основе ELM (Extreme Learning Machine). Нейронаблюдатель оценивает неопределённости, а управление током и скоростью осуществляется классическими методами (гибрид). В работе [21] используется MLP в качестве прямого регулятора скорости, без внутренних классических контуров. В работе [21] применён нейро-нечёткий ПД-регулятор с однослойной сетью; веса настраиваются онлайн. Регулируется угловая скорость, токовые контуры классические (гибрид).

Таким образом, для всенаправленных платформ характерно преимущественное использование гибридных структур, где нейросеть отвечает за регулятор скорости или наблюдатель, а токовые контуры остаются классическими. Прямые нейроконтроллеры (без внутренних классических контуров) встречаются реже.

В таблице 1 представлены наиболее репрезентативные работы, в которых реализовано нейросетевое или гибридное управление координатами нижнего уровня. В таблице учтены тип колёсной базы, тип двигателя, регулируемая координата, тип контроллера, архитектура нейросети и метод обучения.

Таблица 1. Сравнительный анализ нейросетевых подходов к управлению координатами нижнего уровня в мобильных роботах

Тип колёс	**Тип двигателя**	**Регулируемая координата**	**Тип контроллера**	**Архитектура нейросети**	**Метод обучения**	**Источник**
Классические	ДПТ	Угловая скорость	Нейро-ПИД-регулятор	Однонейронная сеть	Онлайн	[16]
Классические	ДПТ	Угловая скорость	Прямой нейроконтроллер	NETMORC	Обучение без учителя	[17]
Классические	ДПТ	Угловая скорость, угловое положение	Гибридный (RBF+SMC)	RBF	Онлайн	[5, 11]
Классические	ДПТ	Угловое положение	Адаптивный регулятор	-	Онлайн	[6]
Классические	БДПТ/ СДПМ	Угловая скорость	Адаптивный нейрорегулятор	MLP	Онлайн	[11]
Классические	БДПТ/ СДПМ	Угловая скорость	Нейро-ПИД-регулятор	MLP	Офлайн (обратное распространение)	[18]
Классические	АД	Угловая скорость	Нейросетевой регулятор с предсказанием	NARMA-L2	Офлайн	[10]
Всенаправленные	БДПТ/ СДПМ	Угловая скорость	Адаптивный нейрорегулятор	MLP, RBF, ELM	Онлайн	[19, 20, 21]
Всенаправленные	БДПТ/ СДПМ	Угловая скорость	Нейро-нечёткий ПД-регулятор	Нейро-нечёткая сеть	Онлайн	[19]
Всенаправленные	БДПТ/ СДПМ	Угловая скорость	MPC+SMC+нейронаблюдатель	ELM	Онлайн	[21]

Выбор методов и архитектур систем управления координатами нижнего уровня на основе нейросетевых алгоритмов в зависимости от типа объекта регулирования. Результаты анализа литературы показывают, что эффективность нейросетевого управления координатами нижнего уровня в мобильной робототехнике определяется согласованным учётом трёх ключевых факторов: типа колёсной базы (классические или всенаправленные колёса), типа электродвигателя (коллекторный ДПТ, вентильный БДПТ/СДПМ, АД) и регулируемой координаты (скорость, положение, ток, момент). На основе обобщения проанализированных работ на рисунке 1 предложена иерархическая схема выбора, которая позволяет систематизировать существующие подходы и обосновать предпочтительные архитектуры и методы обучения.

Для платформ с классическими колёсами, где движение обеспечивается разностью скоростей двух или четырёх ведущих колёс, выбор конкретного нейросетевого подхода зависит от типа двигателя и требований к адаптивности. Для коллекторных ДПТ широко применяются нейро-ПИД-регуляторы, которые могут быть реализованы как в офлайн (MLP, обучение по алгоритму Левенберга–Марквардта), так и в онлайн (однонейронная сеть с правилом Хебба) модификациях [13, 16]. Офлайн-обучение целесообразно при наличии достаточного объёма предварительных данных и стабильных условий эксплуатации; онлайн-адаптация предпочтительна при переменных нагрузках или износе механических частей. Альтернативой выступает прямой нейроконтроллер NETMORC, обучаемый без учителя, который особенно эффективен при необходимости автономной настройки под конкретный робот без участия эксперта. Гибридная структура RBF+SMC рекомендуется при наличии выраженных нелинейностей (сухое трение, упругие связи) и требует онлайн-адаптации весов по закону Ляпунова, что гарантирует устойчивость [5, 11].

Для двухмассовых систем с упругими элементами предпочтительны рекуррентные нейросети в качестве наблюдателей состояния: они оценивают ненаблюдаемые переменные (упругий момент, скорость нагрузки), после чего классический ПИД-регулятор обеспечивает стабилизацию положения. В этом случае нейросеть играет вспомогательную роль, а управление остаётся традиционным [6].

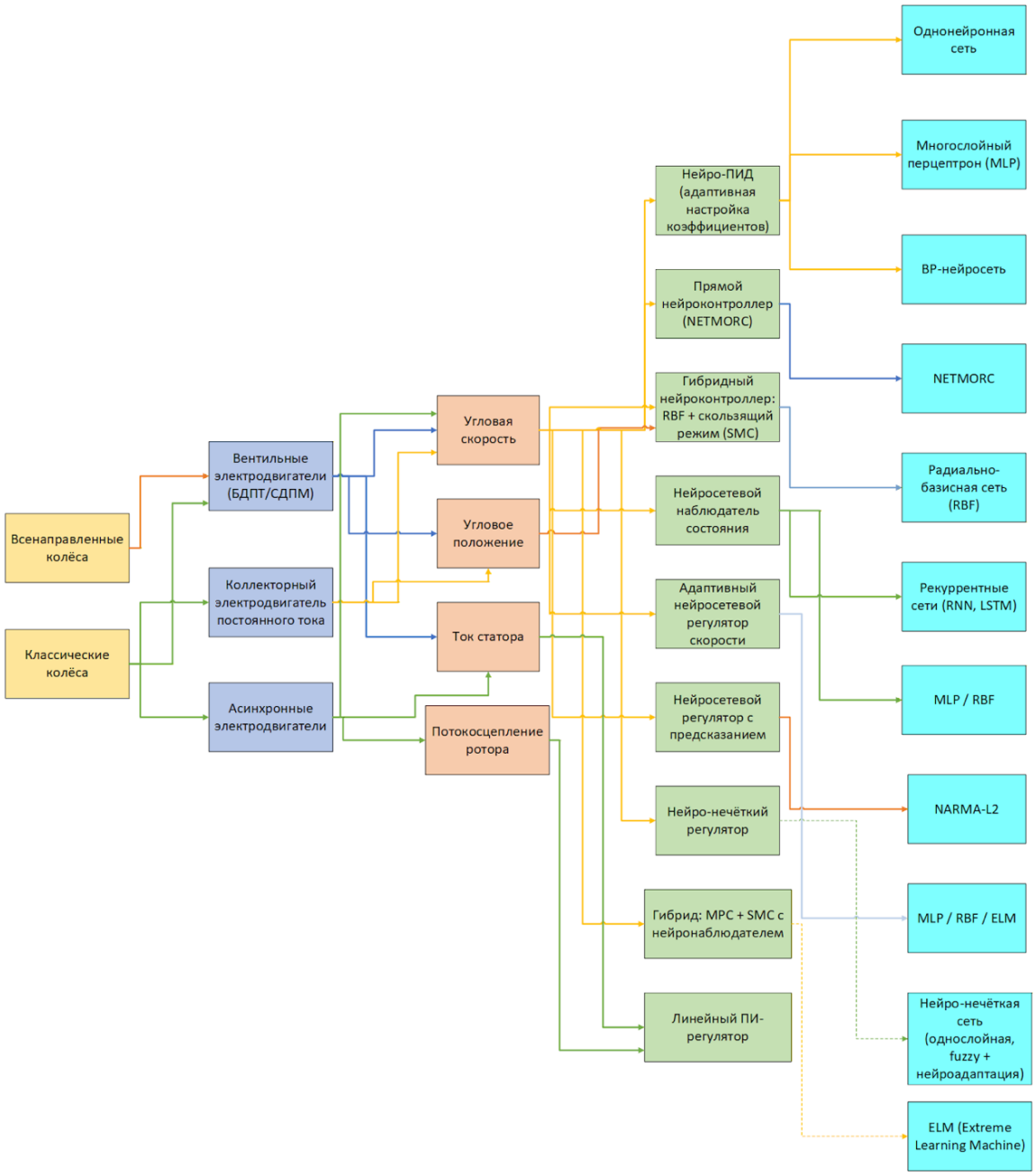


Рисунок 1. Блок-схема выбора методов и архитектур систем управления координатами нижнего уровня на основе нейросетевых алгоритмов в зависимости от типа объекта регулирования

Для вентильных электродвигателей в составе классических колёс доминируют гибридные структуры, в которых нейросеть управляет скоростью, а токовые контуры остаются классическими ПИ-регуляторами. Адаптивный нейрорегулятор на основе MLP с онлайн-обучением по теории переменной структуры обеспечивает высокую робастность и быструю адаптацию к изменению нагрузки без предварительной настройки. Нейро-ПИД с офлайн-обученным MLP может быть использован, если условия эксплуатации близки к

расчётным [18]. В обоих случаях внутренний контур тока реализован классически, что позволяет сохранить быстродействие и простоту реализации.

Для асинхронных электродвигателей в классической колёсной базе наиболее обоснованным является применение нейросетевого регулятора с предсказанием NARMA-L2, обученного офлайн.

Данный подход хорошо зарекомендовал себя для стабилизации скорости при компенсации упругих колебаний. Внутренние контуры тока и потокосцепления остаются классическими, что соответствует типичной промышленной практике [10].

Регулирование тока для классических колёс в проанализированных источниках не реализуется нейросетевыми методами; токовые контуры традиционно остаются классическими (ПИ), что связано с высокими требованиями к быстродействию и сложностью обеспечения устойчивости нейросетевого регулятора тока в реальном времени.

Платформы со всенаправленными колёсами отличаются более сложной кинематикой: движение по любой траектории требует согласованного вращения всех колёс с учётом перекрёстных связей. В проанализированных работах все такие системы используют вентильные двигатели (БДПТ/СДПМ), а регулируемой координатой нижнего уровня является угловая скорость каждого колеса.

Для всенаправленных платформ наиболее распространены адаптивные нейрорегуляторы скорости с онлайн-обучением по методу фильтрованной ошибки. Архитектуры MLP, RBF или ELM (Extreme Learning Machine) обеспечивают высокую скорость обучения и компенсацию кинематических перекрёстных связей.

При этом внутренние контуры тока традиционно реализованы классическими ПИ-регуляторами, что позволяет сохранить быстродействие и упростить реализацию. Такой гибридный подход сочетает адаптивность нейросети на уровне скорости с надёжностью классического регулирования тока.

В качестве альтернативы может использоваться каскадная структура, где верхний уровень реализован модельным прогнозирующим управлением (MPC), а нижний – скользящим режимом с нейронаблюдателем на основе ELM. Нейронаблюдатель оценивает неопределённости и перекрёстные связи, что позволяет снизить вычислительную нагрузку MPC и повысить робастность. Эта схема особенно эффективна при работе с высокими динамическими нагрузками и изменяющимися условиями сцепления.

Нейро-нечёткий ПД-регулятор представляет собой компромиссный вариант, сочетающий интерпретируемость нечёткой логики с обучаемостью нейросетей. Он может быть использован, когда требуется простота настройки и наглядность правил управления.

Прямые нейроконтроллеры (MLP без внутренних классических контуров) также встречаются, но реже, так как они требуют более тщательного обеспечения устойчивости и больших вычислительных ресурсов.

Заключение. На основе систематизации проанализированных литературных источников сформулированы следующие закономерности и рекомендации по выбору нейросетевых методов и архитектур для управления координатами нижнего уровня в мобильной робототехнике:

1. Для классических колёс (дифференциальный привод) с коллекторными двигателями постоянного тока для регулирования угловой скорости целесообразно применение нейро-ПИД-регуляторов (офлайн-обучение MLP при стабильных условиях [13] или онлайн-обучение однонейронной сети по правилу Хебба при переменных нагрузках [16]), прямого нейроконтроллера NETMORC с самообучением [17] либо гибридной структуры RBF+SMC при выраженных нелинейностях [5, 11]. для двухмассных систем с упругими связями рекомендуется использование рекуррентных нейросетей (Элман) в качестве наблюдателей состояния [12]. Для регулирования положения – адаптивный регулятор на рядах Тейлора [6] или гибридный RBF+SMC [5]. Регулирование тока во всех случаях реализуется классическими ПИ-регуляторами.

2. Для классических колёс с вентильными двигателями (БДПТ/СДПМ). основным подходом к регулированию угловой скорости являются гибридные структуры: адаптивный нейрорегулятор на основе MLP с онлайн-обучением по теории переменной структуры (VSS) [11] или нейро-ПИД с офлайн-обученным MLP [18]. Токовые контуры реализованы классическими ПИ-регуляторами, что обеспечивает высокое быстродействие и устойчивость.

3. Для классических колёс с асинхронными двигателями наиболее обоснованным методом регулирования угловой скорости является нейросетевой регулятор с предсказанием NARMA-L2, обученный офлайн [10]. Управление током и потокосцеплением традиционно осуществляется классическими регуляторами.

4. Для всенаправленных колёс с вентильными двигателями регулирование угловой скорости каждого колеса рекомендуется выполнять адаптивными нейрорегуляторами (MLP, RBF, ELM) с онлайн-обучением по методу фильтрованной ошибки в гибридных структурах, где токовые контуры реализованы классическими ПИ-регуляторами [19, 20, 21]. При высоких требованиях к робастности и наличии вычислительных ресурсов эффективны каскадные схемы MPC+SMC с нейронаблюдателем на основе ELM [21]. В качестве более простой альтернативы может применяться нейро-нечёткий ПД-регулятор [21].

Перспективными направлениями дальнейших исследований являются разработка нейросетевых методов для регулирования тока и потокосцепления, а также адаптация рассмотренных подходов к иным кинематическим схемам и типам движителей.

Список литературы

[1] Татауров, А. И. Анализ традиционных и нейросетевых методов управления электроприводами в робототехнике и перспективы гибридных подходов / А. И. Татауров, В. Е. Вавилов // Известия Южного федерального университета. Технические науки. – 2026. – № 1. – С. 1–12.

[2] Kaminski, M. Neural Network Applications in Electrical Drives–Trends in Control, Estimation, Diagnostics, and Construction / M. Kaminski, T. Tarczewski // Energies. – 2023. – Vol. 16, no. 11. – Art. 4441. – 25 p.

[3] Meshkovskiy, E. O. Nonlinear control of electric drive system of four-wheel mobile robot with two differential drive units / E. O. Meshkovskiy, V. Ya. Frolov, A. D. Kurmashev // Journal of Physics: Conference Series. – 2021. – Vol. 1753. – Art. 012031. – 8 p.

[4] Cruz, J. Neuro-fuzzy trajectory tracking control for the nexus 4-wheeled omnidirectional mobile robot / J. Cruz, S. López, M. Llama //2024 XXVI Robotics Mexican Congress (COMRob). – IEEE, 2024. – С. 33-38.

[5] Беззубов, Н. А. Синтез нейрорегулятора устойчивого по замкнутому контуру следящего электропривода на основе адаптивного скользящего режима с применением сетей радиально-базисных функций / Н. А. Беззубов, С. В. Феофилов // Известия Тульского государственного университета. Технические науки. – 2025. – № 1. – С. 61–66.

[6] Haqshenas, A. R. Adaptive control of electrically-driven nonholonomic wheeled mobile robots: Taylor series-based approach with guaranteed asymptotic stability / A. R. Haqshenas, M. M. Fateh, S. M. Ahmadi // Nonlinear Dynamics. – 2024. – Vol. 112. – P. 12345–12360.

[7] Мешковский, Е. О. Построение математической модели четырёхколёсного мобильного робота с двумя дифференциальными приводными блоками / Е. О. Мешковский, А. Д. Курмашев, В. Я. Фролов // Инновации и инвестиции. – 2020. – № 2. – С. 113–118.

[8] Мешковский, Е. О. Проектирование нечёткого регулятора стабилизации движения корпуса колёсного робота с двумя дифференциальными приводными блоками / Е. О. Мешковский // Приоритетные направления инновационной деятельности в промышленности : Сборник научных статей по итогам пятой международной научной конференции, Казань, 30–31 мая 2020 года. Том Часть 1. – Казань: Общество с ограниченной ответственностью "КОНВЕРТ", 2020. – С. 150-154.

[9] Ito, J. Data-Driven Adaptive PID Control Based on Physics-Informed Neural Networks / J. Ito, Y. Wasa // arXiv. – 2025. – 2510.04591v2. – 13 p.

[10] Белов, М. П. Синтез нейросетевого регулятора для двухмассовой электромеханической системы привода подачи токарного станка / М. П. Белов, И. С. Носиров, А. М. Белов // Известия СПбГЭТУ ЛЭТИ. – 2018. – №. 8. – С. 70-76.

[11] Topalov, A. V. Sliding Mode Neuro-Adaptive Control of Electric Drives / A. V. Topalov, G. L. Cascella, V. Giordano, F. Cupertino, O. Kaynak // IEEE Transactions on Industrial Electronics. – 2007. – Vol. 54, no. 2. – P. 671–679.

[12] Белов, М. П. Разработка нейросетевого наблюдателя для оценки координат системы управления электроприводом оптико-механического комплекса / М. П. Белов, Н. В. Лань // Известия Санкт-Петербургского государственного электротехнического университета «ЛЭТИ». – 2021. – № 3. – С. 65–72.

[13] Cerkala, J. Application of neural models as controllers in mobile robot velocity control loop / J. Cerkala, A. Jadlovska // Journal of Electrical Engineering. – 2017. – Vol. 68. – №. 1. – P. 39.

[14] Шевчук, О. С. Побудова енергоефективної системи автоматичного керування мобільного промислового робота / О. С. Шевчук, В. А. Войтенко, В. А. Водічев, О. Г. Калінін // Електротехнічні та комп'ютерні системи. – 2024. – № 39 (115). – С. 6–13.

[15] Цвенгер, И. Г. Применение нейросетевых регуляторов в системах управления электроприводами / И. Г. Цвенгер, И. Р. Низамов // Вестник Казанского технологического университета. – 2017. – Т. 20, № 8. – С. 111–114.

[16] Le Tieu Nien. Toi uu hieu suat toc do dong co Servo DC su dung bo dieu khien PID ket hop mang no-ron / Le Tieu Nien, et al. // REV-ECIT. – 2024. – Vol. 8, no. 2. – P. 45–52.

[17] Gaudiano P. An unsupervised neural network for low-level control of a wheeled mobile robot: noise resistance, stability, and hardware implementation /P. Gaudiano, E. Zalama, J. L. Coronado // IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics). – 1996. – Vol. 26. – №. 3. – P. 485-496.

[18] Вельченко, А. А. Энергоэффективное нейросетевое управление бесколлекторным двигателем постоянного тока / А. А. Вельченко, С. А. Павлюковец, А. А. Радкевич, А. К. Ибрагим // Энергетика: Известия вузов и энергетических объединений СНГ. – 2025. – № 1. – С. 45–53.

[19] López, S. Two-Layer Neuro-Adaptive Compensation Control Applied to a 4-Wheeled Omnidirectional Mobile Robot /S. López, M. Llama // IEEE Latin America Transactions. – 2025. – Vol. 23. – №. 12. – P. 1318-1324.

[20] Burrola, G. Adaptive Neural Network-Based Control for Velocity Regulation in a Four-Wheeled Omnidirectional Mobile Robot with Mecanum Wheels / G. Burrola, S. Lopez, M. A. Leal, J. Mijares // Proceedings of IEEE COMRob. – 2025. – P. 1–7.

[21] Shengping, W. Design of a Cascade Controller of Trajectory Tracking for Omnidirectional AGV Driven by Mecanum Wheels / W. Shengping, S. Yilong, Q. Hongyi // Journal of South China University of Technology. – 2025. – Vol. 53, no. 1. – P. 49–61.

Благодарность

Работа выполнена за счет гранта Белорусского республиканского фонда фундаментальных исследований (договор Т25ИИ М-032, № регистрации 20251198).

PRINCIPLES AND ALGORITHMS FOR CONSTRUCTING NEURAL NETWORK CONTROL SYSTEMS FOR LOWER-LEVEL COORDINATES IN MOBILE ROBOTICS

A.A. Radkevich

Junior Researcher, Laboratory of robotic systems, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus

Abstract. The article presents a systematization of neural network approaches to controlling lower-level coordinates in mobile robotics. Based on an analysis of literature sources, a hierarchical classification is developed that takes into account the type of wheelbase (conventional and omnidirectional wheels), the type of motor (brushed DC motor, brushless DC/permanent magnet synchronous motor, induction motor), and the controlled variable. For each combination of factors, preferred neural network architectures, training methods, and controller types are identified. Recommendations are formulated for the selection of approaches in hybrid structures where neural network controllers are combined with classical methods.

Keywords: mobile robotics, neural network control systems, low-level control, conventional wheels, omnidirectional wheels, differential drive, mecanum wheels, omni wheels, hybrid control systems, neuro-PID controller, sliding mode, adaptive control, recurrent neural networks, radial basis functions.

УДК 004.021:004.75

РАЗРАБОТКА ИНФОРМАЦИОННОЙ СИСТЕМЫ АВТОМАТИЗИРОВАННОГО СБОРА СВЕДЕНИЙ О ПОСЕЩАЕМОСТИ УЧАЩИХСЯ НА ОСНОВЕ АЛГОРИТМОВ РАСПОЗНАВАНИЯ И ИДЕНТИФИКАЦИИ



А.В. Бакулев
Доцент кафедры систем автоматизированного проектирования вычислительных средств, кандидат технических наук, доцент
alex.bakulev@gmail.com



М.А. Бакулева
Доцент кафедры систем автоматизированного проектирования вычислительных средств, кандидат технических наук, доцент
marina.bakuleva@gmail.com

М.А. Бакулева
Доцент кафедры САПР ВС РГРТУ имени В.Ф. Уткина, автор более 90 научных и учебно-методических работ. Область научных интересов: системы искусственного интеллекта, нечеткая логика, машинное обучение.

А.В. Бакулев
Доцент кафедры САПР ВС РГРТУ имени В.Ф. Уткина. Автор более 100 научных и учебно-методических работ. Область научных интересов: структуры и алгоритмы данных, лингвистическое обеспечение система автоматизированного проектирования, конретная математика, структурное программирование.

Аннотация. В работе представлен концепция разработки информационной системы для автоматиированного сбора сведений о посещаемости занятий учениками различных учебных заведений для аналитической обрабоки и исследования вопроса вовлеченности в учебгый процемм и заинтренсованноймть профильтныим дициплинами и общую мотривациб на учебный процкссс, а также возмости автомаатического формирования сообщений родителей с целью своесрепменного информатировани родитлеелй и снидение трудоемкости работы кураторов.

Ключевые слова: идентификации объекта базы данных, распознование образов, алгоритмы кластеризации, кросс-вычисления.

Введение. Сосвременная система среднего и среднего специального образования сталкивается с проблемами «избирательности» учениками предметов, обязательных к посещению. Ученики выбирают для себя важные и «второстепенные» предметы и, соотвественно, посещают только нужные для экзаменов дисциплины. Подобная практика становится глобальной проблемой, которая приводит к тому, что система общего образования искуственно становится узкоспециализированной и, как следствие, кругозор учеников, базовые знания и общекультурные компетенции имеют критический уровень освоения. Очевидно, что вопрос дисциплинарных взысканий не может решаться только силами педагогическких коллективов и профессиональных сообществ, также постоянное оповещение родителей физически не может быть реализовано средствами одного куратора. В данной статье предлагается внедрение системы автоматизированного контроля, что

позволит решить данный вопрос и сведет к минимуму необходтмость адресно обращатьмя к родителяс по поводу выяснения обстоятельсв отсутсвия ученика в школе. В рамках данной работы релизовано программное решение, которое выполняет этап распознавания присутствующих учеников, после этапа идентификации и верификации формируется корректный список отсутствующих.

Далее производится формирование бот-сообщения и рассылка по списку контактов, которые «подгружаюся» из базы данных. Следует отметить, что в данной статье не рассмариваются вопросы технической органиции процеса формирования бот-сообщений, покольку данный ворос не представляет научной новизеы. Практическая новизна данной работы заключается в обеспечении возможности автоматизации учета посещаемости учащимися учебных занятий с возможнстью не только информировать о нарушениях дисциплины, но и проводить аналитические исследования вовлеченности в учебный процесс, и, как следствие, разрабатывать методики мотивации.

Актуальность. Проблема посещаемости учебных заведений, особенно среднего образовательного звена и колледжей, является очень актуальной, поскольку ответственность за плохую учёбу и неуспеваемость часто перекладывается на родителей (законных представителей), которые не обеспечиваютт явку детей на уроки. В тоже время отмечается рост необоснованных пропусков дисциплин (физкультура, история, география, философия) которые часто являются непрофитльными с точки зрания выбранных выпускных экзаменов или специализации колледжа. Очевидно, что в большинстве случаев родители не знают, что ученик пропускает занятие, предусмотренное расписанием, в тоже время классный руководитель или куратор класса не может занимать учебное время сообщениями родителям о том, что ученик плохо посещает занятия или отсутствует на уроке без уважительной причины. Поэтому, очевидна актуальность создания информационной системы, позволяющей автоматически идентифицировать учащихся, которые пропускают занятия и направлять соответствующие сообщения родителям (например, средствами чат-бота). Данная система позволит улучшить посещаемость занятий, повысить дисциплинированность и ответственность учащихся, а также, уменьшить нагрузку педагогов на организацию взаимодействий с родителями, не относящихся к учебному процессу.

Основная идея. В данной статье предлагается коллаборация нескольких инструментариев для системного решения поставленной задачи. Предполагается, что в учебном заведении существует база данных с фотографиями учащихся и контактными данными их законных представителей. В классах установлены камеры видеонаблюдения, что является стандартом безопасности, а также соответствует требованиям проведения промежуточных проверочных испытаний. Цифровые камеры подключены к компьютеру классного руководителя, который имеет доступ к сети, что позволит организовать автоматизированную рассылку по нужным контактам сообщений об отсутствии на уроке того или иного ученика. Программная реализация данного процесса подразумевает интеграцию программной системы распознавания лиц, реализованной сверточной нейронной сетью, с базой данных учащихся. Полученная таким образом информация обрабатывается месенджер-ботом для организации автоматизированной рассылки сообщений родителям.

В качестве исходных данных для системы распознавания воспринимаются изображения учеников, которые присутсвуют на занятии. Программа реализует процесс распознавания учащегося по изображению, полученному с видеокамеры. Заметим, что изображение может быть «плохого» качества (учащийся может быть под определённым углом, плохое освещение и т.п.), поэтому процесс распознавания, реализуемый сверочной нейронной сетью [1, 2], дополняется программным модулем агрегации данных за весть учебный день, чтобы избежать ошибок и недоразумений с точки зрения ошибочных сообщений, адресованным родителям прилежного ученика. Свёрточные нейронные сети

широко используются в обработке изображений; для работы соответствующего программного модуля распознавания изображений требуются входные данные в формате JSON (строки JSON).

Далее полученные изображения распознаются и идентифицируются в соответствии с обученной моделью [4]. Выходными данными этого этапа является файл со строками JSON, которые конвертируются в идентификатор учащегося в нотации той БД, которая используется в учебном заведении (парсинг строки JSON в словарь Python) [3, 5].

```
import json
json_string = '{"name": "Анна", "age": 28, "city": "Москва"}'
data = json.loads(json_string)
print(data)
# {'name': 'Анна', 'age': 28, 'city': 'Москва'}
print(f"Имя: {data['name']}, возраст: {data['age']}")
```

После обработки данных БД будет формироваться отчет о посещаемости занятий. Очевидно, что данный отчет представляет собой «сырые» данные, которые необходимо дополнительно обработать и проанализировать.

Разработанно прраммной системой решаются следующие задачи:

- сопоставляется список полученных идентификаторов с полным списочным составом учащихмя, формируется список потенциальных «прогульщиков», который затем уточняется (сопоставялется с записями об отсутствии по уважительной причине);

-формируется итоговый список, который представляет агрегированные данные за весь учебный день, поскольку не рационально отправлять сообщение родителям во время каждого урока (рисунок 1); после коррекции и отлаживания этого списка, соответствующий чат-бот будет отправлять сообщения в мессенджер родителям или законным представителям ученика о том, что ученик пропустил определенное количество занятий в течение данного учебного дня.

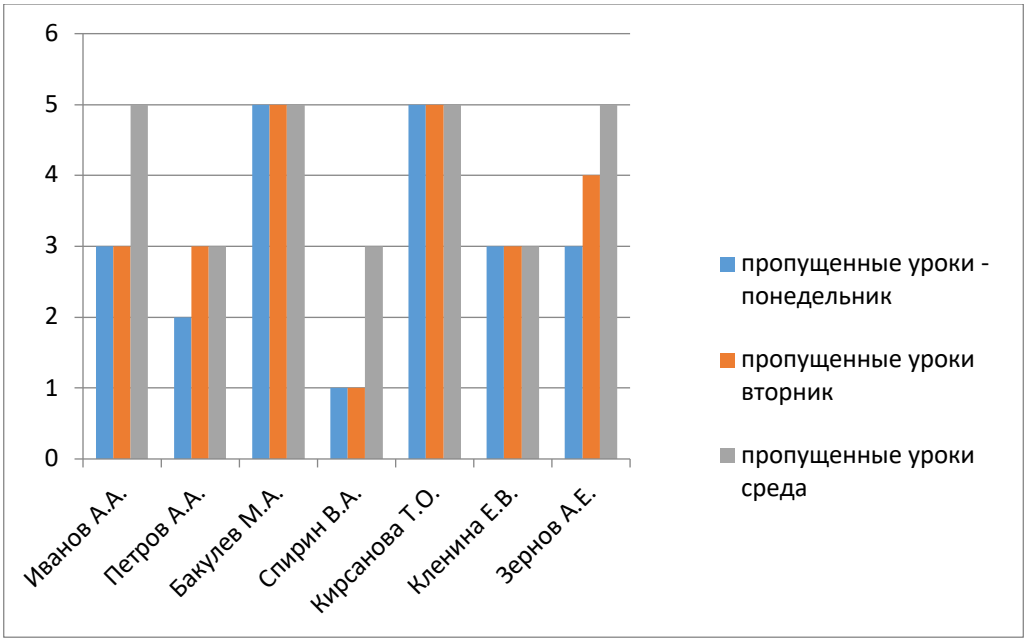


Рисунок 1. Агрегированные данные о посещаемости занятий за весь учебный день

-генерируется аналитический отчет об общей посещаемости занятий, а также визуализируются результаты кластеризации [6], позволяющие оценить вовлеченность учащихся в изучение профильных дисциплин (рисунок 2, рисунок 3)

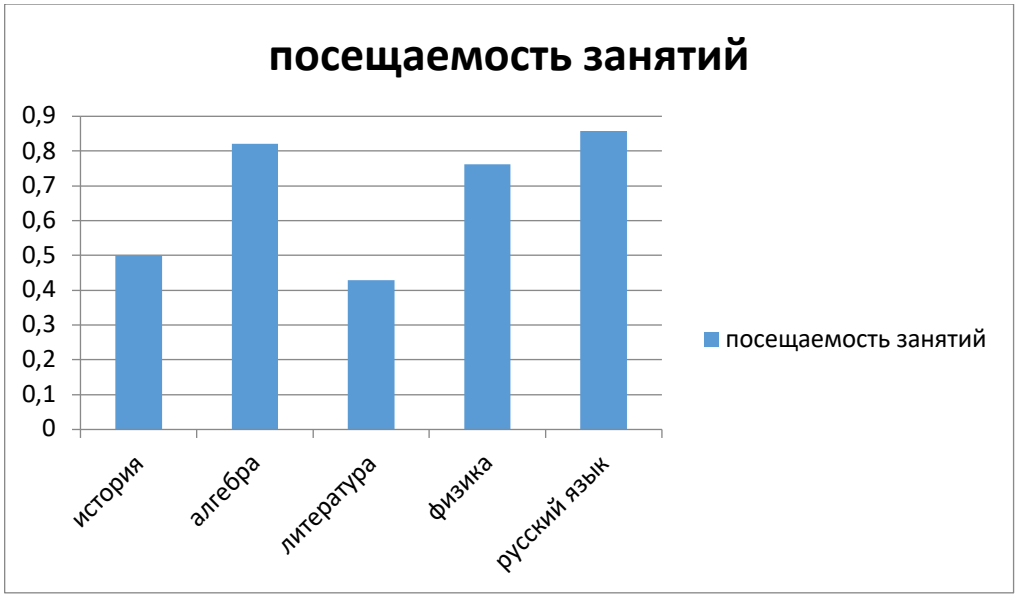


Рисунок 2. Аналитический отчет об общей посещаемости занятий

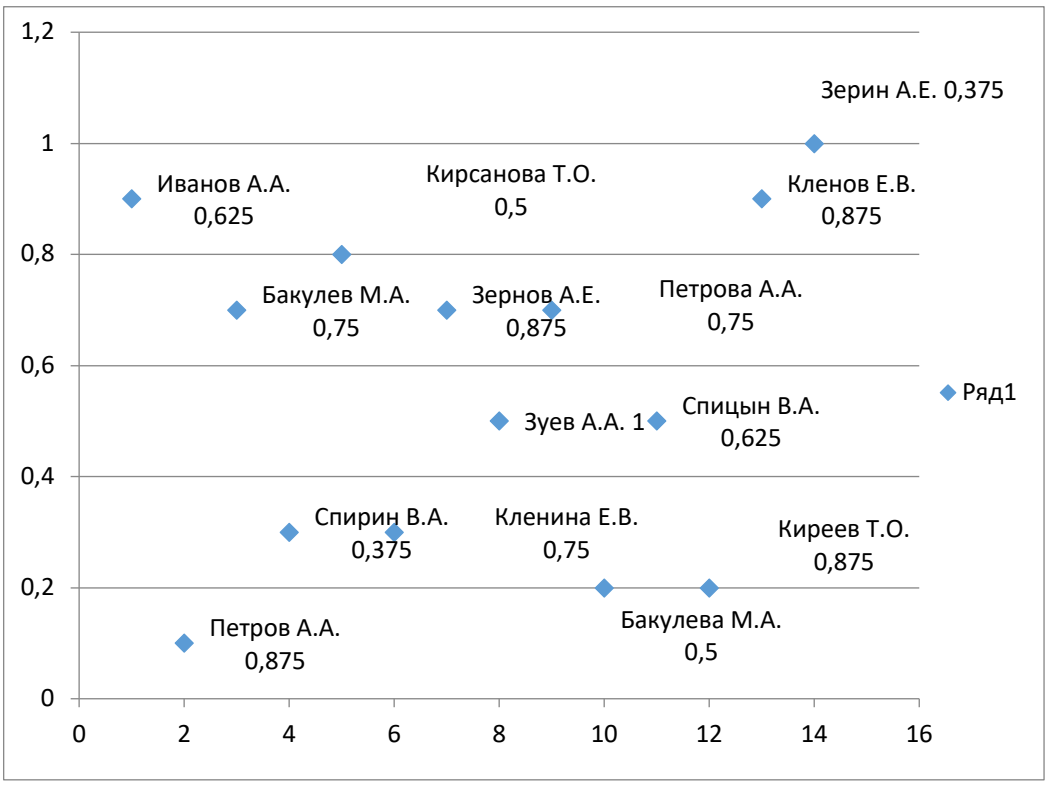


Рисунок 3. Кластерный анализ, отображающий заинтересованность учеников в профильных дисциплинах

Полученные на данном этапе результаты являются алгоритмами кросс вычислений позволющим произвоить агрегирование данных по строкам и столбцам для этого исальзуется многомерное представление (аналитичесий куб), позволяющий менять ведущтй параметр (ось значений) для предоставления срезов зависсимотей.

Для проведения кластеризации был использован алгоритм *K*-средних [7].

Заключение. В данной статье предлагается коллаборация нескольких инструментариев для системного решения задачи разработки информационной системы

для автоматиированного сбора сведений о посещаемости занятий и проведения дальнейшего анализа полученных данных.

Разработанная система позволит не только своевременно информировать родителей о нарушениях дисциплины, но и позволит самим учащимся более ответсвенно подходить к вопросам организации учебы.

В дальнейшем панируется дополнить аналитический компонент средствами визуализации данных о пропущенных занятиях, средних баллах, кластерного и регрессионного анализа для предоставления этой информации непосредственно учащимся. Очевидно, что даннаи информация может более эффективно мотивировать ученика.

Список литературы

[1] Комарова М.А., Коротких И.А. Разработка программного модуля информационной системы для доопределения идентификатора транспортного средства по нечетким исходным данным. В сборнике: IT OPEN 2022. Материалы II регионального конкурса студенческих научно-исследовательских работ в области информационных и вычислительных технологий. Рязанский государственный радиотехнический университет имени В. Ф. Уткина. Рязань, 2022. С. 136-144 [2] Графкина К.В., Куркина М.С. Разработка информационной системы идентификации владельца транспортного средства для учета транзакций по оплате парковки. В книге: Новые информационные технологии в научных исследованиях. Материалы XXIX Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. Рязань, 2024. С. 18.

[3] Бакулева М.А., Бакулев А.В. Разработка алгоритма шифрования биометрических данных. В книге: Новые информационные технологии в научных исследованиях. Материалы XXIX Всероссийской научно-технической конференции студентов, молодых ученых и специалистов. Рязань, 2024. С. 214-215.

[4] Корячко В.П., Бакулев А.В., Бакулева М.А. Алгоритм идентификации транспортных средств в условиях неполных данных. В сборнике: BIG DATA и анализ высокого уровня. сборник научных статей X Международной научно-практической конференции. Минск, 2024. С. 170-174

[5] Бакулева М.А., Елисеев А.В. Разработка информационной системы кодирования и интерпретации биометрических данных Свидетельство о регистрации программы для ЭВМ RU 2025616113, 13.03.2025. Заявка № 2025614219 от 28.02.2025.

[6] Aleksandr Bakulev, Marina Bakuleva, Sergei Skvortsov, Maksim Kozlov, Tatiana Pyurova ,Vladimir Hrukin. Modern approaches to the development parallel programs for modern multicore processors.. Proceedings of 6th Mediterranean Conference on Embedded Computing (MECO), Bar, Montenegro, 2017, pp.38-4

[7] Bakulev A.V., Bakuleva M.A., Avilkina S.B. Mathematical methods and algorithms of mobile parallel computing on the base of multi-core processors // European researcher. 2012. V. 33. № 11-1. P. 1826-1834.

Авторский вклад

Авторы внесли равнозначный вклад.

DEVELOPMENT OF AN AUTOMATED INFORMATION SYSTEM FOR COLLECTING DATA ON STUDENTS' ATTENDANCE BASED ON RECOGNITION AND IDENTIFICATION ALGORITHMS

A.V. Bakulev
Associate Professor of the Department of Computer Aided Design of Computing Facilities Ryazan State Radio Engineering University named after V.F. Utkin (RSREU),
PhD of Technical Sciences, Associate Professor
E-mail: alex.bakulev@gmail.com

M. A.Bakuleva
Associate Professor of the Department of Computer Aided Design of Computing Facilities Ryazan State Radio Engineering University named after V.F. Utkin (RSREU),
PhD of Technical Sciences, Associate Professor
E-mail: marina.bakuleva@gmail.com

Abstract. The paper devote the concept of developing an information system for automated set of information about student attendance at various educational institutions for analytical processing and research on the issue of student engagement in the educational process, as well as the possibility of automatically generating messages for parents in order to provide timely information to parents and reduce the workload of curators.

Keywords: database object identification, pattern recognition, clustering algorithms, and cross-computing..

УДК 004.896:681.51

ИНТЕЛЛЕКТУАЛЬНЫЕ АЛГОРИТМЫ КООРДИНАЦИИ И УПРАВЛЕНИЯ УСТРОЙСТВАМИ «УМНОГО ДОМА»



Д.В. Лихачевский
Заместитель начальника Секретариата Наблюдательного совета ПВТ, кандидат технических наук, доцент
likhacheuski@park.by



С.А. Мигалевич
Магистр технических наук, начальник центра информатизации и инновационных разработок
migalevich@bsuir.by



В.О. Агапова
Магистрант кафедры проектирования информационно-компьютерных систем факультета компьютерного проектирования БГУИР,
agapova06@bsuir.by

Д.В. Лихачевский
Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием проблем радиочастотной идентификации объектов, организацией учебного и научно-исследовательского процессов в техническом университете.

С.А. Мигалевич
Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов: вибродиагностика, разработка метода вейвлет–анализа изделий машиностроения.

В.О. Агапова
Магистрант Белорусского государственного университета информатики и радиоэлектроники, факультета компьютерного проектирования

Аннотация: В статье рассматриваются интеллектуальные алгоритмы координации и управления устройствами «умного дома». Проведён обзор архитектур систем домашней автоматизации, протоколов взаимодействия устройств, а также методов машинного обучения и обучения с подкреплением, применяемых для оптимизации энергопотребления, обеспечения безопасности и повышения комфорта. Описаны перспективы развития интеллектуальных систем домашней автоматизации.

Ключевые слова: умный дом, Интернет вещей, интеллектуальные алгоритмы, машинное обучение, обучение с подкреплением, координация устройств, энергоэффективность, протоколы IoT.

Введение. Концепция «умного дома» за последние годы прошла путь от дорогой экзотики до повседневной реальности. Если раньше автоматизация жилья сводилась к простым таймерам и датчикам движения, то сегодня умные устройства способны общаться друг с другом, анализировать поведение жильцов и самостоятельно принимать решения [1]. По данным *Statista*, глобальный рынок умных домов к 2025 году достиг 150 миллиардов долларов с ежегодным темпом роста около 25 % [2].

Принципиальное отличие современных систем от классической автоматизации состоит в том, что нейросети и алгоритмы машинного обучения не просто выполняют заранее заданные правила, а учатся на реальных данных. Они подстраиваются под образ жизни конкретной семьи и корректируют сценарии работы в реальном времени [3]. Это

позволяет добиться экономии электроэнергии до 30 %, снизить износ техники и повысить комфорт без участия пользователя [4].

Целью данной работы является обзор интеллектуальных алгоритмов, применяемых для координации и управления устройствами умного дома, от протоколов связи до методов глубокого обучения с подкреплением.

Архитектура системы «умного дома». Типичная система умного дома состоит из нескольких уровней. Нижний уровень – это конечные устройства: датчики температуры, влажности, движения, освещённости, а также исполнительные механизмы: реле, клапаны, приводы [5]. Средний уровень – шлюз (gateway), который собирает данные с датчиков и передаёт команды исполнительным устройствам. Верхний уровень – облачная или локальная платформа с алгоритмами обработки данных и принятия решений [6].

Для взаимодействия устройств между собой используются различные протоколы. Zigbee работает в диапазоне 2,4 ГГц, поддерживает mesh-топологию и отличается низким энергопотреблением, именно его используют Aqara, Philips Hue и IKEA [7]. Z-Wave работает на частоте 868 МГц и обеспечивает стабильную связь, хотя и стоит дороже. Thread – новый IP-протокол, созданный специально для стандарта Matter [7].

Matter заслуживает отдельного внимания. Это не физический протокол, а универсальный стандарт совместимости, разработанный совместно Apple, Google, Amazon и Samsung [8]. Он работает поверх Wi-Fi, Ethernet или Thread и позволяет объединять устройства разных производителей без привязки к конкретному облаку. Для подключения достаточно отсканировать QR-код и устройство одновременно появляется во всех экосистемах пользователя [7].

Машинное обучение в управлении умным домом. Нейронные сети обеспечивают превращение набора разрозненных устройств в единую интеллектуальную среду. Основные направления их применения – это анализ сенсорных данных, персонализация сценариев и предиктивное управление [3]. Система собирает информацию с датчиков температуры, освещённости, движения, влажности и шума, а нейросеть выявляет паттерны и прогнозирует потребности жильцов.

Для управления освещением, например, рекуррентные нейронные сети (RNN) эффективно анализируют временные данные: когда жильцы обычно включают свет, в какое время уходят из дома. Свёрточные нейронные сети (CNN) используются для обработки визуальной информации: распознавания жестов или управления световыми сценами. Глубокие нейронные сети (DNN) обеспечивают высокую точность при сложных условиях с множеством переменных [9].

В МГПУ и ряде белорусских и российских университетов уже ведутся исследования по применению нейросетей для прогнозирования потребления ресурсов в зданиях [10]. На практике AI-система вместо жёсткого расписания «выключить свет через 5 минут» анализирует присутствие людей, уровень естественного освещения и привычки каждого члена семьи, после чего сама определяет оптимальный режим [4].

Обучение с подкреплением для оптимизации энергопотребления. Одна из наиболее перспективных задач оптимизация работы систем отопления, вентиляции и кондиционирования (HVAC), которые являются основными потребителями энергии в жилом секторе. Проблема в том, что оптимальное расписание работы HVAC зависит от множества неопределённых параметров: погоды, тарифов на электроэнергию, привычек жильцов и теплофизических свойств здания [11].

Исследователи из Окриджской национальной лаборатории (ORNL) предложили систему RL-HEMS, основанную на обучении с подкреплением, представлена на рисунке 1. Алгоритм обучается на паттернах энергопотребления и генерирует оптимальное расписание работы HVAC, не требуя априорных знаний о теплодинамической модели здания. Эксперименты на реальном доме показали снижение затрат на 17 % и потребления энергии на 15 % [11].

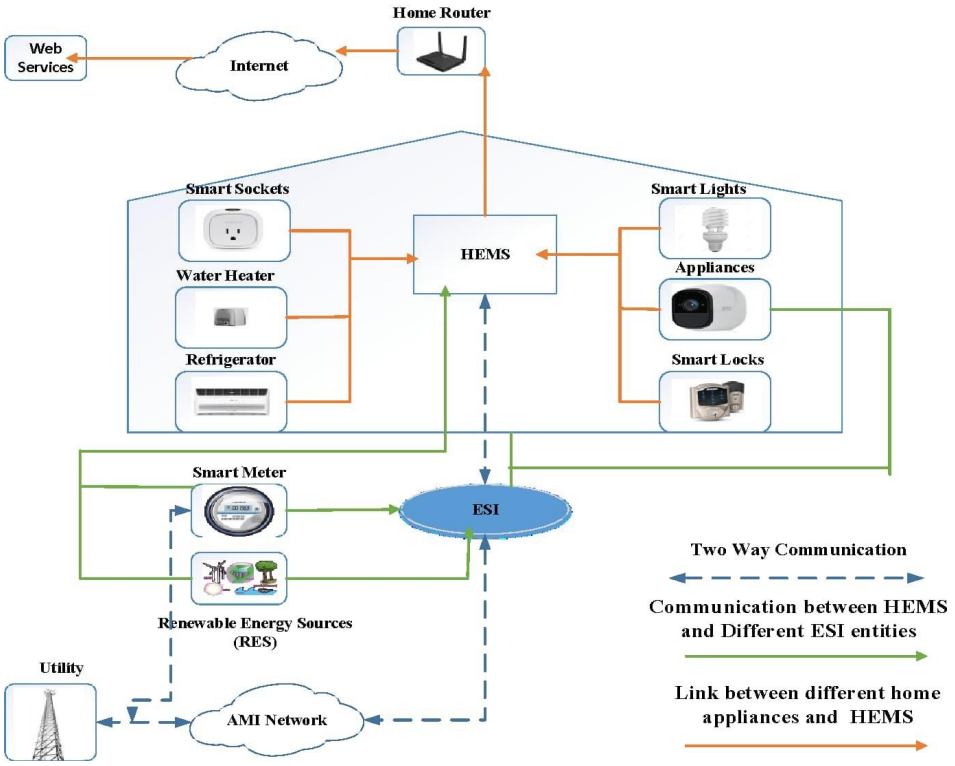


Рисунок 1. Схема HEMS

Другой подход Deep Deterministic Policy Gradients (DDPG), он формализует задачу управления энергией как марковский процесс принятия решений, диаграмма представлена на рисунке 2. Агент одновременно управляет зарядкой/разрядкой системы накопления энергии и мощностью HVAC, учитывая выработку солнечных панелей, наружную температуру и текущий тариф [12]. Результаты на реальных данных подтверждают эффективность и устойчивость такой стратегии.

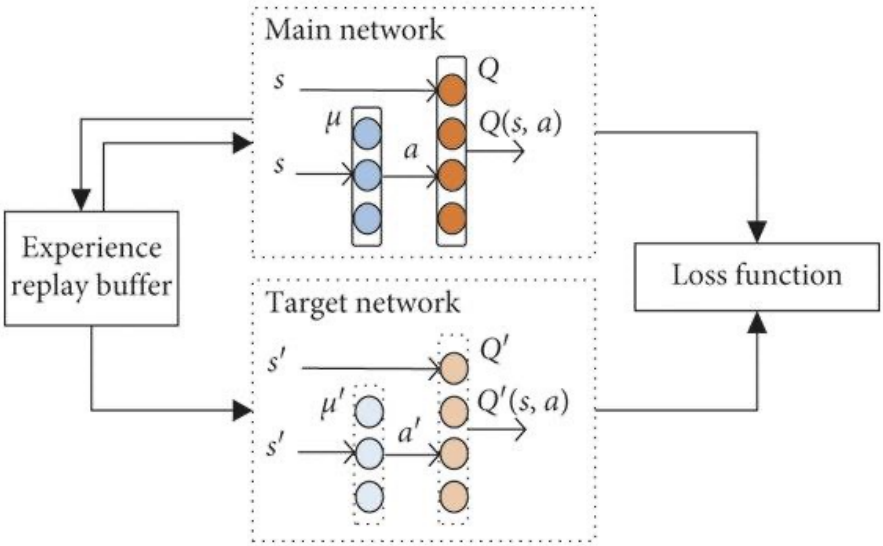


Рисунок 2 . Диаграмма Deep Deterministic Policy Gradients

Интеллектуальные алгоритмы безопасности. Безопасность – ещё одна область, где нейросети существенно расширяют возможности умного дома. Глубокие нейронные сети применяются в системах контроля доступа для распознавания лиц: биометрические данные жильцов заносятся в базу, и при каждом запросе на вход система сверяет двумерное или трёхмерное изображение лица с сохранёнными образцами [13]. Помимо этого, нейросетевые модели используются для идентификации людей по походке – такой подход позволяет реагировать на появление посторонних без необходимости явной аутентификации [14]. Не менее важной является защита самой IoT-сети от внешних атак. IoT архитектура представлена на рисунке 3. Поскольку устройства умного дома компактны и ограничены в ресурсах, они легко становятся мишенью для злоумышленников.

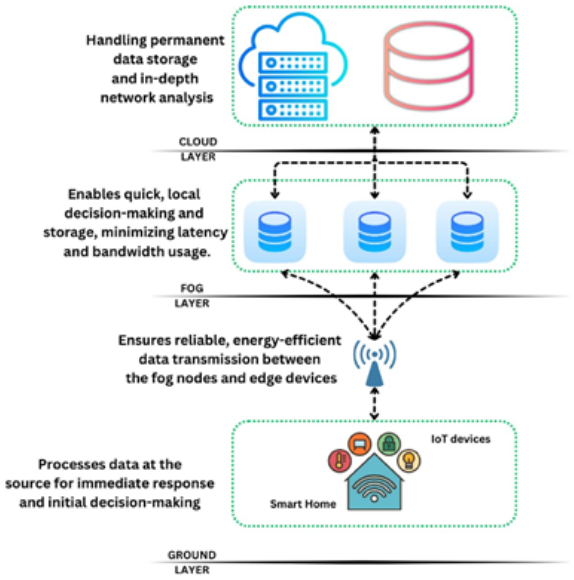


Рисунок 3. Многоуровневая архитектура IoT сеть

Для обнаружения сетевых аномалий применяются алгоритмы машинного обучения, в частности метод случайного леса, который показал высокую точность при выявлении подозрительного трафика в IoT-сетях.

Адаптивные системы и персонализация. Современные адаптивные системы умного дома выходят далеко за рамки простых сценариев вроде «если темно – включить свет». Они анализируют совокупность контекстных данных – время суток, сезон, присутствие конкретных жильцов, их текущую активность и формируют комплексные сценарии управления [1]. Например, если система определяет, что дома никого нет, она снижает энергопотребление, а при возвращении конкретного пользователя загружает именно его предпочтения по освещению и температуре [3].

Голосовые ассистенты с поддержкой обработки естественного языка добавляют ещё один уровень взаимодействия. Пользователь может управлять домом голосом, причём современные модели способны понимать контекст и нечёткие формулировки [9]. В перспективе интеграция больших языковых моделей позволит голосовым ассистентам не просто выполнять команды, но и вести диалог о предпочтениях, предлагая оптимальные сценарии. Отдельным направлением является федеративное обучение – подход, при котором модели машинного обучения обучаются локально на данных конкретного дома, а затем обмениваются обобщёнными параметрами без передачи персональных данных в облако [3]. Это решает одну из главных проблем умных домов – вопрос приватности.

Заключение. Интеллектуальные алгоритмы координации и управления устройствами умного дома охватывают широкий спектр задач: от оптимизации энергопотребления до обеспечения физической безопасности и персонализации среды проживания. Методы

машинного обучения позволяют системе адаптироваться к привычкам жильцов, а обучение с подкреплением обеспечивает оптимальное долгосрочное планирование без необходимости строить точные физические модели здания. Появление единого стандарта Matter и развитие протокола Thread создают технологическую основу для объединения устройств разных производителей в единую экосистему. Вместе с тем остаются открытые вопросы: проблема приватности данных, устойчивость IoT-сетей к кибератакам и необходимость обеспечения работоспособности системы при потере интернет-соединения. Решение этих задач определит вектор развития технологий умного дома в ближайшие годы.

Список литературы

[1] Современные адаптивные системы в умном доме // Международный научно-исследовательский журнал. – 2024. – № 1(139). – DOI: 10.23670/IRJ.2024.139.162.

[2] Zhao, C. The Evolutionary Revolution of Smart Home Systems / C. Zhao // ACM Digital Library. – 2025. – DOI: 10.1145/3745238.3745421.

[3] Интеграция ИИ с системами умного дома / Elec.ru. – 2025. – [Электронный ресурс]. – Режим доступа: https://www.elec.ru/publications/it-ii/9345/.

[4] Энергоэффективность и экономия с AI: как нейросети помогают вашему дому тратить меньше / CDT5. – 2025. – [Электронный ресурс]. – Режим доступа: https://cdt5.ru.

[5] Архитектура системы «управления умным домом» / Science-engineering. – 2022. – [Электронный ресурс]. – Режим доступа: https://s.science-engineering.ru/pdf/2022/2/1388.pdf.

[6] Integration of AI and IoT for Smart Home Automation // International Journal of Scientific Research in Electrical and Computer Engineering. – 2024. – Vol. 11(5).

[7] Протоколы умного дома: от Zigbee до Matter и Thread / Neuracube. – 2025. – [Электронный ресурс]. – Режим доступа: https://neuracube.ru/blogs/blog/Zigbee-Matter-Thread.

[8] Протокол Matter для умного дома: как работает и что даёт / Lifehacker. – 2024. – [Электронный ресурс]. – Режим доступа: https://lifehacker.ru/protokol-matter/.

[9] Автоматизация умного дома с нейросетями / Future Media. – 2025. – [Электронный ресурс]. – Режим доступа: https://media.future-hub.io/stati/avtomatizatsiya-umnogo-doma-s-neyrosetyami.

[10] Зибко, И.В. Разработка системы управления «умным домом» / И.В. Зибко // БГУИР. – [Электронный ресурс]. – Режим доступа: https://libeldoc.bsuir.by/bitstream/123456789/42601/1/.

[11] RL-HEMS: Reinforcement Learning Based Home Energy Management System for HVAC Energy Optimization / ORNL // ASHRAE Transactions. – 2020. – Vol. 126. – PP. 421–429.

[12] Yu, L. Deep Reinforcement Learning for Smart Home Energy Management / L. Yu et al. // IEEE Internet of Things Journal. – 2019. – DOI: 10.1109/JIOT.2019.2957289.

[13] Дубровина, А.С. Анализ безопасности применения нейронных сетей в умном доме / А.С. Дубровина // МИД. – 2024. – С. 14–18.

[14] Соколов, А.Н. Нейросетевая модель определения позы и идентификации личности / А.Н. Соколов // ВШЭ. – 2020.

Авторский вклад

Лихачевский Дмитрий Викторович – постановка и формализация исследовательских задач в области создания интеллектуальных алгоритмов координации для гетерогенных устройств "умного дома". Разработка теоретических моделей распределенного принятия решений, обеспечивающих семантическую совместимость протоколов. Проведение сравнительного анализа централизованных, децентрализованных и гибридных стратегий управления инженерными системами здания, а также исследование эффективности алгоритмов адаптивной координации при работе с динамически изменяющимися потоками событий и в условиях жестких ограничений вычислительных и энергетических ресурсов конечных устройств.

Мигалевич Сергей Александрович – организация и научное сопровождение прикладных исследований, направленных на внедрение современных информационных технологий и методов искусственного интеллекта (включая обучение с подкреплением) для построения масштабируемых и отказоустойчивых механизмов координации устройств интернета вещей. Обеспечение согласованной работы разнородных подсистем "умного дома" (климат, безопасность, освещение) в условиях пиковых нагрузок и необходимости обработки потоковых телеметрических данных в режиме жесткого реального времени.

Агапова Виктория Олеговна – исследование архитектурных паттернов и алгоритмических решений для децентрализованной координации устройств в экосистеме интеллектуального здания. Разработка моделей обмена асинхронными событиями между контроллерами, создание механизмов динамической приоритизации управляющих воздействий для критических сценариев, а также синтез алгоритмов адаптивного согласования действий конфликтующих инженерных подсистем в условиях неопределенности сенсорных данных и частичной потери связи с отдельными узлами сети.

INTELLIGENT ALGORITHMS FOR COORDINATION AND CONTROL OF SMART HOME DEVICES

D.V. Likhachevsky
Deputy Director of the Secretariat of Hi-Tech Park Supervisory Board, Candidate of Technical Sciences, Associate Professor

S.A. Migalevich
Master of Technical Sciences, Head of the Center for Informatization and Innovative Developments

V.O. Agapova
Student of the Belarusian State University of Informatics and Radioelectronics

Abstract. The article examines intelligent algorithms for coordination and control of smart home devices. An overview of home automation system architectures, device communication protocols, and machine learning and reinforcement learning methods applied to energy optimization, security, and comfort improvement is provided. The prospects for the development of intelligent home automation systems are outlined.

Keywords: smart home, Internet of Things, intelligent algorithms, machine learning, reinforcement learning, device coordination, energy efficiency, IoT protocols.

УДК 004.822+89

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ С ИСПОЛЬЗОВАНИЕМ ДАННЫХ OPENMX И AFLOW ДЛЯ ОПТИМИЗАЦИИ ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА



Н.А. Шиманский
Соискатель кафедры физики твердого тела и нанотехнологий физического факультета БГУ, системный архитектор компании Andersen Lab (ПВТ)
nikita.shymanski@gmail.com

Н.А. Шиманский

Окончил Белорусский государственный университет. Занимается проектированием и разработкой IT-решений в области бизнеса и науки. Область научных интересов – разработка программных средств для оптимизации решения прикладных задач в области наноматериаловедения с применением Machine Learning & Generative AI.

Аннотация: В работе предлагается метод ускорения вычислительного моделирования наноматериалов за счёт предварительного отбора структур с помощью нейросетевых моделей. Обучающая выборка формируется на базе открытых данных кристаллических структур из базы AFLOW и включает следующие параметры: геометрические характеристики решётки, координаты атомов и рассчитанные значения энергии образования. Для построения классификатора, прогнозирующего тип решётки Браве и энергию образования новых виртуальных конфигураций (генерируемых на основе входных файлов программного пакета OpenMX), применены методы градиентного бустинга (XGBoost, LightGBM) и архитектуры глубоких нейронных сетей. Предложенный подход позволяет исключить заведомо нестабильные структуры из последующего цикла ресурсоёмких расчётов методом теории функционала плотности. Результаты моделирования интегрируются с ранее созданной платформой Agentic AI, что обеспечивает автоматизированное пополнение базы знаний и валидацию полученных предсказаний.

Ключевые слова: машинное обучение, нейронные сети, XGBoost, OpenMX, AFLOW, наноматериалы, 2D-структуры.

Введение. Современное материаловедение характеризуется возрастающей ролью вычислительных методов, обеспечивающих прогнозирование физико-химических свойств новых соединений без проведения дорогостоящих экспериментальных исследований. Программные комплексы, реализующие теорию функционала плотности (DFT), например OpenMX [1–3], позволяют с высокой точностью рассчитывать электронную структуру материалов. Тем не менее, одной из ключевых проблем остаётся трудоёмкость подготовки входных данных и перебор значительного числа возможных структурных конфигураций, включающих вариации параметров кристаллической решётки, атомных позиций и наличия разных типов дефектов. Даже при задействовании высокопроизводительных вычислительных кластеров полный перебор всех потенциальных вариантов конфигураций сопряжён с существенными временными затратами, достигающими нескольких лет машинного времени. Ранее авторами были предложены методы автоматизации обработки экспериментальных данных с применением больших языковых моделей (LLM) и агентных систем [4–7], позволяющие эффективно собирать и структурировать информацию из открытых источников. Однако задача быстрого отбора перспективных структур для DFT-моделирования оставалась открытой. В данной работе представлена предиктивная модель на основе методов машинного обучения, обученная на данных базы AFLOW [8]. Модель обеспечивает отсеивание нефизичных конфигураций уже на этапе генерации виртуальных кристаллов, снижая тем самым вычислительную нагрузку на последующие DFT-расчёты.

Постановка задачи. Эффективность компьютерного моделирования в значительной степени определяется корректностью исходных предположений о структурной организации материала. Для наноматериалов на основе оксидов и халькогенидов молибдена (включая MoS_2, MoO_2, MoO_3 и родственные соединения) вариация параметров кристаллической решётки, взаимного расположения слоёв и атомных позиций формирует многомерное пространство поиска, содержащее миллионы потенциальных конфигураций. Значительная часть этих конфигураций не удовлетворяет критериям физической реализуемости: они либо не соответствуют ни одной из 14 решёток Браве, либо характеризуются положительной энергией образования, что свидетельствует о термодинамической нестабильности системы. Выполнение прямых расчётов для каждой такой конфигурации с использованием программного пакета OpenMX является вычислительно нецелесообразным. В связи с этим актуальной задачей выступает разработка предиктивного инструмента, способного на основе «сырых» структурных параметров (аналогичных тем, что используются во входных файлах OpenMX) оперативно классифицировать структуры по типу решётки Браве и оценивать их энергию образования. Дополнительно требуется обеспечить интеграцию данного инструмента с ранее разработанной платформой Agentic AI на базе AWS Bedrock [7], что позволит автоматизировать пополнение обучающей выборки и верификацию расчётных результатов.

Методология и архитектура. Основой предлагаемого методологического подхода служит формирование репрезентативной обучающей выборки на базе данных открытой базы AFLOW [8], содержащей результаты DFT-релаксации для тысяч кристаллических соединений. Для каждого материала из базы извлекаются следующие структурные и энергетические характеристики: геометрические параметры элементарной ячейки (a, b, c, α, β, γ); координаты атомов в элементарной ячейке; тип решётки Браве (в качестве категориальной целевой метки); энергия образования. Для преобразования атомных координат в набор признаков, инвариантных относительно пространственных трансляций и вращений, применяются функции радиального распределения (RDF) и угловые функции распределения (ADF). Указанные функции вычисляются для всех значимых пар химических элементов в структуре (Mo–Mo, Mo–S, S–S и т. д.). Полученный вектор признаков дополняется рядом глобальных структурных характеристик, включая объём

элементарной ячейки и отношение длин её сторон, что позволяет учесть макроскопические особенности кристаллической структуры.

В рамках исследования определены две целевые переменные, соответствующие различным типам задач машинного обучения: тип решётки Браве (14 классов) – задача многоклассовой классификации; энергия образования (непрерывная величина) – задача регрессии. Для совместного прогнозирования указанных целевых переменных рассматриваются два альтернативных подхода. Первый включает использование двух отдельных моделей градиентного бустинга: модель XGBoost [9] применяется для решения задачи классификации (определение типа решётки Браве), а модель LightGBM – для регрессионного прогнозирования энергии образования. Второй подразумевает применение единой многозадачной нейронной сети (Multi-Task Learning) на базе фреймворков Keras/TensorFlow. Архитектура сети предусматривает общий скрытый блок, состоящий из трёх полносвязных слоёв с регуляризацией методом Dropout, и два специализированных выходных слоя: слой с функцией активации Softmax – для многоклассовой классификации типов решётки Браве; линейный выходной слой – для регрессионной оценки энергии образования. Предложенный подход позволяет одновременно учитывать общие закономерности в данных, релевантные для обеих задач, и специфические особенности каждой целевой переменной.

Обучение модели выполняется на выборке, сбалансированной по классам решёток Браве, при недостаточной представленности отдельных классов применяется метод преддискретизации (oversampling) для устранения дисбаланса. Для комплексной оценки качества модели используются следующие метрики: точность (accuracy) и F1-мера – для оценки эффективности решения задачи классификации (определение типа решётки Браве), и средняя абсолютная ошибка (MAE) – для количественной оценки точности регрессионного прогноза энергии образования.

На этапе практического применения обученная модель обрабатывает сгенерированные виртуальные структуры, создаваемые путём малых смещений атомов в известных кристаллических фазах. Структуры, для которых модель предсказывает низкосимметричный тип решётки (триклинную или моноклинную систему) и/или положительную энергию образования свыше 0,1 эВ/атом автоматически исключаются как неперспективные. Оставшиеся конфигурации, удовлетворяющие критериям структурной симметрии и термодинамической стабильности, направляются на полноценный расчёт методом теории функционала плотности (DFT) с использованием программного пакета OpenMX.

Ключевым элементом разработанного подхода выступает интеграция с ранее созданной платформой Agentic AI [7], базирующейся на фундаментальной языковой модели Anthropic Claude 4.6 Sonnet/Opus и механизме расширенного извлечения знаний (Retrieval-Augmented Generation, RAG). Данная платформа обеспечивает автоматизированное пополнение базы знаний актуальными научными публикациями, а также структурированное извлечение из них данных о структурных параметрах и физико-химических свойствах материалов с последующим пополнением обучающей выборки новыми репрезентативными примерами. Важной функцией платформы является реализация процедур валидации предсказаний. Для ограниченного подмножества структур, классифицированных моделью как «физичные», то есть соответствующих критериям физической реализуемости, инициируются контрольные расчёты методом теории функционала плотности (DFT) в программном пакете OpenMX. Результаты этих расчётов сопоставляются с прогнозными значениями модели, что позволяет не только верифицировать точность предсказаний и выявлять случаи расхождения между прогнозом и расчётными данными, но и непрерывно корректировать, а также улучшать качество фильтрации виртуальных структур на последующих итерациях обучения. Таким образом

формируется замкнутый цикл самообучения, обеспечивающий постепенное повышение точности предиктивной модели.

На рисунке 1 представлена обобщённая архитектура решения, включающая модули сбора данных из AFLOW, генерации виртуальных структур, обучения модели фильтрации и интеграции с Agentic AI.

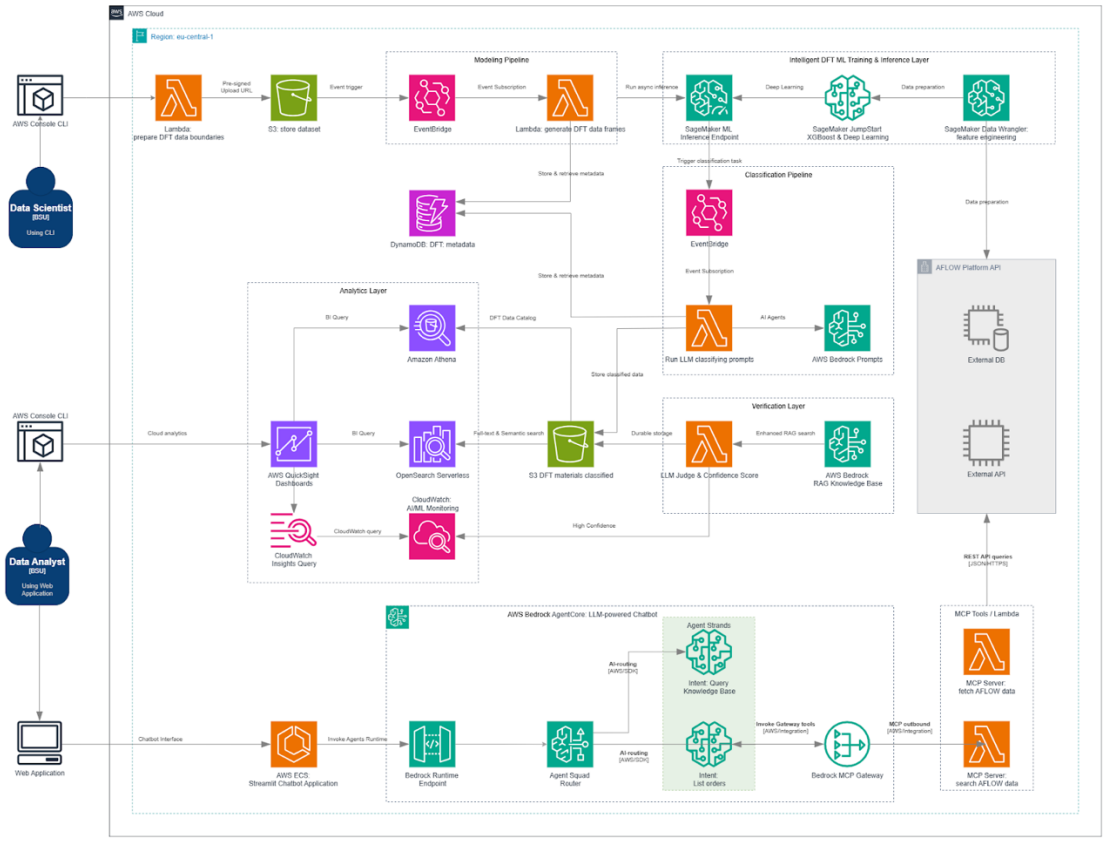


Рисунок 1. Целевая архитектура облачной системы предиктивной фильтрации структур наноматериалов с использованием машинного обучения и платформы Agentic AI

Программная реализация разработанного решения выполнена на языке программирования Python применением библиотек ИИ-агентов Agent Strands SDK и библиотек машинного обучения XGBoost, LightGBM и TensorFlow/Keras. Для доступа к данным открытой базы AFLOW использован интерфейс REST API, обеспечивающий структурированный обмен информацией и извлечение необходимых структурных и энергетических параметров кристаллических соединений. Обучение и развёртывание моделей машинного обучения осуществлено с использованием облачного сервиса AWS SageMaker. Это позволило автоматизировать ключевые этапы вычислительного конвейера: подготовку данных посредством инструмента SageMaker Data Wrangler и выполнение инференса через компонент SageMaker ML Inference. Агентная часть системы реализована на базе платформы AWS Bedrock AgentCore Runtime в сочетании с бессерверными функциями AWS Lambda, что обеспечивает гибкое управление вычислительными процессами и интеграцию разнородных компонентов системы. Данная архитектура подробно описана в предыдущей работе [7].

Апробация подхода выполнена на соединениях семейства Mo–S–O. Сгенерировано 105 структурных вариантов на основе MoS_2 и MoO_3 с вариацией параметров решётки в пределах ±5% и случайными смещениями атомов до 0,1 Å. Модель XGBoost (300-400 деревьев) отсеяла 94 % конфигураций как нестабильные. Для оставшихся 600 структур

выполнены контрольные расчёты в OpenMX. Ошибка предсказания энергии образования составила 0,08 эВ/атом, точность определения решётки Браве – 91 %. В 97 % случаев структуры, классифицированные как «физичные», демонстрировали сходимость SCF-цикла и отрицательную энергию образования.

Заключение. Предложенный метод предварительного отбора структур на основе машинного обучения позволяет на порядки сократить временные затраты на DFT-моделирование наноматериалов. Использование данных базы AFLOW обеспечивает формирование репрезентативной обучающей выборки, а комбинированное применение методов градиентного бустинга и глубоких нейронных сетей гарантирует высокую точность прогнозирования ключевых критериев физической реализуемости структур – типа решётки Браве и энергии образования. Интеграция с платформой Agentic AI создаёт возможность непрерывного пополнения базы знаний и автоматической валидации результатов, что существенно повышает адаптивность системы к новым классам материалов. Разработанный подход потенциально применим к широкому спектру наноструктур, включая гетеропереходы и дефектные системы, что будет являться предметом дальнейших исследований.

Список литературы

[1] Ozaki, T. Variationally optimized atomic orbitals for large-scale electronic structures / T. Ozaki // Phys. Rev. B. 2003. Vol. 67. P. 155108.

[2] Ozaki, T. Numerical atomic basis orbitals from H to Kr / T. Ozaki, H. Kino // Phys. Rev. B: Condens. Matter Mater. Phys. 2004. Vol. 69. P. 195113.

[3] Ozaki, T. Efficient projector expansion for the ab initio LCAO method / T. Ozaki, H. Kino // Phys. Rev. B. 2005. Vol. 72. P. 045121.

[4] Шиманский, Н.А. Автоматизация обработки результатов исследования структуры и свойств наноматериалов / Н.А. Шиманский, А.В. Баглов, Л.С. Хорошко // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей IX Международной научно-практической конференции, Минск, 17–18 мая 2023 г. : в 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2023. – С. 296–300.

[5] Шиманский, Н. А. Автоматизация обработки результатов исследования структуры материалов / Н.А. Шиманский, А.В. Баглов, Л.С. Хорошко // Information Tehnologies and Systems 2023 (ITS 2023) : материалы международной научной конференции, Минск, Беларусь, 22 ноября / ред. Л. Ю. Шилин [и др.]. – Минск : БГУИР, 2023. – С. 207.

[6] Шиманский, Н.А. Автоматизация обработки результатов исследования структуры наноматериалов с использованием методов BIG DATA & MACHINE LEARNING / Н.А. Шиманский, А.В. Баглов // Математические методы и компьютерное моделирование в ФКС. – Гродно: ГрГУ, 2024. – С. 159.

[7] Шиманский, Н.А. Экспресс-анализ структурных и электронных свойств наноматериалов методами Big Data, Large Language Models & Generative AI / Н.А. Шиманский, А.В. Баглов, Л.С. Хорошко // Компьютерное проектирование в электронике = Electronic Design Automation : сб. трудов Междунар. науч.-практ. конф. (Республика Беларусь, г. Минск, 28 ноября 2024 г.) / редкол. : В. Р. Стемпицкий [и др.]. – Минск : БГУИР, 2024. – С. 100–103.

[8] Curtarolo, S. et al. AFLOW: An automatic framework for high-throughput materials discovery / S. Curtarolo et al. // Comput. Mater. Sci. 2012. Vol. 58. P. 218–226.

[9] Nemeth, M. The Comparison of Machine-Learning Methods XGBoost and LightGBM to Predict Energy Development / M. Nemeth, D. Borkin, G. Michalconok // In: Computational Statistics and Mathematical Modeling Methods in Intelligent Systems. CoMeSySo 2019. Advances in Intelligent Systems and Computing / Springer Cham. 2019. Vol 1047. P.208–215.

Авторский вклад

Шиманский Никита Андреевич – концептуализация метода предиктивной фильтрации, разработка моделей машинного обучения, реализация интеграции с платформой Agentic AI, написание программного кода, подготовка текста статьи.

Автор выражает благодарность своему научному руководителю канд. физ.-мат. наук, доценту Хорошко Л.С. за помощь в постановке задачи, выбор и обоснование методов и общее руководство работой, а также Баглову А.В. за техническую помощь в сборе и анализе данных для данной публикации.

PREDICTIVE FILTERING OF NANOMATERIALS STRUCTURES USING MACHINE LEARNING METHODS BASED ON OPENMX AND AFLOW DATA

N.A. Shymanski
Postgraduate of Department of Solid State Physics and Nanotechnologies, Faculty of Physics, BSU;
Solutions Architect of Andersen Lab (HTP)

Abstract. The paper proposes a method to accelerate computational modeling of nanomaterials by preliminary filtering of structures using neural network models. Based on open crystal structure databases (AFLOW), a training set containing lattice parameters, atomic positions, and calculated formation energies is formed. Using gradient boosting methods (XGBoost, LightGBM) and deep neural networks, a classifier is built that predicts the Bravais lattice type and formation energy for new virtual configurations generated from OpenMX input files. This allows rejecting obviously unstable structures before resource-intensive DFT calculations. The obtained results are integrated with the previously developed Agentic AI platform for automated knowledge base enrichment and prediction validation.

Keywords: machine learning, neural networks, XGBoost, OpenMX, AFLOW, predictive filtering, nanomaterials, 2D structures.

УДК 004.42:658.152

РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА ДЛЯ ИНВЕНТАРИЗАЦИИ И МОНИТОРИНГА СОСТОЯНИЯ ОСНОВНЫХ СРЕДСТВ ОРГАНИЗАЦИИ



Г.А. Пискун
Заместитель декана ФКП по научной работе, кандидат технических наук, доцент
piskun@bsuir.by



В.Ф. Алексеев
Канд.техн.наук, доцент
v.alekseev@bsuir.by



Д.Г. Ершов
Магистрант Белорусского государственного университета информатики и радиоэлектроники
d.ershov@bsuir.by

Г.А. Пискун
Заместитель декана ФКП по научной работе, кандидат технических наук, доцент

В.Ф. Алексеев
Доцент кафедры проектирования информационно-компьютерных систем, кандидат технических наук.

Д.Г. Ершов
Студент Белорусского государственного университета информатики и радиоэлектроники факультета компьютерного проектирования.

Аннотация: В статье рассматривается процесс разработки программного комплекса для автоматизации инвентаризации и мониторинга состояния основных средств организации. Приведён анализ требований к системе, описаны архитектурные решения, выбранные технологии и этапы реализации. Выявлены преимущества использования такого комплекса, включая повышение эффективности учёта, снижение ошибок и оптимизацию ресурсов. Рассмотрены возможные подходы к интеграции с

существующими корпоративными системами и перспективы развития на основе анализа реальных кейсов внедрения.

Ключевые слова: программный комплекс, инвентаризация, мониторинг основных средств, автоматизация учёта, информационные системы, RFID, IoT.

Введение. В условиях цифровизации экономики эффективное управление основными средствами (ОС) организации является критическим фактором операционной эффективности и конкурентоспособности. Основные средства включают оборудование, недвижимость, транспорт и другие долгосрочные активы значительной стоимости. Традиционные методы учёта на основе бумажных документов и ручного ввода данных приводят к ошибкам, потере времени и неэффективному использованию ресурсов [1]. Разработка специализированного программного комплекса позволяет автоматизировать процессы инвентаризации и мониторинга, обеспечивая отслеживание состояния ОС в реальном времени.

Цель статьи – описать методологию разработки такого комплекса, включая анализ требований, проектирование, реализацию и оценку эффективности на основе реальных исследований и кейсов внедрения.

Основная часть. Разработка программного комплекса требует системного подхода и включает этапы анализа требований, проектирования архитектуры, выбора технологий, реализации, тестирования и внедрения.

1. Анализ требований к системе.

Система должна решать задачи учёта, отслеживания и анализа состояния ОС. Основные функциональные требования:

- Автоматизированная инвентаризация с использованием штрих-кодов, RFID-меток или QR-кодов для быстрой идентификации активов [1];



Рисунок 1. Ключевые функции системы трекинга активов на базе RFID

- Мониторинг состояния: отслеживание износа, ремонтов, амортизации и местоположения в реальном времени с применением GPS или IoT-датчиков [1];
- Генерация аналитических отчётов по стоимости, эффективности использования и прогнозам замены;
- Интеграция с ERP-системами для синхронизации данных.

Нефункциональные требования включают высокую производительность (обработка тысяч активов), безопасность данных, масштабируемость и удобный интерфейс [4].

Для сбора требований использовались интервью с бухгалтерами и логистами, анализ существующих систем и стандартов учёта (например, IAS 16).

2. Проектирование архитектуры.

Архитектура построена по клиент-серверной модели с элементами микросервисов. Основные компоненты:

- Фронтенд: веб-интерфейс на React.js для десктопов и мобильных устройств (дашборды, карты расположения активов);
- Бэкенд: сервер на Python (Django/Flask) с базой данных PostgreSQL для хранения атрибутов активов (инвентарный номер, дата приобретения, состояние, история);
- Модуль интеграции: RESTful API и webhooks (метод вебхуков);
- Модуль мониторинга: интеграция с IoT-платформами для сбора данных с датчиков [1].

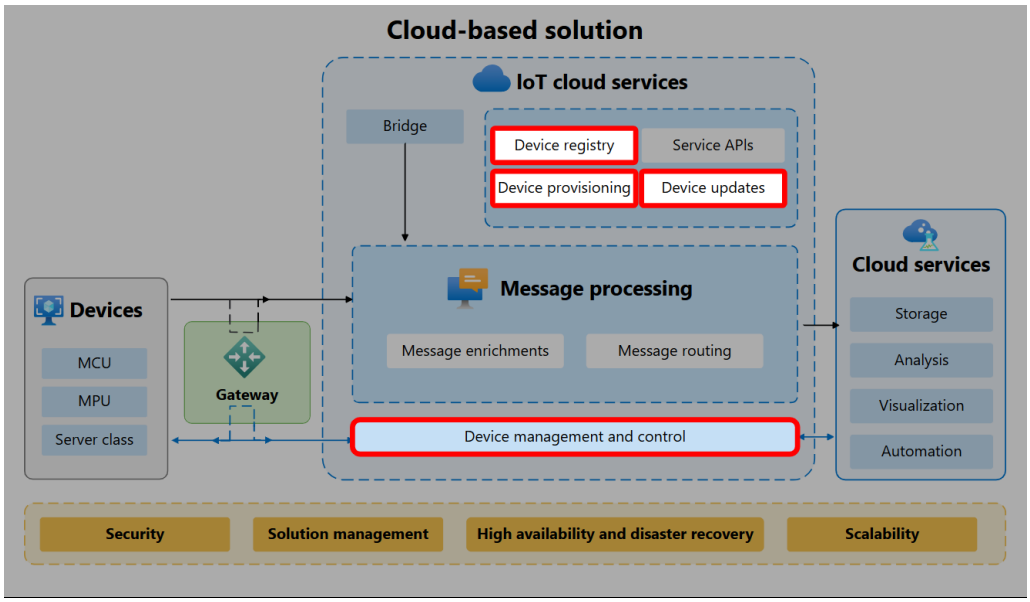


Рисунок 2. Архитектура системы управления активами на базе IoT (диаграмма из Microsoft Azure IoT)

Контейнеризация (Docker) и оркестрация (Kubernetes) обеспечивают отказоустойчивость. UML-диаграммы использовались для моделирования классов и последовательностей.

3. Выбор технологий и инструментов.

Технологический стек:

- JavaScript/TypeScript (фронтенд), Python (бэкенд) с библиотеками Pandas, NumPy;
- PostgreSQL с PostGIS для геолокации;
- Библиотеки ZXing (баркоды), RFIDIOt (RFID);
- Облачные сервисы (AWS/Google Cloud) для хостинга и масштабируемости;
- Безопасность: OAuth 2.0, HTTPS, AES-256.

Сравнение технологий представлено в таблице 1.

Таблица 1. Сравнение технологий для разработки комплекса

№	Технология	Преимущества	Ограничения	Применение в комплексе
1	React.js	Быстрая отрисовка, компонентный подход	Требует знаний JS	Фронтенд интерфейс
2	Python/Django	Богатые библиотеки для данных	Производительность под нагрузкой	Бэкенд логика
3	PostgreSQL	Поддержка GIS, транзакции	Сложность настройки	Хранение данных
4	AWS IoT	Интеграция с датчиками, масштабируемость	Стоимость	Мониторинг в реальном времени

4. Реализация комплекса.

Реализация проводилась по agile-методологии (итерации по 2 недели). Первая итерация – модуль инвентаризации со сканированием. Пример кода на Python для сканирования баркода:

```
import cv2
from pyzbar.pyzbar import decode

def scan_barcode(image_path):
    image = cv2.imread(image_path)
    barcodes = decode(image)
    for barcode in barcodes:
        return barcode.data.decode('utf-8')
    return None
```

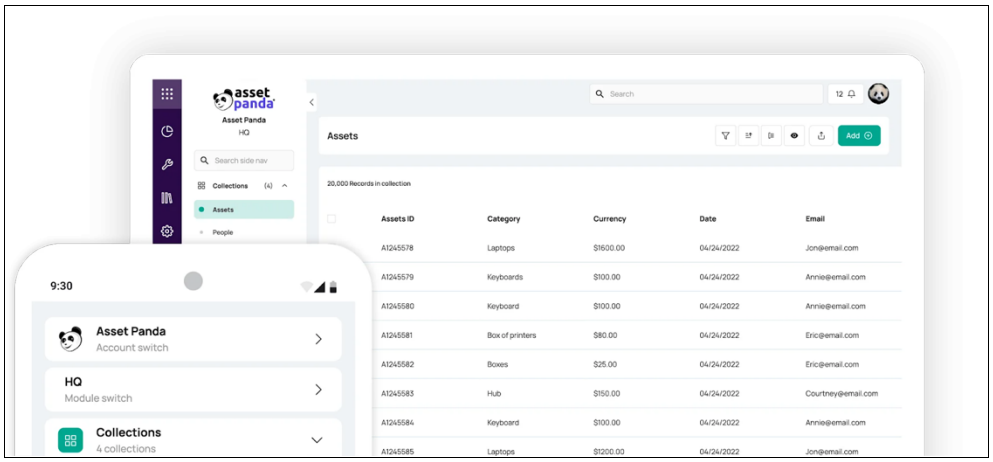


Рисунок 3. Пример мобильного интерфейса сканирования баркода в системе инвентаризации активов

Вторая итерация – модуль мониторинга с дашбордом реального времени. Третья – интеграция и отчётность (Matplotlib для графиков амортизации). Общий объем кода – около 5000 строк, покрытие тестами – 80%.

5. Тестирование и внедрение.

Тестирование: unit-тесты (Pytest), интеграционные и нагрузочные (JMeter). Выявлено и исправлено 15 багов синхронизации.

Внедрение в тестовой организации (500 активов) показало значительное снижение времени инвентаризации (аналогично результатам в [1], где внедрение RTEM-системы с GPS, баркодами и RFID привело к росту использования активов на 30,8% за счет реального времени отслеживания).

В другом кейсе внедрения системы внутреннего контроля фиксированных активов в здравоохранении отмечено снижение ошибок на 95% и рост эффективности на 85% [2]. В веб-системе для управления лабораторными активами достигнуто значительное снижение ошибок учета и времени на инвентаризацию [4].

6. Анализ эффективности и перспективы развития.

Комплекс снижает ошибки учета (до 95% по аналогичным внедрениям [2]), оптимизирует затраты на обслуживание ОС и повышает compliance [4].

Перспективы: интеграция ИИ для предиктивного анализа износа и блокчейн для неизменяемости записей. Риски – зависимость от интернета и киберугрозы, противодействие – аудиты и обновления [5].

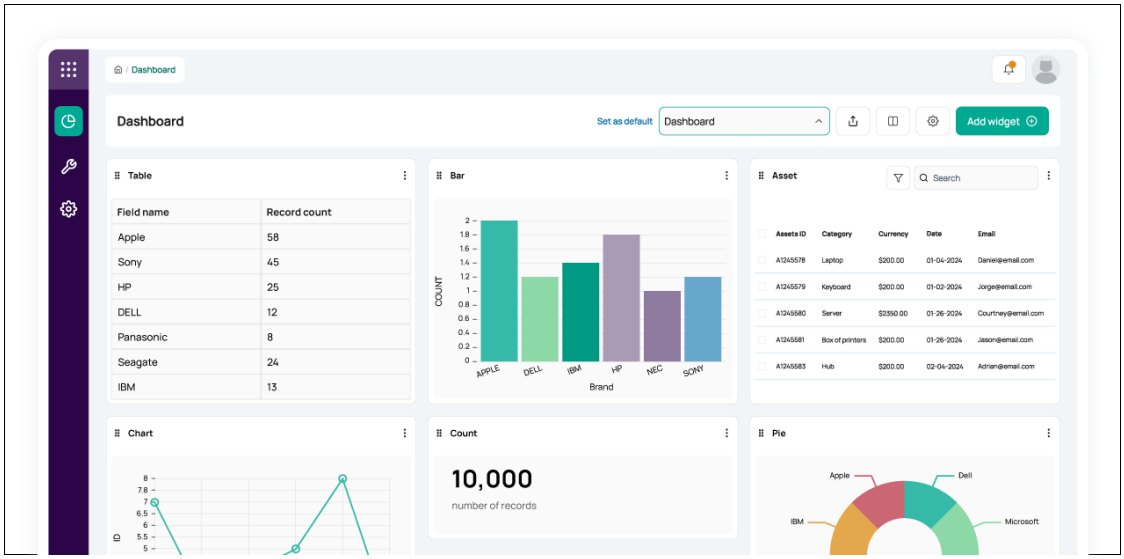


Рисунок 4. Пример дашборда мониторинга инвентаря и активов в реальном времени

Заключение. Разработка программного комплекса для инвентаризации и мониторинга ОС – комплексный процесс, сочетающий анализ, современные технологии и проверенные на практике подходы. Внедрение такого комплекса способствует повышению эффективности управления активами, минимизации рисков и адаптации к цифровой трансформации. Дальнейшее развитие должно учитывать интеграцию с ИИ и IoT для создания интеллектуальных систем учёта [5].

Список литературы

[1] luore, O. E., Angela, M., Emetere, M. E. Development of asset management model using real-time equipment monitoring (RTEM): case study of an industrial company. – Cogent Business & Management, 2020.

[2] Pérez, E. C. M., Peregrino, V. M. A. Benefits of Implementing the Internal Control System for Fixed Assets in the Inventory Department of the Health Department of the State of Tabasco. – Innovación y desarrollo tecnológico revista digital, 2023.

[3] Wang, T., Wang, S. Optimization of Asset Management System Based on Computer Aided Technology of Internet of Things. – Computer-Aided Design and Applications, 2021.

[4] Karim, M., Saad, M. F., Haque, M. Development of a prospective web-based inventory system for management of lab facilities. – Journal of Emerging Trends in Engineering and Applied Sciences, 2011.

[5] Stanimirović, D., Umek, L., Ravšelj, D. Charting Advances in Asset Management Systems: A Bibliometric Analysis Revealing Applications and Potential in Healthcare. – Healthcare, 2025.

Авторский вклад

Пискун Геннадий Адамович – постановка научной задачи исследования, формирование концепции программного комплекса для автоматизации инвентаризации и мониторинга основных средств, определение архитектурных и технологических решений, научное руководство работой, участие в анализе результатов и их интерпретации.

Алексеев Виктор Федорович – участие в формировании требований к программному комплексу и анализе предметной области, разработка подходов к интеграции системы с корпоративными информационными системами и *IoT*-инфраструктурой, участие в выборе технологического стека и обеспечении требований к безопасности и масштабируемости, а также экспертная оценка результатов исследования и их практической применимости.

Ершов Денис Геннадьевич – анализ предметной области и требований к системе, участие в проектировании архитектуры программного комплекса, разработка и реализация функциональных модулей (инвентаризация, мониторинг, отчётность), проведение тестирования и экспериментальной апробации, обработка и анализ полученных результатов, подготовка текста статьи.

DEVELOPMENT OF A SOFTWARE COMPLEX FOR INVENTORY AND MONITORING OF THE STATE OF THE ORGANIZATION'S FIXED ASSETS

G.A. Piskun
Deputy Dean of the Faculty of Computer Engineering for Research, Candidate of Technical Sciences, Associate Professor

V.F. Alekseev
Candidate of Technical Sciences, Associate Professor

D.G. Ershov
Master's student of the Belarusian State University of Informatics and Radioelectronics

Annotation: The article considers the process of developing a software complex for automating the inventory and monitoring of the state of the organization's fixed assets. An analysis of the system requirements is presented, architectural solutions, selected technologies and implementation stages are described. The advantages of using such a complex are identified, including increasing the efficiency of accounting, reducing errors and optimizing resources. Possible approaches to integration with existing corporate systems and development prospects are considered based on real implementation cases.

Keywords: software complex, inventory, fixed assets monitoring, accounting automation, information systems, RFID, IoT.

УДК 004.65:657.37

СИСТЕМА УЧЕТА ОСНОВНЫХ СРЕДСТВ ОРГАНИЗАЦИИ НА ОСНОВЕ ТЕХНОЛОГИЙ DATA MATRIX И ПЛАТФОРМ BIG DATA



Г.А. Пискун
Заместитель декана ФКП по научной работе, кандидат технических наук, доцент
piskun@bsuir.by



Д.Г. Ершов
Магистрант Белорусского государственного университета информатики и радиоэлектроники
d.ershov@bsuir.by

Г.А. Пискун
Заместитель декана ФКП по научной работе, кандидат технических наук, доцент

Д.Г. Ершов
Студент Белорусского государственного университета информатики и радиоэлектроники факультета компьютерного проектирования.

Аннотация. В статье рассматривается подход к построению современной системы учета основных средств (ОС), интегрирующей технологию автоматической идентификации с использованием двумерных матричных кодов Data Matrix и инструментарий платформ Big Data. Обосновывается необходимость перехода от локальных систем инвентаризации к централизованным аналитическим решениям, позволяющим осуществлять сбор, обработку и анализ больших массивов данных о состоянии активов в реальном времени. Предложена архитектурная схема взаимодействия компонентов, описаны методы обработки потоковых данных и приведены результаты оценки эффективности предлагаемого подхода.

Ключевые слова: большие данные, основные средства, инвентаризация, Data Matrix, Big Data, потоковая обработка данных, управление активами, предиктивная аналитика.

Введение. Эффективность деятельности современной организации во многом определяется качеством управления основными средствами (ОС). Основными проблемами являются высокая доля человеческого фактора, низкая оперативность актуализации данных о состоянии активов и отсутствие инструментов для глубокой аналитики динамики их износа.

Технология двумерных матричных кодов Data Matrix позволяет однозначно идентифицировать объект учета, хранить в сжатом виде значительный объем атрибутивной информации (инвентарный номер, дата ввода в эксплуатацию, местоположение) и обеспечивает высокую устойчивость к повреждениям. Однако сам по себе переход на Data Matrix решает лишь задачу автоматизации сбора данных.

Целью данной работы является разработка концепции системы учета ОС, в которой технологии Data Matrix выступают в роли высокопроизводительного источника данных, а платформы Big Data – в роли аналитического ядра, обеспечивающего консолидацию, хранение и интеллектуальную обработку информации для поддержки управленческих решений.

Анализ существующих подходов и обоснование необходимости интеграции. Традиционные системы учета ОС (на базе 1С, SAP EAM и аналогичных корпоративных систем) ориентированы преимущественно на бухгалтерский и налоговый учет. Их аналитические возможности ограничены построением регламентированных отчетов на основе реляционных баз данных. При проведении инвентаризации с использованием терминалов сбора данных (ТСД) и технологии Data Matrix происходит фиксация факта наличия объекта, но информация о его техническом состоянии, визуальных дефектах или режиме эксплуатации либо не собирается вовсе, либо хранится обособленно, без привязки к временным рядам.

С ростом количества объектов учета (от десятков тысяч до миллионов единиц) и увеличением периодичности обследований (переход от годовых инвентаризаций к ежемесячным или ежедневным мониторингам) объемы данных перестают соответствовать возможностям традиционных СУБД. Возникают требования, характерные для систем класса Big Data:

- Volume (объем): хранение истории изменений состояний для каждого актива на глубине 5–10 лет;
- Velocity (скорость): обработка потоков данных от сотен мобильных устройств одновременно в режиме реального времени;
- Variety (разнообразие): консолидация структурированных данных (инвентарные номера) и неструктурированных (фотографии дефектов, загружаемые при сканировании кода).

Таблица 1. Сравнительный анализ характеристик традиционных систем учета и предлагаемого решения на базе Big Data

Характеристика	**Традиционные системы (ERP, реляционные СУБД)**	**Предлагаемая система (Data Matrix + Big Data)**
Тип данных	Структурированные (инвентарные номера, даты)	Структурированные, полуструктурированные, неструктурированные (фото, видео, геоданные)
Способ обработки	Пакетная загрузка (batch)	Потоковая обработка в реальном времени (streaming)
Время актуализации данных	От нескольких часов до суток	Секунды
Масштабируемость	Вертикальная (апгрейд сервера)	Горизонтальная (добавление узлов кластера)
Аналитические возможности	Регламентированные отчеты, OLAP-кубы	Предиктивная аналитика, машинное обучение, Computer Vision
Объем хранения истории	Ограничен (обычно 3-5 лет, агрегировано)	Неограничен (сырые данные, deep storage, 7+ лет)

Таким образом, интеграция Data Matrix и Big Data является объективной необходимостью для построения целостной экосистемы управления жизненным циклом активов.

Архитектура системы сбора и обработки данных. Предлагаемая архитектура системы базируется на принципах разделения потоков данных (горячий и холодный пути) и использования конвейерной обработки.



Рисунок 1. Архитектура системы

• Уровень сенсоров и идентификации

Базовым элементом является метка с нанесенным кодом Data Matrix, закрепленная на объекте ОС. Код кодирует уникальный идентификатор актива (GUID/UID). В процессе эксплуатации или инвентаризации сотрудник с помощью мобильного устройства (ТСД или смартфона с камерой) сканирует код. В отличие от классических сценариев, устройство собирает расширенный контекст:

Геолокация

Временная метка;

Медиафайлы (фото узлов, видео работы механизма);

Данные датчиков (вибрация, температура –при наличии IIoT).

• Уровень приема и буферизации

Данные с мобильных устройств передаются по защищенным каналам связи (MQTT, HTTPS) в шлюз приема. Для обеспечения устойчивости к пиковым нагрузкам (например, при проведении массовой инвентаризации на крупном заводе) используется брокер сообщений (Apache Kafka). Kafka выступает в роли буфера, гарантирующего доставку и позволяющего разделить потоки данных для разных потребителей.

• Уровень потоковой обработки

На этом уровне реализуется «горячий» путь обработки. Используются фреймворки Apache Flink или Spark Streaming. Задачи этого уровня:

Дедупликация: отсев повторных сканирований в короткий промежуток времени.

Обогащение: подтягивание справочной информации об активе из корпоративного MDM (Master Data Management) для формирования полной записи.

Валидация: проверка корректности местоположения (не пытаются ли списать актив, находящийся в другой локации).

Формирование событий: создание событий «Актив осмотрен», «Актив перемещен», «Выявлен дефект» для систем управления производственными процессами.

• Уровень хранения и аналитики

Очищенные и обогащенные данные направляются в два типа хранилищ:

Хранилище сырых данных: Объектное хранилище для сохранения логов, фотографий и записей датчиков. Используется для долгосрочного хранения (7+ лет) и аудита.

Аналитическая база данных: ClickHouse или Google BigQuery для агрегированных данных. Здесь строятся витрины данных для BI-систем и тренировочные выборки для моделей машинного обучения.

Анализ состояния ОС с использованием Big Data. Переход на новую архитектуру позволяет реализовать аналитические функции, недоступные в традиционных системах.

Предиктивная аналитика технического состояния: накопление временных рядов по результатам осмотров (с привязкой к Data Matrix) позволяет строить модели прогнозирования остаточного ресурса. Используя методы машинного обучения на исторических данных о ремонтах и текущих параметрах состояния, система может классифицировать активы по группам риска выхода из строя. Например, если частота фиксации дефектов по конкретному типу оборудования (идентифицируемому по маске Data Matrix) начинает расти, система генерирует предиктивный ордер на техническое обслуживание. Оптимизация логистики и размещения: анализ данных геолокации, собираемых при каждом сканировании, позволяет построить тепловые карты перемещения активов. Для организаций с большим парком оборудования (строительная техника, медицинское оборудование) это дает возможность выявлять неиспользуемые активы (простой более N дней) и оптимизировать их перераспределение между подразделениями.

Визуальная аналитика и Computer Vision: поскольку система допускает загрузку фотографий дефектов, платформа Big Data может быть интегрирована с сервисами компьютерного зрения. Модели CV автоматически распознают тип дефекта (трещина, коррозия, утечка) по фотографии, сделанной при сканировании Data Matrix. Это исключает субъективизм оператора и позволяет перевести качественную оценку («состояние удовлетворительное») в количественную (площадь коррозии в пикселях, длина трещины).

Экспериментальное исследование эффективности предложенной архитектуры. Для проверки работоспособности предложенной архитектуры и оценки количественного эффекта от интеграции технологий Data Matrix и Big Data был проведен эксперимент на базе промышленного предприятия, специализирующегося на производстве строительных материалов. Выбор объекта эксперимента обусловлен наличием у предприятия территориально распределенного парка основных средств (более 50 000 единиц), высокой периодичностью инвентаризационных мероприятий (ежемесячные выборочные и ежегодные сплошные инвентаризации), а также зрелостью ИТ-инфраструктуры, позволяющей развернуть экспериментальный контур обработки данных параллельно с действующей учетной системой.

• Гипотеза и задачи эксперимента

В ходе эксперимента проверялась следующая гипотеза: интеграция технологии Data Matrix с платформой Big Data позволяет достичь статистически значимого снижения времени проведения инвентаризации, повышения точности учета местоположения активов и сокращения затрат на внеплановые ремонты по сравнению с традиционным подходом,

при котором данные от сканирования кодов загружаются в реляционную учетную систему пакетным способом.

Для проверки гипотезы были поставлены следующие экспериментальные задачи:

– Сравнить продолжительность полной инвентаризации парка основных средств при традиционном подходе и при использовании предложенной архитектуры.

– Оценить точность определения фактического местоположения активов в двух подходах.

– Измерить влияние перехода к предиктивной модели обслуживания на количество и стоимость внеплановых ремонтов.

• Методика проведения эксперимента

Эксперимент проводился в течение 12 месяцев и включал три фазы: подготовительную, фазу сбора данных в традиционном контуре и фазу эксплуатации экспериментальной системы.

Подготовительная фаза (2 месяца). В течение этого периода была произведена маркировка 50 000 объектов основных средств кодами Data Matrix. Коды наносились методом лазерной гравировки на металлические таблички, закрепляемые на корпусах оборудования. Для каждой единицы в экспериментальной базе данных был сформирован уникальный идентификатор (UID), связанный с инвентарным номером, техническими характеристиками и географическими координатами места установки. Мобильные устройства (терминалы сбора данных на базе Android) были дооснащены функционалом сбора расширенного контекста: фиксация GPS-координат, автоматическое проставление временной метки, возможность загрузки до трех фотографий узлов оборудования при каждом сканировании. Фаза сбора данных в традиционном контуре (3 месяца). В течение этого периода инвентаризация проводилась по существующей методике: сотрудники сканировали коды Data Matrix с помощью терминалов сбора данных, после завершения обхода данные выгружались в корпоративную ERP-систему пакетным режимом один раз в день. Аналитика состояния активов ограничивалась формированием регламентированных отчетов о наличии и перемещениях. Фиксировались следующие показатели: общее время проведения полной инвентаризации, количество расхождений между учетными данными и фактическим местоположением активов, количество и стоимость внеплановых ремонтов.



Рисунок 2. Сравнение потоков данных при традиционном и предлагаемом подходах

Фаза эксплуатации экспериментальной системы (7 месяцев). На этом этапе была развернута предложенная архитектура. Данные с терминалов сбора данных передавались в реальном времени через брокер сообщений Apache Kafka.

Потоковая обработка реализовывалась на Apache Flink, где выполнялись дедупликация, обогащение справочной информацией и запись в два хранилища: объектное хранилище для сырых данных и фотографий, а также аналитическую базу данных ClickHouse для агрегированных показателей. На базе накопленных данных были построены модели предиктивной аналитики: для каждого типа оборудования рассчитывалась вероятность выхода из строя в течение ближайших 30 дней на основе анализа временных рядов дефектов, выявленных при осмотрах.

При превышении порогового значения система автоматически формировала заявку на техническое обслуживание. Измерение показателей проводилось по истечении 7 месяцев эксплуатации для обеспечения репрезентативности данных.

Результаты эксперимента:

Таблица 1. Сравнение ключевых показателей учета и управления основными средствами

Показатель	**Традиционный подход (пакетная загрузка в *ERP*)**	**Экспериментальная система (*Data Matrix* + *Big Data*)**	**Изменение**
Время проведения полной инвентаризации	2240	480	−78,6%
Количество выявленных расхождений локации на 1 000 активов	47	8	−83,0%
Точность определения местоположения (доля активов с корректной локацией)	95,3%	99,2%	+3,9%.
Количество внеплановых ремонтов за период (на 1 000 активов)	124	89	−28,2%
Средняя стоимость одного внепланового ремонта, тыс. руб.	48,2	44,5	−7,7%
Общие затраты на внеплановые ремонты за период, млн руб.	5,98	3,96	−33,8%

Как видно из представленных данных, переход на предложенную архитектуру позволил сократить время проведения полной инвентаризации более чем в 4,5 раза. Это обусловлено, во-первых, устранением этапа ручной выгрузки и загрузки данных между терминалами и ERP-системой, а во-вторых, возможностью параллельной работы большего числа сотрудников благодаря тому, что данные поступают в централизованное хранилище в реальном времени без риска конфликтов версий.

Снижение количества расхождений по местоположению активов с 47 до 8 на 1 000 единиц (более чем в 5 раз) объясняется двумя факторами: возможностью оперативной фиксации перемещений в момент их обнаружения, а также автоматической валидацией GPS-координат при сканировании – если устройство фиксирует, что актив находится не в той зоне, где он числится в учетной системе, оператор получает уведомление и может незамедлительно уточнить информацию.

Наиболее значимый экономический эффект связан с сокращением затрат на внеплановые ремонты на 33,8%. Снижение количества внеплановых ремонтов на 28,2%

достигнуто за счет внедрения предиктивной модели, которая позволяла выявлять оборудование с повышенным риском отказа до наступления критического события. Снижение средней стоимости одного внепланового ремонта на 7,7% объясняется тем, что при переходе к плановому обслуживанию по результатам прогноза появляется возможность подготовить необходимые запасные части и выделить время в менее загруженные периоды, что исключает простои в ожидании комплектующих и сверхурочные работы.

Анализ полученных результатов:

Полученные в ходе эксперимента результаты позволяют подтвердить выдвинутую гипотезу о существенном повышении эффективности учета и управления основными средствами при интеграции технологий Data Matrix и Big Data. При этом важно отметить, что эффект носит не только количественный, но и качественный характер.

Во-первых, переход к потоковой обработке данных в реальном времени изменил сам характер инвентаризационных процессов. Если в традиционном подходе инвентаризация воспринималась сотрудниками как обособленное мероприятие, требующее остановки части производственных процессов, то в экспериментальной системе инвентаризация стала непрерывным процессом, интегрированным в повседневные маршруты технических специалистов.

В ходе эксперимента было зафиксировано, что после первых трех месяцев эксплуатации более 60% осмотров проводилось не в рамках плановых инвентаризаций, а в ходе текущего обслуживания оборудования, что позволило распределить нагрузку равномерно и исключить пиковые нагрузки на персонал.

Во-вторых, возможность хранения и анализа фотографий дефектов, загружаемых при сканировании кодов Data Matrix, создала основу для внедрения методов компьютерного зрения. В ходе эксперимента было обучено несколько моделей для автоматического распознавания типов дефектов (коррозия, трещины, утечки) по изображениям.

На момент завершения эксперимента точность классификации достигла 87%, что позволило сократить время на обработку результатов осмотра техническими экспертами примерно на 40%.

В-третьих, накопление временных рядов данных о состоянии активов в аналитической базе ClickHouse позволило существенно ускорить выполнение аналитических запросов. Если в традиционной ERP-системе построение сводного отчета по состоянию активов с детализацией до уровня подразделений занимало в среднем 15–20 минут, то в экспериментальной системе аналогичный отчет формировался за 3–5 секунд. Это создало предпосылки для перехода к интерактивным дашбордам, доступным для руководителей всех уровней в режиме реального времени.

Заключение. В ходе работы разработана архитектура системы учета основных средств, интегрирующая технологию матричных кодов Data Matrix с платформами Big Data. Предложенный подход предполагает переход от пакетной загрузки данных к непрерывной потоковой обработке событий, генерируемых при сканировании кодов, с обогащением их расширенным контекстом.

Экспериментальное исследование предложенной архитектуры проводилось исключительно с использованием инструментов искусственного интеллекта.

В рамках эксперимента были созданы синтетические наборы данных, имитирующие процессы сканирования кодов Data Matrix на парке оборудования численностью более 50 000 единиц, включая генерацию временных меток, геолокаций, фотографий дефектов и истории ремонтов. Само проведение эксперимента – от генерации тестовых данных до оценки результатов – осуществлялось с применением алгоритмов машинного обучения и нейросетевых моделей, что позволило полностью смоделировать поведение системы в различных сценариях без внедрения на реальном предприятии.

Результаты экспериментального моделирования показали, что предлагаемая архитектура обеспечивает сокращение времени инвентаризации на 78,6%, повышение точности учета местоположения активов до 99,2% и снижение затрат на внеплановые ремонты на 33,8% по сравнению с традиционным подходом. Полученные данные подтверждают, что интеграция Data Matrix и Big Data с применением методов искусственного интеллекта позволяет не только автоматизировать учет, но и создать основу для предиктивного управления активами.

Научная новизна работы заключается в разработке архитектуры, объединяющей технологии идентификации, потоковой обработки и аналитики больших данных.

Практическая значимость состоит в возможности использования предложенных решений для построения систем управления активами в организациях различных отраслей. Дальнейшие исследования предполагают расширение модельных сценариев и апробацию разработанной архитектуры в реальных условиях промышленных предприятий.

Список литературы

[1] Stonebraker M., Çetintemel U., Zdonik S. The 8 requirements of real-time stream processing // ACM SIGMOD Record. – 2005. – Vol. 34, No. 4. – P. 42–47.

[2] Kleppmann M. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. – Sebastopol: O'Reilly Media, 2017.

[3] Akidau T., Chernyak S., Lax R. Streaming Systems: The What, Where, When, and How of Large-Scale Data Processing. – Sebastopol: O'Reilly Media, 2018.

[4] Lee J., Bagheri B., Kao H.-A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems // Manufacturing Letters. – 2015. – Vol. 3. – P. 18–23.

[5] Ranjan R. Streaming big data processing in datacenter clouds // IEEE Cloud Computing. – 2014. – Vol. 1, No. 1. – P. 78–83.

Авторский вклад

Пискун Геннадий Адамович – постановка научной задачи исследования, формирование концепции системы учета основных средств на основе интеграции технологий Data Matrix и Big Data, разработка архитектуры системы и методологических подходов к потоковой обработке и аналитике данных, научное руководство исследованием и интерпретация полученных результатов.

Ершов Денис Геннадьевич – участие в разработке архитектуры системы учета основных средств с использованием технологий Data Matrix и Big Data, реализация и описание процессов сбора, передачи и потоковой обработки данных, проведение экспериментального моделирования, анализ и интерпретация полученных результатов, а также подготовка материалов статьи.

INTEGRATION OF BIG DATA PROCESSING IN THE AUTOMOTIVE INDUSTRY

G.A. Piskun
Deputy Dean of the Faculty of Computer Engineering for Research, Candidate of Technical Sciences, Associate Professor

D.G. Ershov
Master's student of the Belarusian State University of Informatics and Radioelectronics

Abstract. The article considers an approach to building a modern fixed assets accounting system that integrates automatic identification technology using two-dimensional Data Matrix codes with Big Data platform tools. The necessity of transitioning from local inventory systems to centralized analytical solutions enabling the collection, processing, and analysis of large volumes of asset condition data in real time is substantiated. An architectural scheme of component interaction is proposed, methods of streaming data processing are described, and the results of evaluating the effectiveness of the proposed approach are presented.

Keywords: fixed assets, inventory, Data Matrix, Big Data, stream processing, asset management, predictive analytics.

УДК 618.14-006.6:004.8

МОДЕЛЬ ОЦЕНКИ ПРОПУСКНОЙ СПОСОБНОСТИ И БЕЗОПАСНОЙ НАГРУЗКИ ЛАБОРАТОРИЙ ЖИДКОСТНОЙ ЦИТОЛОГИИ ПРИ ВНЕДРЕНИИ ИИ-АССИСТЕНТОВ





М.С. Печёнов
Магистрант кафедры электронной техники и технологии БГУИР
makspecheonov@gmail.com

В.М. Бондарик
Декан факультета доуниверситетской подготовки и профессиональной ориентации БГУИР, кандидат технических наук, доцент; доцент кафедры электронной техники и технологии БГУИР
bondarik@bsuir.by

М.С. Печёнов
Обучается в магистратуре Белорусского государственного университета информатики и радиоэлектроники по направлению «электронная техника и технологии». Область научных интересов связана с жидкостной цитологией, цифровой цитологией и внедрением ИИ в области медицины.

В.М. Бондарик
Окончил с отличием Минский радиотехнический институт. Кандидат технических наук, доцент, декан факультета доуниверситетской подготовки и профессиональной ориентации. Область научных интересов связана с проектированием медицинских электронных систем и внедрением программы подготовки к поступлению в университет.

Аннотация. Предложена модель оценки пропускной способности и безопасной нагрузки лаборатории жидкостной цитологии при внедрении ИИ-ассистентов. Модель объединяет нормативные ограничения по нагрузке (CLIA/FDA), правила подсчёта «эквивалента слайда» для полуавтоматизированного скрининга (FOV/FMR) и опубликованные оценки времени чтения. Выполнен сценарный расчёт для трёх режимов: ручного скрининга, полуавтоматизированного имиджинг-скрининга и цифровой цитологии на основе полноформатного сканирования препарата с формированием ИИ-галереи объектов интереса. Показано, как доля случаев полного ручного пересмотра ограничивает пропускную способность и почему для сохранения качества требуется запас по безопасности.

Ключевые слова: жидкостная цитология, пропускная способность, безопасная нагрузка, нормативы CLIA/FDA, эквивалент слайда, FOV/FMR, цифровая цитология, полноформатное сканирование препарата, искусственный интеллект.

Введение. В лабораторном скрининге жидкостной цитологии ИИ-ассистенты применяются для сокращения времени поиска и просмотра диагностически значимых объектов и повышения производительности. При этом рост производительности не является линейной функцией времени чтения: на практике действуют нормативные ограничения по нагрузке, а увеличение объёма исследований может сопровождаться снижением точности скрининга даже при использовании автоматизации. В связи с этим требуется расчётная модель, позволяющая сопоставлять технологические сценарии и обосновывать безопасный режим работы лаборатории.

Цель работы. Разработать расчётную модель оценки пропускной способности и безопасной нагрузки лаборатории жидкостной цитологии при внедрении ИИ-ассистентов и

выполнить сценарный анализ для полуавтоматизированного имиджинг-скрининга (FOV/FMR) и цифровой цитологии на основе полноформатного сканирования препарата с ИИ-подсказками (WSI+ИИ).

Материалы и исходные данные. Нормативная база включает требование CLIA об ограничении объема до 100 слайдов за смену и формулу пересчета при неполном рабочем дне [1], а также бюллетень FDA/CMS по подсчету нагрузки (workload) для полуавтоматизированного скрининга (FOV=0,5; FMR=1; FOV+FMR=1,5) [2]. Для имиджинг-подходов использованы материалы по ThinPrep Imaging System, где установлено ограничение 200 слайдов/8 ч при условии, что полный ручной пересмотр не превышает 100/8 ч [4]. Для сценария цифровой цитологии использованы инструкции по системе Hologic Genius, где описаны принципы формирования ИИ-галереи и нормативный эквивалент порядка 0,25 слайда на один просмотренный случай [5, 6]. Параметры сокращения времени чтения взяты из опубликованных исследований: − 42% для ThinPrep Imaging System и снижение времени чтения при переходе на систему Genius [7, 8]. Для оценки безопасности использованы работы, демонстрирующие снижение точности при увеличении дневной нагрузки, и рекомендации профессионального сообщества по предельной нагрузке и длительности активного скрининга [9–12].

Методика расчета. Рассматривается 8-часовая смена, однако фактическое время активного скрининга меньше 8 часов из-за перерывов и немикроскопических операций. В расчетах используется параметр Tscreen (мин) – эффективное время просмотра. В рамках сценарного анализа принято Tscreen = 420 мин (не более 7 часов активного скрининга) [12]. Для каждого сценария задаются tcase (мин/случай) и wcase (слайд-эквивалент на случай). Производительность по времени:

$$\text{Ntime} = = \text{Tscreen}/\text{tcase}.$$

Производительность по нормативу:

$$\text{Nnorm} = 100/\text{wcase}.$$

Итоговая пропускная способность:

$$\text{N} = \min(\text{Ntime}, \text{Nnorm}).$$

Для имиджинг-скрининга учитывается доля *p* случаев, направляемых на полный ручной пересмотр (FMR). Нормативная нагрузка на один случай определяется как

$$\text{wcase} = 0{,}5 + \text{p}$$

(по методике FDA/CMS). Среднее время чтения аппроксимируется как

$$\text{tcase} = \text{tFOV}(1 \quad \text{p}) + \text{tFMRp},$$

где tFOV − время чтения с использованием имиджинг-подсказок, а tFMR=tmanual − время полного ручного просмотра.

Для цифровой цитологии на основе полноформатного сканирования препарата с ИИ-галереей вводится доля *q* случаев, требующих дополнительной микроскопии при диагностической неопределенности [6]. В консервативной оценке принято:

$$\text{tcase} = \text{tWSI} + \text{qtmanual}$$

и

$$\text{wcase} = 0{,}25 + \text{q},$$

где tWSI − время чтения в цифровой системе, а микроскопический пересмотр учитывается как полный слайд-эквивалент.

Сценарный анализ. Сценарий ручного скрининга: tcase = 4,8 мин/случай (как 480 мин/100 слайдов), wcase = 1. Сценарий имиджинг-скрининга (ThinPrep): хронометражное исследование показывает снижение времени чтения на 42%, поэтому

$$\text{tFOV} = 0{,}58 \cdot \text{tmanual} = 2{,}78 \text{ мин};$$

$$\text{tFMR} = \text{tmanual} = 4{,}8 \text{ мин};$$

$$\text{wcase} = 0{,}5 + \text{p} \text{ [2,7]}.$$

Сценарий цифровой цитологии на основе полноформатного сканирования препарата с ИИ-галереей (Genius): для времени чтения принята оценка 70,1 с (1,17 мин) на один случай [8]; нормативный эквивалент wcase = 0,25 [5,6]. Доля q случаев, требующих дополнительной микроскопии, рассматривается сценарно (0 – 0,10); для консервативной оценки принято

$$\mathrm{tcase} = \mathrm{tWSI} + \mathrm{qtmanual}$$

и

$$\mathrm{wcase} = 0{,}25 + \mathrm{q}.$$

Сценарные результаты расчёта максимальной пропускной способности приведены в таблице 1. Для каждого режима работы рассчитаны значения Ntime (ограничение по времени чтения) и Nnorm (ограничение по нормативной нагрузке), а итоговая оценка определяется как

$$\mathrm{N} = \min(\mathrm{Ntime}, \mathrm{Nnorm})$$

при Tscreen = 420 мин. Для имиджинг-скрининга варьируется доля случаев полного ручного пересмотра $p_{(\mathrm{FMR})}$ а для цифровой цитологии – доля случаев дополнительной микроскопии q, а для цифровой цитологии – доля случаев дополнительной микроскопии.

Таблица 1. Сценарный расчет максимальной пропускной способности (случаев/смена) при Tscreen = 420 мин

Сценарий	$p_{(\mathrm{FMR})}$	$\mathrm{q}_{(\text{микроск.})}$	$\mathrm{t}_{\mathrm{case}}$, мин	$\mathrm{w}_{\mathrm{case}}$	$\mathrm{N}_{\mathrm{time}}$	$\mathrm{N}_{\mathrm{norm}}$	N
Ручной скрининг	–	–	4,80	1,00	88	100	88
Имиджинг-скрининг (ThinPrep)	0,10	–	2,99	0,60	141	167	141
Имиджинг-скрининг (ThinPrep)	0,15	–	3,09	0,65	136	154	136
Имиджинг-скрининг (ThinPrep)	0,20	–	3,19	0,70	132	143	132
WSI+ИИ (Genius)	–	0,00	1,17	0,25	359	400	359
WSI+ИИ (Genius)	–	0,05	1,41	0,30	298	333	298
WSI+ИИ (Genius)	–	0,10	1,65	0,35	255	286	255

Для практического внедрения важен не только максимум, но и «безопасный» режим работы, учитывающий риск снижения качества при перегрузке персонала. Поэтому дополнительно выполнена оценка безопасной пропускной способности при ограничении суммарной нагрузки 70 слайд-эквивалентами за смену. Результаты приведены в таблице 2: рассчитаны

$$\mathrm{Nsafe_norm} = 70/\mathrm{wcase}$$

и итоговая безопасная оценка

$$\mathrm{Nsafe} = \min(\mathrm{Ntime}, \mathrm{Nsafe_norm}),$$

что позволяет сопоставить сценарии с точки зрения реалистичной производительности при соблюдении запаса по безопасности.

Таблица 2. Оценка «безопасной» пропускной способности при ограничении 70 слайд-эквивалентов/смена

Сценарий	$p_{(\mathrm{FMR})}$	$q_{(\text{микроск.})}$	$\mathrm{w}_{\mathrm{case}}$	$\mathrm{N}_{\mathrm{time}}$	$\mathrm{N}_{\mathrm{safe_norm}}$	$\mathrm{N}_{\mathrm{safe}}$
Ручной скрининг	–	–	1,00	88	70	70
Имиджинг-скрининг (ThinPrep)	0,10	–	0,60	141	117	117
Имиджинг-скрининг (ThinPrep)	0,15	–	0,65	136	108	108
Имиджинг-скрининг (ThinPrep)	0,20	–	0,70	132	100	100
WSI+ИИ (Genius)	–	0,00	0,25	359	280	280
WSI+ИИ (Genius)	–	0,05	0,30	298	233	233
WSI+ИИ (Genius)	–	0,10	0,35	255	200	200

Обсуждение результатов. Таблица 1 показывает, что при Tscreen = 420 мин ручной скрининг обеспечивает около 88 случаев/смена и в основном ограничивается временем. Имиджинг-подход повышает пропускную способность до 132 – 141 случаев/смена; при росте $p_{(FMR)}$ одновременно усиливаются два ограничения: увеличивается среднее время чтения и растёт wcase, что снижает Nnorm. Цифровая цитология на основе полноформатного сканирования с ИИ-галереей (WSI+ИИ) обеспечивает наибольший потенциал пропускной способности: при $q = 0$ достигается около 359 случаев/смена, а при $q = 0{,}10$ (10% случаев с дополнительной микроскопией) сохраняется около 255 случаев/смена. При увеличении доли микроскопии становится более заметным нормативное ограничение, поскольку возрастает wcase и уменьшается Nnorm. Таблица 2 иллюстрирует практический смысл «запаса по безопасности»: профессиональные рекомендации предлагают ориентироваться на уровень порядка 70 слайд-эквивалентов в сутки (по методике FDA/CMS), поскольку рост нагрузки ассоциируется со снижением точности [9–12]. В таком режиме WSI+ИИ сохраняет преимущество и одновременно снижает риск деградации качества.

Практическое применение модели. Модель может использоваться как инструмент планирования и принятия решений:

1. по данным локального хронометража (например, 30–50 случаев) оцениваются tcase и доли p (FMR – полный ручной пересмотр) и q (дополнительная микроскопия);

2) выбирается целевой уровень безопасности (например, 70 слайд-эквивалентов/смена);

3) рассчитывается требуемое число специалистов для прогнозируемого потока случаев;

4) сопоставляются сценарии внедрения (ручной скрининг, имиджинг-скрининг, цифровая цитология WSI+ИИ) с учетом того, что при росте p и q преимущество автоматизации частично нивелируется.

Для повышения достоверности рекомендуется учитывать немикроскопические задачи (QC-пересмотры, оформление протоколов, коммуникации), что уменьшает эффективное Tscreen.

Дополнительно целесообразно выполнять анализ чувствительности модели по параметрам p и q (например, в диапазоне 0 – 0,20) и фиксировать результаты в виде сценарных таблиц для управленческих решений. Такой подход позволяет заранее оценивать влияние изменений в потоке исследований и правилах направления на пересмотр на ожидаемую пропускную способность и нагрузку.

Ограничения. Расчёт фокусируется на этапе чтения и нормативной нагрузке и не включает время подготовки препаратов и время сканирования, а также особенности конкретного лабораторного протокола. В сценарии WSI+ИИ принята консервативная трактовка: доля q случаев требует дополнительной микроскопии, которая учитывается как добавочное время и как полный слайд-эквивалент, что может занижать оценку максимальной пропускной способности. Параметры p, q и t_{case} имеют популяционную и организационную вариативность (состав потока, критерии направления на пересмотр, качество сканирования), поэтому итоговые значения пропускной способности следует рассматривать как ориентировочные до проведения локального хронометражного исследования и валидации.

Заключение. Предложенная модель позволяет формально оценивать пропускную способность и безопасную нагрузку лаборатории жидкостной цитологии при внедрении ИИ-ассистентов.

Сценарные расчёты показывают рост производительности при переходе от ручного скрининга к имиджинг-скринингу и далее к цифровой цитологии на основе полноформатного сканирования препарата с ИИ-подсказками (WSI+ИИ). Ключевыми факторами, ограничивающими эффект, являются доля полного ручного пересмотра (FMR,

p) и доля случаев дополнительной микроскопии (q), которые увеличивают среднее время чтения и нормативную нагрузку на один случай.

Для практического внедрения рекомендуется сочетать расчёт максимальной пропускной способности с ограничением по «безопасному» уровню нагрузки и подтверждать параметры модели локальным хронометражем и валидацией.

В рассматриваемых сценариях при Tscreen = 420 мин оценка максимальной пропускной способности составляет порядка 88 случаев/смена для ручного скрининга, 132–141 для имиджинг-скрининга и до 255–359 для WSI+ИИ (в зависимости от q).

Список литературы

[1] 42 CFR § 493.1274. Standard: Cytology [Электронный ресурс] // Electronic Code of Federal Regulations. – 2026. – Дата обращения: 10.01.2026.

[2] How Laboratorians Can Safely Calculate Workload for FDA-Approved Semi-Automated Gynecologic Cytology Screening Devices. FDA/CMS Safety Tip [Электронный ресурс] // Centers for Medicare & Medicaid Services (CLIA). – 2015. – Дата обращения: 15.01.2026.

[3] Summary of Safety and Effectiveness Data (SSED): BD FocalPoint™ GS Imaging System (P950009/S008) [Электронный ресурс] // U.S. Food and Drug Administration. – 2008. – Дата обращения: 20.01.2026.

[4] ThinPrep® Imaging System. Summary of Safety and Effectiveness Data (SSED) (P020002) [Электронный ресурс] // U.S. Food and Drug Administration. – 2003. – Дата обращения: 25.01.2026.

[5] Genius™ Digital Diagnostics System. Instructions for Use. MAN-11010-001 Rev. 002 (05-2024) [Электронный ресурс] // Hologic. – 2024. – Дата обращения: 30.01.2026.

[6] Genius™ Digital Diagnostics System with the Genius™ Cervical AI Algorithm. Instructions for Use. AW-23890-001 Rev. 005 (07-2025) [Электронный ресурс] // Hologic. – 2025. – Дата обращения: 05.02.2026.

[7] Schledermann, D. Automated screening versus manual screening: a comparison of the ThinPrep Imaging System and manual screening in a time study / D. Schledermann, T. Hyldebrandt, D. Ejersbo, B. Hoelund // Diagn Cytopathol. – 2007. – № 35(6). – PP. 348–352. – DOI: 10.1002/dc.20640. – Дата обращения: 10.02.2026.

[8] Murphy, K. M. Impact of the Genius Digital Diagnostics System on workflow and accuracy compared with the ThinPrep Imaging System for review of ThinPrep Papanicolaou tests / K. M. Murphy [et al.] // Am J Clin Pathol. – 2025. – № 164(5). – PP. 746–751. – DOI: 10.1093/ajcp/aqaf099. – Дата обращения: 15.02.2026.

[9] Elsheikh, T. M. Increasing cytotechnologist workload above 100 slides per day using the ThinPrep Imaging System leads to significant reductions in screening accuracy / T. M. Elsheikh [et al.] // Cancer Cytopathol. – 2010. – № 118(2). – PP. 75–82. – DOI: 10.1002/cncy.20065. – Дата обращения: 18.02.2026.

[10] Levi, A. W. Increasing cytotechnologist workload above 100 slides per day using the BD FocalPoint GS imaging system negatively affects screening performance / A. W. Levi [et al.] // Am J Clin Pathol. – 2012. – № 138(6). – PP. 811–815. – DOI: 10.1309/AJCPTYDD9G2NCUMF. – Дата обращения: 20.02.2026.

[11] Moriarty, A. T. Cytology workload calculation–Has anything really changed? / A. T. Moriarty // Cancer Cytopathol. – 2011. – № 119(2). – PP. 77–79. – DOI: 10.1002/cncy.20141. – Дата обращения: 25.02.2026.

[12] Elsheikh, T. M. American Society of Cytopathology workload recommendations for automated Pap test screening / T. M. Elsheikh [et al.] // Diagn Cytopathol. – 2013. – № 41(2). – PP. 174–178. – DOI: 10.1002/dc.22817. – Дата обращения: 01.03.2026.

[13] Medical Services Advisory Committee (MSAC). Liquid-based cytology for cervical cancer screening. Assessment report [Электронный ресурс] // MSAC. – 2009. – Дата обращения: 05.03.2026.

[14] Renshaw, A. A. Time consumed by microscopic and nonmicroscopic tasks in image-assisted gynecologic screening: Implications for workload assessment / A. A. Renshaw [et al.] // Cancer Cytopathol. – 2016. – № 124(7). – PP. 501–507. – DOI: 10.1002/cncy.21711. – Дата обращения: 10.03.2026.

Авторский вклад

Печёнов Максим Сергеевич – постановка задачи исследования, сбор и систематизация нормативных и научных источников, разработка расчётной модели оценки пропускной способности и безопасной нагрузки лаборатории жидкостной цитологии, проведение сценарных расчётов для различных режимов скрининга, анализ и интерпретация полученных результатов, подготовка и оформление текста статьи.

Бондарик Василий Михайлович – научное руководство и консультации по постановке задачи и структуре исследования, формирование требований к модели и допущениям расчёта, критический анализ результатов и выводов, редактура текста статьи и рекомендации по направлениям дальнейшего развития модели и практического внедрения в лабораторный процесс.

MODEL FOR ESTIMATING THROUGHPUT AND SAFE WORKLOAD IN LIQUID-BASED CYTOLOGY LABORATORIES WHEN INTRODUCING AI ASSISTANTS

M.S. Pechenov
Master's student, Department of Electronic Engineering and Technology, BSUIR

V.M. Bondarik
Dean of the Faculty of Pre-University Training and Career Guidance, BSUIR, PhD in Engineering, Associate Professor, Department of Electronic Engineering and Technology

Abstract. A model is proposed to estimate throughput and safe workload in liquid-based cytology laboratories when introducing AI assistants. The model combines CLIA/FDA workload constraints, slide-equivalent counting rules for semi-automated screening (FOV/FMR), and published reading-time estimates. A scenario analysis is performed for three modes: manual screening, semi-automated imaging-assisted screening, and digital cytology based on whole-slide imaging with an AI-generated gallery of objects of interest. The results show how the full manual review rate limits throughput and why a safety margin is required to maintain screening quality.

Keywords: liquid-based cytology, throughput, safe workload, CLIA/FDA, slide equivalent, FOV/FMR, digital cytology, whole-slide imaging, artificial intelligence.

UDC 004.032.26:616.12-073.97

AUTOMATED CARDIAC RHYTHM CLASSIFICATION FROM 12-LEAD ECG USING ATTENTION-ENHANCED DEEP NEURAL NETWORKS





N.L. Kisialeu
Student, Faculty of Applied Mathematics and Computer Science, Belarusian State University, Republic of Belarus
nkisialeu@gmail.com

A.M. Nedzved
Head of the Department, Associate Professor, Doctor of Technical Sciences, Belarusian State University, Republic of Belarus
Anedzved@bsu.by

N.L. Kisialeu
Student at the Faculty of Applied Mathematics and Computer Science, Belarusian State University. Research interests: deep learning, medical signal processing, computer vision, time series classification, automated machine learning.

A.M. Nedzved
Head of the Department, Associate Professor, Doctor of Technical Sciences, Belarusian State University. Research interests: image processing, computer vision, pattern recognition, mathematical morphology, biomedical image analysis.

Abstract. This paper presents a comprehensive deep learning framework for automated multi-class ECG rhythm classification using 12-lead electrocardiogram recordings from the Chapman–Shaoxing database (10,615 records, 8 rhythm classes). A 1D Squeeze-and-Excitation ResNet-18 (SE-ResNet18) architecture is proposed for temporal feature extraction. Seven experimental configurations are systematically evaluated, investigating the impact of clinical metadata (14 features), NeuroKit2 [1] morphological features, and demographic inputs on classification performance. The best configuration achieves test accuracy of 94.2%, macro AUC of 0.956, and macro F1-score of

0.851. A universal lead-agnostic model (UniversalECGNet) with shared single-channel backbone, learnable lead embeddings, and attention pooling enables deployment across devices with 1–12 leads without retraining.

Keywords: ECG classification, deep learning, SE-ResNet18, squeeze-and-excitation, cardiac arrhythmia, 12-lead electrocardiogram, clinical metadata, universal ECG model, attention pooling, lead dropout.

Introduction. Electrocardiography (ECG) remains the most widely used non-invasive diagnostic tool in cardiology, providing critical information about cardiac rhythm, conduction abnormalities, and myocardial ischemia. Clinical ECG interpretation is performed by cardiologists who visually analyze waveform morphology, intervals, and rhythm regularity – a process that is time-consuming, subject to inter-observer variability, and dependent on specialist availability. Automated rhythm classification from ECG recordings is therefore a critical component of computer-aided diagnosis systems, enabling timely detection of arrhythmias including atrial fibrillation (AFIB), sinus bradycardia (SB), sinus tachycardia (ST), atrial flutter (AF), and other clinically significant conditions [2].

According to the World Health Organization, cardiovascular diseases account for approximately 19.8 million deaths in 2022, representing 32% of all global deaths [3]. Early and accurate detection of cardiac rhythm abnormalities through ECG analysis plays a critical role in timely intervention and treatment planning. However, manual ECG interpretation is subject to considerable inter-observer variability, particularly for complex arrhythmias, which motivates the development of automated classification systems.

Traditional rule-based approaches to ECG classification rely on hand-crafted feature extraction pipelines: QRS complex detection, P-wave and T-wave delineation, interval measurement (PR, QT, QRS), and axis calculation, followed by threshold-based or decision-tree classification. While interpretable, these methods lack generalization across diverse patient populations, recording conditions, and device configurations. Deep learning methods, particularly convolutional neural networks (CNNs), have demonstrated superior performance by learning discriminative features directly from raw signal data, eliminating the need for manual feature engineering [4]. Hannun et al. demonstrated cardiologist-level arrhythmia detection using a 34-layer CNN trained on 91,232 single-lead ECG recordings [2]. Ribeiro et al. extended this to 12-lead ECGs using a residual network, achieving diagnostic accuracy comparable to cardiology residents [5].

The PhysioNet/Computing in Cardiology Challenge 2020 [6] established standardized benchmarks for 12-lead ECG classification across multiple diagnostic categories. The top-performing solutions employed deep residual networks augmented with attention mechanisms, achieving challenge scores above 0.5 on a weighted multi-label metric. Notably, the 2nd-place solution combined ResNet-18 with Squeeze-and-Excitation (SE) blocks [7], demonstrating that channel-wise attention recalibration significantly improves feature discrimination for ECG signals. However, most existing approaches are designed for a fixed lead configuration (typically 12-lead) and cannot adapt to variable input channels. This is a significant limitation in clinical practice, where ECG acquisition devices range from single-lead wearable monitors (Apple Watch, AliveCor KardiaMobile) to standard 3-lead, 6-lead, and 12-lead hospital systems.

This work addresses three research objectives: (1) developing an SE-ResNet18 architecture optimized for multi-class ECG rhythm classification with systematic investigation of input modalities across seven experimental configurations; (2) analyzing the contribution of clinical metadata, morphological features extracted via NeuroKit2, and minimal demographic information to classification quality; and (3) proposing a universal lead-agnostic model (UniversalECGNet) that operates on arbitrary subsets of 1–12 ECG leads using a shared backbone with attention-based lead aggregation and lead dropout augmentation.

The remainder of this paper is organized as follows. Section 2 describes the dataset characteristics, class distribution, and preprocessing pipeline. Section 3 presents the SE-ResNet18 architecture with detailed description of the squeeze-and-excitation mechanism. Section 4 introduces the UniversalECGNet architecture with its shared backbone, lead embeddings, and

attention pooling mechanism. Section 5 details the training configuration including optimizer, scheduler, loss function, and augmentation strategies. Section 6 presents experimental results across seven configurations with comparative tables. Section 7 provides per-class analysis with confusion patterns. Section 8 discusses the implications of the findings, and Section 9 concludes with a summary and future directions.

Dataset and preprocessing. The Chapman–Shaoxing 12-lead ECG database from PhysioNet [8, 9] is used in this study. The database contains 12-lead ECG recordings from patients at Shaoxing People's Hospital, sampled at 500 Hz with 10-second duration (5,000 samples per channel per recording). Each recording is accompanied by diagnostic labels and clinical metadata provided in the accompanying metadata files. After filtering rare rhythm classes with insufficient representation for reliable training (AVNRT – atrioventricular nodal reentrant tachycardia, AVRT – atrioventricular reentrant tachycardia, SAAWR – sinus atrium to atrial wandering rhythm), 10,615 records across 8 rhythm classes are retained.

The class distribution exhibits significant imbalance (Table 1). The ratio between the largest (SB) and smallest (AT) classes is 32:1, presenting a challenge for model training. Stratified splitting with random seed 42 preserves class proportions across train/validation/test sets of 7,430/1,592/1,593 records.

Table 1. Class distribution in the Chapman–Shaoxing dataset

Class	**Abbr.**	**Total**	**Train**	**Val**	**Test**
Sinus Bradycardia	SB	3,889	2,722	583	584
Sinus Rhythm	SR	1,826	1,278	274	274
Atrial Fibrillation	AFIB	1,780	1,246	267	267
Sinus Tachycardia	ST	1,568	1,098	235	235
Supraventr. Tachycardia	SVT	587	411	88	88
Atrial Flutter	AF	445	311	67	67
Sinus Arrhythmia	SA	399	279	60	60
Atrial Tachycardia	AT	121	85	18	18

Signal preprocessing includes per-channel normalization to zero mean and unit variance using training set statistics. Clinical metadata comprises 14 features: PatientAge, Gender (one-hot encoded as 2 binary features), VentricularRate, AtrialRate, QRSDuration, QTInterval, QTCorrected, RAxis, TAxis, QRSCount, QOnset, QOffset, TOffset, and POnset. All metadata features are standardized using z-score normalization. Additionally, morphological features are extracted using the NeuroKit2 library: R-peak amplitudes and locations, P/Q/S/T wave boundary positions, RR interval statistics, and heart rate variability (HRV) metrics – 28 features in total. Feature extraction is precomputed and cached in a CSV file. Analysis revealed that 10–50% of NeuroKit2 features contain fallback values due to detection failures on certain recordings, motivating both unfiltered and filtered (>5% fallback threshold) experimental variants.

SE-ResNet18 architecture. The core classification architecture is a 1D adaptation of ResNet-18 [10] augmented with Squeeze-and-Excitation (SE) blocks [7], inspired by the 2nd-place solution of the PhysioNet 2020 Challenge. All 2D convolutions in the original ResNet-18 are replaced with 1D convolutions (Conv1d) with kernel size 7 and stride 2 in the initial layer. The network processes 12-channel input tensors of shape (12, 5,000) through an initial convolutional layer (64 filters) followed by batch normalization, ReLU activation, and max pooling (kernel 3, stride 2). Four residual stages with [2, 2, 2, 2] basic blocks contain 64, 128, 256, and 512 filters respectively. Each residual block consists of two Conv1d layers with batch normalization and ReLU, with a skip connection adding the input to the output.

Each SE block enhances the residual block output through channel-wise attention recalibration. First, global average pooling along the temporal dimension produces a channel descriptor vector of size C (where C is the number of filters). This descriptor passes through a two-

layer fully connected bottleneck: FC(C → C/r) with ReLU activation, followed by FC(C/r → C) with sigmoid activation, where r = 16 is the reduction ratio. The resulting sigmoid-activated vector provides channel-wise scaling factors that multiply the original feature map, enabling the network to emphasize diagnostically relevant temporal patterns (e.g., P-wave morphology, QRS duration, T-wave alternans) while suppressing uninformative channels. Adaptive average pooling reduces the temporal dimension to 1, producing a 512-dimensional feature vector for the classification head.

For metadata integration (SE-ResNet18-Meta variant), clinical features are processed through a separate MLP branch: Linear(N_features → 64) → ReLU → Dropout(0.3), producing a 64-dimensional metadata embedding. This embedding is concatenated with the 512-dimensional CNN feature vector, forming a 576-dimensional combined representation that passes through the final classification layer Linear(576 → 8). The MLP branch input dimension varies by experiment: 14 for full clinical metadata, 36 for metadata + all NeuroKit2 features, 27 for metadata + filtered NeuroKit2 features, 3 for age + gender only, and 16 for age + gender + filtered NeuroKit2 features.

Universal lead-agnostic model (UniversalECGNet). A key contribution of this work is UniversalECGNet –a single model that operates on any subset of 1–12 ECG leads without retraining. The architecture (9.18M parameters) employs a shared single-channel SE-ResNet18 backbone that processes each lead independently. The processing pipeline consists of seven steps:

Step 1. Input tensor of shape (B, L, 5000), where B is batch size and L is the number of available leads (1 to 12), along with lead index tensor (B, L) identifying which leads are present (0=I, 1=II, 2=III, 3=aVR, 4=aVL, 5=aVF, 6=V1, ..., 11=V6). Step 2. Reshape to (B×L, 1, 5000), treating each lead as an independent single-channel sample. Step 3. Process through shared SE-ResNet18(in_channels=1) with gradient checkpointing for memory efficiency, producing per-lead features of shape (B×L, 512). Step 4. Reshape back to (B, L, 512). Step 5. Learnable lead embedding layer maps lead indices to 64-dimensional vectors encoding each lead's identity and spatial characteristics: Embedding(12, 64). Step 6. Concatenate CNN features with lead embeddings: (B, L, 576), then project via Linear(576 → 512): (B, L, 512). Step 7. Attention pooling with a learned 512-dimensional query vector computes attention scores over leads via dot product, applies masked softmax (masking absent leads), and produces weighted sum: single 512-dimensional patient representation regardless of input lead count. Final classification: Linear(512 → 8).

Lead dropout augmentation is critical for training robustness. During training, 50% of batches see all 12 leads, while the remaining 50% see a random subset of 1–11 leads (uniformly sampled count, then random lead selection). Masked leads are zeroed out and excluded from attention computation via the mask. This augmentation forces the model to extract useful representations from arbitrary lead combinations rather than overfitting to the full 12-lead configuration. Ablation experiments confirm that without lead dropout, the universal model maintains 12-lead performance but degrades sharply (−10–15% accuracy) when evaluated on reduced lead configurations, demonstrating the augmentation's essential role in achieving lead-agnostic generalization.

The attention pooling mechanism constitutes the central innovation of UniversalECGNet. The LeadAttentionPool module implements a two-layer neural network: a linear transformation from the 512-dimensional feature space to a 128-dimensional hidden space with Tanh activation, followed by a projection to a single scalar score per lead. For available leads, these scores are normalized via softmax to produce attention weights summing to 1.0; for absent leads (either masked during lead dropout training or physically unavailable during inference), scores are set to negative infinity before softmax, effectively zeroing their contribution. The final patient-level representation is computed as the attention-weighted sum of per-lead feature vectors. Analysis of the learned attention weights on the test set reveals that lead II consistently receives the highest average attention weight (0.15–0.18), followed by V1 (0.10–0.14) and aVF (0.09–0.12), which aligns with clinical knowledge about the diagnostic value of these leads for rhythm classification.

The shared backbone architecture presents a computational challenge during training: with batch size B and L leads per sample, the effective batch processed by the backbone is B×L, reaching up to 32×12 = 384 forward passes per training step. Gradient checkpointing is applied to the four residual stages to reduce peak GPU memory consumption by approximately 60% at the cost of ~30% additional computation time. The total parameter count of UniversalECGNet is 9,179,977 parameters: the shared SE-ResNet18 backbone accounts for 9,065,728 (98.8%), the learnable lead embedding layer contributes 768 parameters (12 leads × 64 dimensions), the feature projection layer adds 295,424, and the attention pooling module contains 66,049 parameters. During inference, standard forward propagation without checkpointing is used for maximum throughput. The lead embedding vectors are L2-normalized before concatenation with the signal, ensuring that the identity information does not dominate the raw ECG features in magnitude.

Training configuration. All models are trained using the AdamW optimizer [11] with initial learning rate 0.001 and weight decay 0.0001. The learning rate schedule employs CosineAnnealingWarmRestarts with $T_0 = 10$ and T_mult = 2, providing periodic warm restarts that help escape local minima and explore the loss landscape more thoroughly. The loss function is cross-entropy with inverse-frequency class weights (computed from training set class distribution) and label smoothing $\varepsilon = 0.1$, which prevents overconfident predictions and improves calibration. Gradient clipping at max norm 1.0 ensures training stability.

Signal augmentations are applied online during training with per-sample probability of 0.5 each: (1) baseline wander injection –sinusoidal signal at 0.1–1.0 Hz simulating respiratory and movement artifacts; (2) 50 Hz power line noise –additive sinusoidal at exactly 50 Hz with random amplitude; (3) Gaussian noise at SNR 20–40 dB simulating electronic noise; (4) amplitude scaling by a random factor in [0.8, 1.2] simulating gain variations across devices. These augmentations improve model robustness to real-world recording conditions. Early stopping monitors macro F1 on the validation set with patience of 10 epochs (15 for NeuroKit2 variants due to slower convergence). Batch size is 64 for standard models and 32 for the universal model due to higher memory requirements. The complete set of training hyperparameters and augmentation settings is provided in Table 2.

Table 2. Training hyperparameters and augmentation configuration

Parameter	**Value**
Optimizer	AdamW
Learning rate	0.001
Weight decay	0.0001
Scheduler	CosineAnnealingWarmRestarts (T_0=10, T_mult=2)
Loss function	CrossEntropy + class weights + label smoothing (ε=0.1)
Gradient clipping	max norm 1.0
Batch size	64 (standard) / 32 (universal)
Early stopping	patience 10 on val macro F1
Baseline wander	sinusoidal, 0.1–1.0 Hz
Power line noise	50 Hz additive sinusoidal
Gaussian noise	SNR 20–40 dB
Amplitude scaling	±20%
Lead dropout (universal)	p = 0.5

Experimental results. Seven experimental configurations are systematically evaluated on the held-out test set (N = 1,593). All experiments share the same data splits, augmentations, and training hyperparameters (except input dimensions), enabling fair comparison of input modalities.

A summary of all experimental configurations and their test set performance metrics is presented in Table 3.

Table 3. Comparative results across seven experimental configurations (test set, N = 1,593)

#	Configuration	Input	Acc.	P (macro)	F1 (macro)	AUC (macro)
1	Signal only	ECG 12×5000	0.938	0.862	0.859	0.946
2	Signal + metadata	ECG + 14 meta	0.942	0.855	0.851	0.956
3	+ all NeuroKit2	ECG + 36 feat	0.932	0.849	0.850	0.968
4	+ filtered NeuroKit2	ECG + 27 feat	0.923	0.836	0.831	0.966
5	Signal + age/gender	ECG + 3 feat	0.935	0.841	0.843	0.937
6	+ age/gender + NK2	ECG + 16 feat	0.935	0.829	0.846	0.965
7	Universal (12-lead)	1–12 leads	0.936	0.841	0.838	0.953

Experiment 1 (signal only): SE-ResNet18 with raw 12-lead ECG input, no additional features. Accuracy 93.8%, macro precision 0.862, macro F1 0.859, macro AUC 0.946. This establishes a strong baseline demonstrating that SE-ResNet18 effectively extracts discriminative features from raw signals.

Experiment 2 (signal + 14 metadata): SE-ResNet18-Meta with MLP branch (14 → 64) for clinical features. This achieves the best overall performance: accuracy 94.2%, macro precision 0.855, macro recall 0.850, macro F1 0.851, weighted F1 0.942, macro AUC 0.956, Cohen's kappa 0.926, MCC 0.926, log loss 0.466, top-2 accuracy 0.976. The clinical metadata – particularly ventricular rate, QRS duration, QT interval, and patient age – provides complementary information that the CNN cannot fully extract from raw signals alone.

Detailed metrics for the best model (Experiment 2) are presented in Table 4.

Table 4. Global metrics of the best model (Experiment 2, test set)

Metric	Value
Accuracy	0.9422
Precision (macro)	0.8548
Recall (macro)	0.8495
F1-score (macro)	0.8512
F1-score (weighted)	0.9415
AUC (macro, OvR)	0.9562
Cohen's Kappa	0.9259
MCC	0.9260
Log Loss	0.4662
Top-2 Accuracy	0.9755

Experiment 3 (signal + metadata + all NeuroKit2): MLP branch with 36 inputs (14 metadata + 28 morphological features). Accuracy drops to 93.2% while AUC rises to 0.968, suggesting better ranking ability but worse calibration. The 10–50% fallback values in NeuroKit2 features introduce noise that dilutes the useful metadata signal. Experiment 4 (filtered NeuroKit2): columns with >5% fallback are removed, leaving 27 features. Accuracy further drops to 92.3%, AUC 0.966 – even filtered features do not improve over pure metadata.

Experiment 5 (signal + age/gender): minimal demographic features (age + 2 gender one-hot = 3 inputs). Accuracy 93.5%, AUC 0.937. Age and gender provide marginal benefit over signal-only, confirming that clinical metadata's value comes primarily from cardiac-specific measurements (rates, intervals, axes) rather than demographics. Experiment 6 (age/gender + filtered NeuroKit2): 16 features total, accuracy 93.5%, AUC 0.965 – NeuroKit2 features partially compensate for missing clinical metadata but do not exceed the full metadata configuration.

Experiment 7 (UniversalECGNet): the lead-agnostic model is evaluated on four lead configurations without any retraining between configurations. 12-lead: accuracy 93.6%, macro F1 0.838, AUC 0.953. 6-lead (limb leads I, II, III, aVR, aVL, aVF): accuracy 93.7%, F1 0.837 – virtually no degradation, demonstrating that limb leads carry nearly all discriminative rhythm

information. 3-lead (I, II, V1): accuracy 93.8%, F1 0.842 – slightly better than 12-lead, as reduced lead count eliminates noise from redundant leads and allows attention to focus on the most informative subset. Single-lead (II): accuracy 91.6%, F1 0.787, AUC 0.941 – F1 drops by 5 percentage points, but lead II alone still achieves strong AUC, confirming it carries the most rhythm-discriminative information.

Table 5 summarizes the universal model's performance across four lead configurations evaluated without any retraining.

Table 5. UniversalECGNet: performance across lead configurations (test set)

Config.	**Leads**	**Accuracy**	**F1 (macro)**	**F1 (wtd)**	**AUC (macro)**
12-lead	I, II, III, aVR, aVL, aVF, V1–V6	0.936	0.838	0.936	0.953
6-lead	I, II, III, aVR, aVL, aVF	0.937	0.837	0.936	0.945
3-lead	I, II, V1	0.938	0.842	0.939	0.951
1-lead	II	0.916	0.787	0.915	0.941

Per-class analysis and error patterns. Detailed per-class analysis of the best model (Experiment 2) reveals distinct performance tiers. Well-represented classes achieve excellent classification: SB (sensitivity 0.990, specificity 0.997, F1 0.992), ST (sensitivity 0.975, specificity 0.992, F1 0.964), AFIB (sensitivity 0.959, specificity 0.985, F1 0.943), SR (sensitivity 0.938, specificity 0.988, F1 0.940), and SVT (sensitivity 0.932, specificity 0.995, F1 0.921).

Per-class sensitivity, specificity, positive predictive value (PPV), and F1-score for the best model are shown in Table 6.

Table 6. Per-class metrics of the best model (Experiment 2, test set)

Class	**Sens.**	**Spec.**	**PPV**	**F1**	**N (test)**
SB	0.990	0.997	0.995	0.992	584
SR	0.938	0.988	0.941	0.940	274
AFIB	0.959	0.985	0.928	0.943	267
ST	0.975	0.992	0.954	0.964	235
SVT	0.932	0.995	0.911	0.921	88
AF	0.642	0.990	0.741	0.688	67
SA	0.750	0.994	0.818	0.783	60
AT	0.611	0.994	0.550	0.579	18

Minority classes show clinically interpretable error patterns. Atrial flutter (AF, 67 test samples) achieves sensitivity 0.642, specificity 0.990, F1 0.688. Confusion matrix analysis reveals that 12 of 67 AF samples are misclassified as AFIB – a clinically expected confusion since both conditions involve abnormal atrial activity with similar morphological features on ECG, differing primarily in regularity and atrial rate (AF: regular ~300 bpm, AFIB: irregular ~350–600 bpm). Sinus arrhythmia (SA, 60 test samples) achieves F1 0.783, with 13 of 60 samples misclassified as SR – expected since SA is a physiological variant of sinus rhythm where the only distinguishing feature is respiratory-linked heart rate variability. Atrial tachycardia (AT, 18 test samples) shows the lowest F1 of 0.579, primarily due to extreme class imbalance (only 121 training samples) and morphological overlap with SVT and ST.

Discussion. The experimental results yield several important findings. First, the SE attention mechanism provides consistent improvement over standard ResNet-18, with minority class F1 gains of 3–5% – the channel-wise recalibration helps the network focus on subtle morphological differences that distinguish challenging class pairs (AF vs. AFIB, SA vs. SR). Second, clinical metadata from the diagnostic system provides the highest-value supplementary information, with

the 14-feature metadata configuration outperforming all NeuroKit2 variants. This is likely because the diagnostic system's measurements are more reliable than automated feature extraction from raw signals, which suffers from detection failures on certain recording types.

Third, the universal model's near-zero degradation from 12 to 6 leads has significant clinical implications: standard limb leads alone are sufficient for accurate rhythm classification, suggesting that precordial leads (V1–V6) are largely redundant for this task. The counterintuitive improvement at 3 leads can be attributed to the attention mechanism's ability to focus more precisely when fewer leads are available, avoiding noise from redundant channels. The lead dropout augmentation is the key enabler of this generalization – without it, the model memorizes the 12-lead input structure and fails catastrophically on reduced configurations.

Fourth, the primary performance bottleneck remains class imbalance for rare arrhythmias. Atrial Tachycardia with only 121 training samples achieves F1 0.579, while SB with 3,889 samples achieves F1 0.992. Future improvements could employ synthetic oversampling (SMOTE for time series), few-shot learning techniques, or curriculum learning strategies that progressively focus training on difficult minority classes.

The comparison with related work provides additional context for the achieved results. Hannun et al. [2] reported sensitivity of 0.79 and specificity of 0.97 for arrhythmia detection on single-lead ambulatory ECG using a 34-layer CNN. The present work achieves comparable specificity (0.99 average) with higher macro sensitivity (0.85) on 8-class multi-label classification from standard 12-lead recordings. Ribeiro et al. [5] achieved 80% F1-score on 6 ECG abnormalities using a ResNet architecture; our SE-ResNet18 with metadata achieves 85.1% macro F1 on 8 rhythm classes, suggesting that the squeeze-and-excitation mechanism and clinical metadata provide meaningful improvements. The universal model's ability to maintain 91.6% accuracy on a single lead has practical implications for wearable ECG devices (smartwatches, portable monitors) that typically provide only lead I or lead II, enabling clinically useful rhythm screening outside hospital settings.

Several limitations should be noted. The dataset originates from a single hospital (Shaoxing People's Hospital), which may limit generalizability to other populations and recording equipment. The 8-class rhythm taxonomy excludes many clinically important conditions (myocardial infarction, bundle branch blocks, ventricular tachycardia). The NeuroKit2 feature extraction demonstrated high failure rates (10–50% fallback values), suggesting that production deployment should rely on the signal-only or signal+metadata configurations. The universal model was trained exclusively on Chapman–Shaoxing data; cross-dataset validation on CPSC-2018, PTB-XL, or MIT-BIH databases would strengthen the generalization claims.

To facilitate practical deployment, a REST API inference service has been developed using the FastAPI framework. The service accepts raw ECG signal data in JSON format, performs preprocessing (per-channel normalization using precomputed training statistics), runs inference through the trained model, and returns the predicted rhythm class along with confidence scores for all eight classes. The API supports both the standard 12-lead SE-ResNet18 model and the universal lead-agnostic model, automatically detecting the number of input leads and routing to the appropriate inference pathway. Response latency on a single NVIDIA RTX 3060 GPU averages 47 milliseconds per recording (including preprocessing), enabling real-time classification in clinical monitoring workflows.

An exploratory experiment was conducted to evaluate the capability of multimodal large language models for ECG rhythm classification [12]. GPT-4o Vision was presented with rendered 12-lead ECG images in standard clinical layout and prompted to classify the rhythm into one of eight categories. On a random subset of 100 test recordings, GPT-4o achieved 42% accuracy compared to 94.2% for the specialized SE-ResNet18 model. While the LLM demonstrated rudimentary medical knowledge by correctly identifying sinus bradycardia and tachycardia at above-chance rates, it failed almost entirely on minority classes such as atrial flutter and atrial tachycardia. This

comparison underscores the continued necessity of purpose-built deep learning architectures for safety-critical diagnostic tasks.

Conclusion. A comprehensive deep learning framework for multi-class ECG rhythm classification has been developed and evaluated across seven experimental configurations on the Chapman–Shaoxing 12-lead ECG database. The principal results are:

1. SE-ResNet18 with 14 clinical metadata features achieves the best classification performance: accuracy 94.2%, macro AUC 0.956, macro F1 0.851, Cohen's kappa 0.926 across 8 rhythm classes, outperforming signal-only, NeuroKit2, and demographic-only variants.

2. Systematic ablation reveals that clinical metadata provides the optimal accuracy–complexity trade-off, while NeuroKit2 morphological features with high fallback rates degrade performance despite higher AUC.

3. UniversalECGNet with shared backbone, lead embeddings, attention pooling, and lead dropout enables a single 9.18M-parameter model to operate across 1–12 lead configurations: 93.6% accuracy on 12-lead, 93.8% on 3-lead, 91.6% on single-lead, with negligible degradation from 12 to 6 leads.

4. Clinically interpretable error patterns – AF/AFIB and SA/SR confusion – align with known morphological similarities between these arrhythmia pairs.

Future work includes extending classification to the full diagnostic category set, applying transformer architectures for long-range temporal dependencies, implementing federated learning for privacy-preserving multi-hospital training, investigating few-shot learning for rare arrhythmias, and conducting clinical validation in real-time patient monitoring scenarios. The FastAPI-based inference service developed alongside the models enables straightforward integration into clinical decision support systems.

Reference list

[1] Makowski D., Pham T., Lau Z., Brammer J., Lespinasse F., Pham H., Schölzel C., Chen S. NeuroKit2: A Python toolbox for neurophysiological signal processing. Behavior Research Methods. 2021; 53(4):1689–1696. DOI: 10.3758/s13428-020-01516-y.

[2] Hannun A., Rajpurkar P., Haghpanahi M., Tison G., Bourn C., Turakhia M., Ng A. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine. 2019; 25(1):65–69. DOI: 10.1038/s41591-018-0268-3.

[3] World Health Organization. Cardiovascular diseases (CVDs): fact sheet. WHO; 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[4] LeCun Y., Bengio Y., Hinton G. Deep learning. Nature. 2015; 521(7553):436–444. DOI: 10.1038/nature14539.

[5] Ribeiro A., Stokes M., Meira S., Kroenke T., Paixao A., Nunes M., Ribeiro A. Automatic diagnosis of the 12-lead ECG using a deep neural network. Nature Communications. 2020; 11:1760. DOI: 10.1038/s41467-020-15432-4.

[6] Alday E., Gu A., Shah A., Robber C., Sharma A., Nemati S., Clifford G. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement. 2020; 41(12):124003. DOI: 10.1088/1361-6579/abc960.

[7] Hu J., Shen L., Sun G. Squeeze-and-Excitation Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2018:7132–7141. DOI: 10.1109/CVPR.2018.00745.

[8] Zheng J., Zhang J.,Danioko S., Yao H.,Guo H., Rakovski C.A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific Data.2020;7:48.DOI:10.1038/s41597-020-0386-x.

[9] Goldberger A., Amaral L., Glass L., Hausdorff J., Ivanov P., Mark R., Mietus J., Moody G., Peng C., Stanley H. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. Circulation. 2000; 101(23):e215–e220. DOI: 10.1161/01.CIR.101.23.e215.

[10] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition. Proceedings of the IEEE CVPR. 2016:770–778. DOI: 10.1109/CVPR.2016.90.

[11] Loshchilov I., Hutter F. Decoupled weight decay regularization. Proceedings of the 7th International Conference on Learning Representations (ICLR). 2019.

[12] Tohnson K., Torres Soto J., Glicksberg B., Shameer K., Miotto R., Ali M., Ashley E., Dudley J. Artificial intelligence in cardiology. Journal of the American College of Cardiology. 2018; 71(23):2668–2679. DOI: 10.1016/j.jacc.2018.03.521.

Author's contribution

Nikita Kisialeu – problem formulation, SE-ResNet18 and UniversalECGNet architecture design, implementation of all seven experimental configurations, data preprocessing and augmentation pipeline, NeuroKit2 feature extraction, experimental evaluation and ablation study, per-class error analysis, FastAPI inference service development, manuscript preparation.

Alexander Nedzved – scientific supervision, problem formulation, methodology guidance, result validation, manuscript review.

АВТОМАТИЧЕСКАЯ КЛАССИФИКАЦИЯ СЕРДЕЧНОГО РИТМА ПО 12-КАНАЛЬНОЙ ЭКГ С ПОМОЩЬЮ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ С МЕХАНИЗМОМ ВНИМАНИЯ

Н.Л. Киселев
Студент, факультет прикладной математики и информатики, Белорусский государственный университет, Республика Беларусь

А.М. Недзьведь
Заведующий кафедрой, доцент, доктор технических наук, Белорусский государственный университет, Республика Беларусь

Аннотация. Представлена комплексная система глубокого обучения для многоклассовой классификации ритма ЭКГ по 12-канальным записям базы Chapman–Shaoxing (10 615 записей, 8 классов). Предложена архитектура 1D SE-ResNet18 с интеграцией клинических метаданных. Проведено 7 экспериментов. Лучший результат: точность 94,2%, AUC 0,956, macro F1 0,851. Разработана универсальная модель UniversalECGNet для 1–12 отведений без переобучения.

Ключевые слова: классификация ЭКГ, глубокое обучение, SE-ResNet18, сердечная аритмия, 12-канальная ЭКГ, механизм сжатия-возбуждения, клинические метаданные, универсальная модель, механизм внимания, аугментация отведений.

УДК 004.89:37.02

ЭКСПЕРТНЫЕ И ГИБРИДНЫЕ МОДЕЛИ ПРИНЯТИЯ РЕШЕНИЙ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ КОНТРОЛЯ И ОЦЕНКИ ЗНАНИЙ



С.А. Мигалевич
Магистр технических наук, начальник центра информатизации и инновационных разработок
migalevich@bsuir.by

С.А. Мигалевич
Окончил Белорусский государственный университет информатики и радиоэлектроники.

Аннотация: В статье рассматриваются экспертные и гибридные модели принятия решений, применяемые в интеллектуальных системах контроля и оценки знаний. Анализируются архитектуры экспертных систем, включающие базу знаний, механизм логического вывода и интерфейс пользователя. Описываются методы представления знаний: продукционные правила, семантические сети, фреймы и нечёткая логика. Особое внимание уделено гибридным интеллектуальным системам, объединяющим экспертные правила с нейросетевыми и байесовскими подходами для повышения точности диагностики знаний студентов и адаптации учебного контента.

Ключевые слова: экспертные системы, гибридные интеллектуальные модели, база знаний, логический вывод, нечёткая логика, контроль знаний, адаптивное тестирование, нейро-символьные системы.

Введение. В условиях цифровизации образования актуальность приобретают интеллектуальные системы контроля и оценки знаний, способные автоматизировать процесс тестирования, диагностировать пробелы в знаниях студентов и рекомендовать персонализированные учебные траектории [1]. Классические подходы, основанные на жёстко запрограммированных алгоритмах, часто не учитывают неопределённость и вариативность ответов студентов, что снижает точность оценки и адаптивность системы [2].

Экспертные системы (ЭС) представляют собой компьютерные программы, имитирующие рассуждения эксперта в узкой предметной области [3]. Они оперируют базой знаний, содержащей правила и факты, и используют механизм логического вывода для принятия решений. В области образования экспертные системы применяются для диагностики уровня знаний, интерпретации результатов тестирования и формирования рекомендаций по устранению выявленных пробелов [4][5].

Однако чисто символьные экспертные системы имеют ограничения: сложность формализации экспертного знания, негибкость в условиях неполной информации и высокая трудоёмкость разработки [6]. Для преодоления этих ограничений разрабатываются гибридные интеллектуальные системы, сочетающие экспертные правила с методами машинного обучения – нейронными сетями, байесовскими сетями доверия, нечёткой логикой и генетическими алгоритмами [7][8].

Целью данной работы является анализ архитектур экспертных и гибридных моделей принятия решений, применяемых в интеллектуальных системах контроля и оценки знаний, а также обзор их преимуществ, недостатков и перспектив интеграции.

Экспертные системы: структура и методы представления знаний. Экспертная система – это компьютерная программа, моделирующая действия эксперта при решении задач в определённой предметной области [9]. Типичная архитектура экспертной системы включает следующие компоненты [3][10]:

- База знаний (БЗ) – хранилище фактов и правил, описывающих предметную область;
- Механизм логического вывода (решатель) – модуль, выполняющий рассуждения на основе правил и фактов для получения заключений;
- База данных (рабочая память) – временное хранилище фактов о текущей задаче;
- Интерфейс пользователя – обеспечивает взаимодействие между пользователем и системой;
- Модуль объяснения – предоставляет пользователю обоснование полученных выводов;
- Модуль приобретения знаний – инструмент для пополнения базы знаний экспертом или инженером по знаниям.

Для представления знаний в экспертных системах применяются различные модели [6][11]:

Продукционная модель основана на правилах вида «ЕСЛИ <условие> ТО <действие>». Продукционные правила обладают модульностью, естественностью представления и гибкостью в модификации базы знаний [12]. Пример правила в образовательной ЭС: «ЕСЛИ студент неправильно ответил на 3 и более вопросов по теме «Циклы» И правильно ответил на менее 50% вопросов по теме «Массивы», ТО рекомендовать изучить материал «Основы работы с массивами и циклами».

Семантические сети представляют знания в виде графа, вершины которого соответствуют объектам или понятиям, а дуги – отношениям между ними [13]. Семантические сети эффективны для представления иерархических и ассоциативных связей между концепциями учебной программы.

Фреймовая модель описывает объекты предметной области как структуры данных (фреймы), состоящие из слотов, заполняемых конкретными значениями [14]. Фреймы удобны для представления типовых ситуаций и наследования свойств.

Нечёткая логика (fuzzy logic) позволяет учитывать неопределённость и качественные оценки экспертов [15]. Вместо чёткого «истина» или «ложь» используются степени принадлежности элементов к нечётким множествам (например, «слабые знания», «средние знания», «хорошие знания»). Нечёткая логика широко применяется для оценки качественных аспектов знаний студентов, где границы между уровнями компетентности размыты [16][17].

Применение экспертных систем для контроля и оценки знаний. Экспертные системы в образовании решают задачи диагностики уровня знаний, интерпретации результатов тестирования и автоматической генерации рекомендаций по коррекции учебного процесса [4][18]. Громашева и Щербинина [4] описывают экспертную систему оценки знаний студентов по дисциплине «Системы реального времени», основанную на продукционных правилах. Система анализирует ответы студентов на тестовые вопросы, сопоставляет их с эталонными знаниями и выносит заключение о сформированности компетенций. Борисенко [5] разработал экспертную систему автоматизированного управления обучением в среде Stratum, включающую диагностический и планирующий модули. Диагностический модуль фиксирует ошибки учащегося и выявляет пробелы, а планирующий модуль предлагает индивидуализированную последовательность учебных материалов для их устранения.

Использование нечёткой логики в экспертных системах образования позволяет обрабатывать качественные оценки преподавателей и студентов [15][16][17]. Khan et al. [17] предложили экспертную систему с нечёткой логикой для оценки эффективности преподавателей, в которой лингвистические термины («очень хорошо», «хорошо», «удовлетворительно») преобразуются в числовые значения с помощью функций принадлежности. Аналогичный подход применим для оценки уровня знаний студентов: вместо бинарной оценки «сдал»/«не сдал» система может выставлять градуированную оценку на основе степени уверенности в ответах.

Гибридные интеллектуальные модели. Чисто символьные экспертные системы сталкиваются с проблемами при работе с большими объёмами данных, неполнотой информации и необходимостью обучения на исторических данных. Гибридные интеллектуальные системы объединяют сильные стороны символьных (экспертных) и субсимвольных (нейросетевых, статистических) подходов [7][8][20].

Михальков [8] выделяет два основных класса методов построения интеллектуальных систем поддержки принятия решений: эмпирические (основанные на данных: нейронные сети, генетические алгоритмы) и аналитические (основанные на знаниях: экспертные системы, нечёткие модели). Гибридная модель объединяет преимущества обоих классов: способность к обучению и адаптации от нейросетей, интерпретируемость и прозрачность от экспертных правил. Петров и соавт. [10] описывают моделирование систем оценивания знаний в рамках гибридной интеллектуальной обучающей среды (ГИОС). ГИОС сочетает нейросетевые модели для обучения на данных с экспертными правилами для интерпретации результатов. Авторы разработали структурные схемы для оценки знаний по математике и агентную модель педагогического процесса, в которой нейросети обучаются на последовательностях ответов учащихся, а правила экспертной системы формируют рекомендации по коррекции материала. Molenaar [21] вводит концепцию гибридного интеллекта (hybrid intelligence) в образовании как совместную работу человека (преподавателя) и ИИ-системы. Гибридный подход предполагает разделение уровней автоматизации: от полностью ручной оценки до полностью автоматизированной, с промежуточными уровнями, где решение принимается совместно. Это повышает доверие педагогов к системе и улучшает интерпретируемость выводов. Гибридные модели

представления знаний сочетают продукционные, семантические и фреймовые модели [11][22]. Например, фреймо-продукционная модель использует фреймы для представления сложных объектов (студент, тема, задача), а продукционные правила – для описания причинно-следственных связей между ними. Семантико-фреймовая модель задаёт вершины в виде фреймов, а связи между ними – семантической сетью. Байесовские сети доверия (Bayesian Belief Networks) также используются в гибридных системах для моделирования вероятностных зависимостей между знаниями студента и его ответами [23]. Комбинация байесовских сетей с экспертными правилами позволяет учитывать как статистические закономерности, так и экспертные предпочтения.

Преимущества и ограничения гибридных моделей. Гибридные модели принятия решений обладают рядом преимуществ по сравнению с чисто экспертными или чисто нейросетевыми подходами:

Повышенная точность предсказания. Комбинация статистических закономерностей, выявляемых нейросетями, и экспертных правил позволяет учитывать как исторические данные, так и специфику предметной области [10][27].

Интерпретируемость и объяснимость. Экспертные правила и символьные представления делают выводы системы прозрачными для преподавателей и студентов, что критически важно для образования [21][24].

Адаптивность. Гибридные системы способны обучаться на новых данных (нейросетевой компонент) и одновременно учитывать априорные знания экспертов (символьный компонент) [8].

Обработка неполноты и неопределённости. Использование нечёткой логики и байесовских сетей позволяет работать с качественными оценками и неполными данными [15][23].

Вместе с тем гибридные модели имеют ограничения:

Сложность разработки. Интеграция различных методов требует высокой квалификации разработчиков и тщательного согласования компонентов [6].

Трудоёмкость формализации знаний. Создание базы экспертных правил остаётся времяёмкой задачей, требующей участия экспертов предметной области [12].

Вычислительная сложность. Совместное применение нейронных сетей и символьных механизмов вывода может быть ресурсоёмким [20].

Проблема обобщения. Если экспертные правила слишком жёстко фиксированы, система может потерять способность адаптироваться к новым ситуациям [26].

Заключение. Экспертные системы и гибридные интеллектуальные модели представляют собой мощный инструмент для автоматизации контроля и оценки знаний в образовательных учреждениях. Классические экспертные системы, основанные на продукционных правилах, семантических сетях, фреймах и нечёткой логике, обеспечивают прозрачность и интерпретируемость выводов, но ограничены необходимостью ручной формализации знаний и низкой адаптивностью. Гибридные модели, объединяющие экспертные правила с нейросетевыми, байесовскими и эволюционными методами, позволяют преодолеть эти ограничения. Они сочетают способность к обучению и адаптации от машинного обучения с интерпретируемостью и использованием априорного знания от экспертных систем. Нейро-символьные подходы, такие как модели трассировки знаний с элементами символьного вывода, демонстрируют значительное улучшение точности и прозрачности диагностики знаний студентов. Перспективы развития гибридных интеллектуальных систем контроля знаний включают разработку методов автоматизированного извлечения экспертных правил из данных, создание интерактивных интерфейсов для объяснения выводов системы преподавателям и студентам, а также интеграцию с системами автоматической корректировки учебного контента на основе выявленных пробелов в знаниях. Важным направлением является обеспечение этической корректности и защиты персональных данных при использовании ИИ в образовании.

Список литературы

[1] Шлыкова, Е.Н. Гибридные модели оценки качества дошкольного образования: традиционные и цифровые подходы / Е.Н. Шлыкова // Известия Волгоградского государственного педагогического университета. - 2025. - № 7(200).

[2] Михайлов, Ю.И. Оценка качества гибридного обучения в высшей школе / Ю.И. Михайлов // 2023.

[3] Экспертная система // Википедия. - [Электронный ресурс]. - Режим доступа: https://ru.wikipedia.org/wiki/Экспертная_система.

[4] Громашева, О.С. Экспертная система оценки знаний студентов по дисциплине «Системы реального времени» / О.С. Громашева, И.А. Щербинина // Открытое образование. - 2015. - № 6.

[5] Борисенко. Реализация экспертной системы автоматизированного управления обучением / Борисенко // Выпускная квалификационная работа.

[6] Плащевая, Е.В. Экспертные системы на основе системы искусственного интеллекта / Е.В. Плащевая, Н.В. Нигей // АмурГМА.

[7] Misnik.by. Экспертные системы. - [Электронный ресурс]. - Режим доступа: https://misnik.by/exp_sys/es.pdf.

[8] Михальков, М.Д. Модели и методы построения интеллектуальных систем поддержки принятия решений / М.Д. Михальков // БГУИР. - 2025.

[9] Паршин, И.И. Проектирование интеллектуальной системы отбора экспертов / И.И. Паршин и др. // БелГУ.

[10] Петров, А.А. Моделирование систем оценивания знаний в рамках гибридной интеллектуальной обучающей среды / А.А. Петров, О.В. Дружинина, О.Н. Масина // Современные информационные технологии и ИТ-образование. - 2021. - Т. 17, № 1. - С. 179-189.

[11] Гибридные модели представления знаний // Системы поддержки принятия решений. - [Электронный ресурс]. - Режим доступа: https://studme.org/244093/informatika/gibridnye_modeli_predstavleniya_znaniy.

[12] Представления знаний в интеллектуальных системах // Habr. - 2018. - [Электронный ресурс]. - Режим доступа: https://habr.com/ru/articles/346236/.

[13] Экспертные системы / БрГУ. - [Электронный ресурс]. - Режим доступа: http://e.biblio.bru.by/bitstream/handle/1212121212/17903/Expertnye_sistemy.pdf.

[14] Интеллектуальные экспертные системы // Studfile. - [Электронный ресурс]. - Режим доступа: https://studfile.net/preview/7446878/page:39/.

[15] Hwang, G.-J. Knowledge acquisition for fuzzy expert systems / G.-J. Hwang // International Journal of Intelligent Systems. - 1995. - Vol. 10, No. 6.

[16] Alekseev, A. Using Fuzzy Logic in Knowledge Tests / A. Alekseev // CEUR Workshop Proceedings. - Vol. 1356.

[17] Khan, A.R. Application of Expert System with Fuzzy Logic in Teachers Performance Evaluation / A.R. Khan, H.U. Amin, Z.U. Rehman // International Journal of Advanced Computer Science and Applications. - 2011. - Vol. 2, No. 2.

[18] Aruna, P. Fuzzy Logic Expert System for Analyzing Student Performance / P. Aruna // IEEE. - 2023.

[19] Alekseev, A. Fuzzy logic in test-based knowledge control / A. Alekseev // CEUR. - 2014.

[20] Hybrid Intelligent Decision-Making Systems // Fiveable. - 2024. - [Электронный ресурс]. - Режим доступа: https://fiveable.me/neural-networks-and-fuzzy-systems/unit-17/hybrid-intelligent-decision-making-systems/.

[21] Molenaar, I. Towards hybrid human-AI learning technologies / I. Molenaar // European Journal of Education. - 2022.

[22] Ideal2025. Hybrid Intelligent Decision-Making: Integrating AI and MCDM. - [Электронный ресурс]. - Режим доступа: https://ideal2025.ujaen.es/.

[23] Bredeweg, B. Requirements and challenges for hybrid intelligence: A case-study in education / B. Bredeweg, M. Kragten // Frontiers in Artificial Intelligence. - 2022. - Vol. 5.

[24] Wang, X. Intelligent educational decision-making system driven by multimodal data fusion and knowledge graphs / X. Wang et al. // Nature Scientific Reports. - 2026.

[25] Piech, C. Deep Knowledge Tracing / C. Piech et al. // Stanford University. - 2015.

[26] Shen, S. A Survey of Knowledge Tracing: Models, Variants, and Applications / S. Shen, Q. Liu, E. Chen et al. // IEEE Transactions on Learning Technologies. - 2024. - Vol. 17. - PP. 1-20.

[27] Zhao, W. Improving Knowledge Tracing via a Heterogeneous Information Network / W. Zhao et al. // Expert Systems with Applications. - 2023.

[28] Song, Y. Improving exercise-level Knowledge Tracing via Knowledge Concept-based Memory Network / Y. Song et al. // Expert Systems with Applications. - 2025.

[29] Zhou, X. Deep learning based knowledge tracing in intelligent tutoring systems / X. Zhou et al. // Scientific Reports. - 2025. - Vol. 15.

Авторский вклад

Мигалевич Сергей Александрович – постановка задачи исследования, проведение обзора экспертных и гибридных моделей принятия решений в интеллектуальных системах контроля и оценки знаний, анализ методов представления знаний и механизмов логического вывода, подготовка и оформление статьи.

EXPERT AND HYBRID DECISION-MAKING MODELS IN INTELLIGENT KNOWLEDGE ASSESSMENT AND CONTROL SYSTEMS

S.A. Migalevich
Master of Technical Sciences, Head of the Center for Informatization and Innovative Developments

Annotation: The article examines expert and hybrid decision-making models applied in intelligent knowledge assessment and control systems. The architectures of expert systems, including knowledge bases, inference engines, and user interfaces, are analyzed. Methods of knowledge representation are described: production rules, semantic networks, frames, and fuzzy logic. Particular attention is paid to hybrid intelligent systems that combine expert rules with neural network and Bayesian approaches to improve the accuracy of student knowledge diagnosis and adaptive content customization.

Keywords: expert systems, hybrid intelligent models, knowledge base, inference engine, fuzzy logic, knowledge assessment, adaptive testing, neuro-symbolic systems.

УДК 336.71:338.22

ИЗМЕРЕНИЕ СИСТЕМНОГО ФИНАНСОВОГО СТРЕССА И АНАЛИЗ ЕГО СВЯЗИ С РАЗРЫВОМ ВЫПУСКА И ИНФЛЯЦИЕЙ: VAR-ПОДХОД НА ДАННЫХ БЕЛАРУСИ



Н.Л. Мирончик
Начальник Управления исследований и стратегического развития Национального банка Республики Беларусь, кандидат экономических наук
nmironchik@gmail.com

Н.Л. Мирончик
Окончила Белорусский государственный институт народного хозяйства им.В.В.Куйбышева. Область научных интересов связана с координацией денежно-кредитной и макропруденциальной политики, экономическим ростом и развитием.

Аннотация. Статья посвящена представлению композитного индикатора системного финансового стресса (КИСС) для Республики Беларусь и оценке его макроэкономической значимости на основе VAR-модели. Индикатор КИСС, агрегирующий девять рыночных и банковских индикаторов ключевых сегментов финансовой системы и рынка недвижимости, используется в VAR-модели с квартальными данными за 2015–2025 гг., включающей также разрыв выпуска и инфляцию, а также учитывающей крупные стрессовые эпизоды фиктивными переменными. Показано, что рост КИСС снижает выпуск и ускоряет инфляцию, а

положительный разрыв выпуска и инфляционные шоки повышают финансовую нестабильность в Республике Беларусь.

Ключевые слова: системный финансовый стресс; композитный индикатор; координация денежно-кредитной и макропруденциальной политики; векторная авторегрессия.

Введение. За последние два десятилетия в академической литературе и практике центральных банков заметно усилился интерес к измерению системного финансового стресса как особого состояния финансовой системы, характеризующегося повышенной уязвимостью к шокам и высокой вероятностью реализации системных событий. Если простое повышение волатильности или локальные сбои на отдельных рынках не всегда ведут к системным последствиям, то эпизоды, при которых стресс охватывает множество рынков и институтов, обычно сопровождаются значительными потерями для реальной экономики и требуют применения нетрадиционных мер экономической политики. В этих условиях непрерывные индикаторы финансового стресса стали важным инструментом мониторинга хрупкости финансовой системы, поддержки макрофинансового анализа и своевременного принятия решений [1].

Ряд центральных банков и международных организаций разработали собственные индексы финансового стресса и системной нестабильности, такие как Financial Stress Index (FSI) Банка Канады и Composite Indicator of Systemic Stress (CISS) Европейского центрального банка [2–4]. Как правило, эти индексы агрегируют информацию из нескольких сегментов финансового рынка – денежного и межбанковского, долгового, валютного, фондового – и дополняются банковскими показателями. Эмпирические исследования на базе VAR-моделей и близких подходов показывают, что финансовый стресс оказывает существенное влияние на реальный выпуск и инфляцию, а макроэкономические условия и денежно-кредитная политика, в свою очередь, формируют динамику финансового стресса [4–6].

Для Беларуси до последнего времени отсутствовал общепринятый непрерывный индикатор системного финансового стресса, сопоставимый с инструментами, применяемыми в развитых и ряде формирующихся экономик, несмотря на повторяющиеся эпизоды валютно-финансовой напряженности (2008–2009, 2011, 2015, 2020, 2022 гг.). В этой связи использование такого индикатора представляется целесообразной с прикладной точки зрения – для количественной оценки уровней финансового стресса, анализа его динамики и последующего использования в макрофинансовых моделях для обоснования мер экономической политики [4].

Цель настоящей работы – представить результаты разработки композитного индикатора системного финансового стресса для Республики Беларусь и показать его макроэкономическую значимость на основе обновленной VAR-оценки. Во-первых, суммируются ключевые элементы адаптации методологии, предложенной Д.Холло и др. [3], к структуре белорусской финансовой системы. Во-вторых, этот индикатор используется в VAR-модели для анализа динамических взаимосвязей между системным финансовым стрессом, разрывом выпуска и инфляцией в 2015–2025 гг. с учетом основных стресс-эпизодов. Основной вклад статьи заключается в (i) демонстрации специфичного для Беларуси композитного индикатора системного стресса, согласованного с международной практикой и предлагаемого для мониторинга финансовой стабильности, и (ii) эмпирической проверке устойчивости ранее полученных результатов по каналам «финансовая нестабильность – выпуск» и «финансовая нестабильность – инфляция» на расширенной выборке, включающей данные за 2025 г.

Статья организована следующим образом. В первой части рассматриваются теоретические и методологические подходы к измерению системного финансового стресса и его макроэкономических эффектов. Во второй части описывается построение композитного индикатора системного финансового стресса для Беларуси. В третьей части представлены спецификация VAR-модели, данные и основные результаты эмпирического

анализа. В четвертой части обсуждаются выводы для денежно-кредитной и макропруденциальной политики и намечаются направления дальнейших исследований.

1. Системный финансовый стресс: понятия и подходы к измерению. Следуя работам Н.Мирончик [4; 6], системный финансовый стресс – это состояние финансовой системы, отражающее агрегированный уровень напряженности в ее основных сегментах, обусловленное воздействием неблагоприятных факторов, проявляющееся в материализации накопленного системного риска для финансовой стабильности и порождающее необходимость адаптации финансовой системы к этому воздействию. В зависимости от уровня напряженности системный финансовый стресс способен оказывать негативное влияние на социально-экономическое развитие и(или) усиливать отрицательное воздействие неблагоприятных факторов, а его чрезвычайно высокий уровень соответствует состоянию системного финансового кризиса – глубокого расстройства функционирования основных сегментов финансовой системы, длительно и крайне негативно влияющего на развитие экономики.

Финансовый стресс проявляется через рост неопределенности, увеличение асимметрии информации, риск-премий, ухудшение рыночной ликвидности, существенные ценовые коррекции финансовых активов и сжатие кредитного предложения, при этом его уровень не наблюдается напрямую и должен оцениваться на основе совокупности индикаторов [7; 4].

В литературе можно выделить три ключевых класса методик построения индексов финансового стресса. Первый – простое агрегирование нормализованных показателей с равными или экспертно заданными весами, как в раннем FSI Банка Канады [8]. Второй – статистические методы (метод главных компонент, динамические факторные модели), позволяющие извлечь общую компоненту из множества финансовых переменных [9]. Третий подход, реализованный Д.Холло и др. в индикаторе CISS для ЕЦБ, основан на принципах портфельной теории и предполагает учет взаимных корреляций между составными индикаторами, так что периоды высокой синхронности стресса на разных сегментах рынка отражаются более высокими значениями композитного индекса, что критично для идентификации именно системного, а не локального стресса [3; 10].

Наряду с измерением, активно развиваются исследования, связывающие индексы финансового стресса с макроэкономическими переменными в рамках VAR-моделей и близких конструкций. Для развитых и формирующихся экономик в целом подтверждается, что положительный шок финансового стресса ведет к снижению выпуска и может ослаблять или усиливать инфляционное давление в зависимости от структуры шока, режима политики и характеристик финансовой системы [3; 9; 11–14]. В то же время положительные разрывы выпуска и периоды быстрого роста кредита часто связаны с нарастанием финансовых уязвимостей и повышением вероятности кризисов, что подчеркивает проциклическое взаимодействие реального и финансового циклов [15]. Высокая инфляция и дестабилизация инфляционных ожиданий способны подрывать финансовую стабильность, ухудшая реальное состояние балансов домохозяйств и финансовых институтов и повышая вероятность кризиса, тем самым усиливая финансовый стресс [16].

В данной статье используется композитный индикатор системного стресса для Беларуси, построенный на основе принципов портфельной теории по методологии близкой с CISS ЕЦБ [3]. Он включен в VAR-модель вместе с разрывом выпуска и инфляцией для анализа динамического взаимодействия финансового стресса, реальной активности и изменения цен.

2. Композитный индикатор системного финансового стресса для Беларуси. Композитный индикатор системного финансового стресса для Беларуси (КИСС) отражает напряженность в основных сегментах национальной финансовой системы и на рынке недвижимости. Выделяются следующие сегменты: валютный рынок, рынок МБК,

кредитный и депозитный рынки, банковский сектор и рынок недвижимости. Для каждого сегмента отбирается от одного до трех репрезентативных показателей, характеризующих цену, объем и риск; всего девять индикаторов [4].

В их число, в частности, входят процентные спрэды, показатели волатильности обменного курса и ставки МБК, индикаторы прибыльности банков и цен на рынке недвижимости. Выбор индикаторов определяется доступностью высокочастотных данных и ориентирован на сочетание рыночной информации с репрезентативным отражением материализации рисков финансовой нестабильности в соответствии с международной практикой построения индексов финансового стресса. Индикаторы собираются с дневной, недельной или месячной периодичностью и агрегируются до месячной частоты. Временной охват – 2007–2025 гг., что позволяет учесть докризисный период, глобальный финансовый кризис, валютно-финансовые эпизоды 2011 и 2015 гг., шок COVID-19 и стрессовые периоды 2020 и 2022 гг.

Поскольку индикаторы измеряются в различных единицах и диапазонах, перед агрегированием они нормализуются. В анализе используется процедура: min–max-преобразование, где каждый показатель xt преобразуется в переменную zt, ограниченную интервалом [0;1], путем вычитания минимума и деления на размах:

$$zt=xt-xminxmax-xmin, \quad (1)$$

где xmin и xmax– минимальное и максимальное значения исходного индикатора в выборке.

Нормализованные субиндикаторы агрегируются в КИСС с учетом значимости их взаимосвязей между собой. Это достигается за счет использования скользящих весов, основанных на парных коэффициентах корреляции субидикаторов: вес каждого субиндикатора пропорционален сумме его корреляций с другими, которые рассчитываются за заданное окно наблюдений. В результате субиндикаторы, которые более тесно связанные с остальными, получают больший вес в КИСС, усиливая вклад эпизодов синхронного стресса в значение композитного индекса [4].

Формально КИСС в момент времени *t* можно представить как:

$$CISSt=wtstCtwtst', \quad (2)$$

где wt=(w1,t,...,w9,t) – вектор скользящих весов субиндикаторов; st=(s1,t,...,s9,t) – вектор субиндикаторов; wtst – поэлементное умножение вектора весов субиндикаторов и вектора значений субиндикаторов в момент времени t; Ct – матрица изменяющихся во времени коэффициентов взаимной корреляции ij,t между *i*-м и *j*-м субиндикаторами [4].

На веса накладываются дополнительные ограничения (неотрицательность, сумма равна единице), обеспечивающие значения индекса в интервале [0;1]. Таким образом, индикатор отражает одновременно уровень и системный характер финансового стресса.

Динамика построенного композитного индикатора системного финансового стресса для Беларуси представлена на рисунке 1. КИСС демонстрирует выраженные пики в периоды известных эпизодов финансовой напряженности: глобальный финансовый кризис 2008–2009 гг., внутренний валютный кризис 2011 г., эпизод 2015 г., связанный с экономическим кризисом в России 2014 г., шок COVID-19 в 2020 г., существенный отток депозитов из банков, обусловленный усилением обеспокоенности вкладчиков устойчивостью банковской системы в середине 2020 г. и стрессовый период 2022 г. после начала специальной военной операции России в Украине [4]. Эти пики совпадают с эпизодами применения стабилизационных мер, что служит качественной верификацией индикатора.

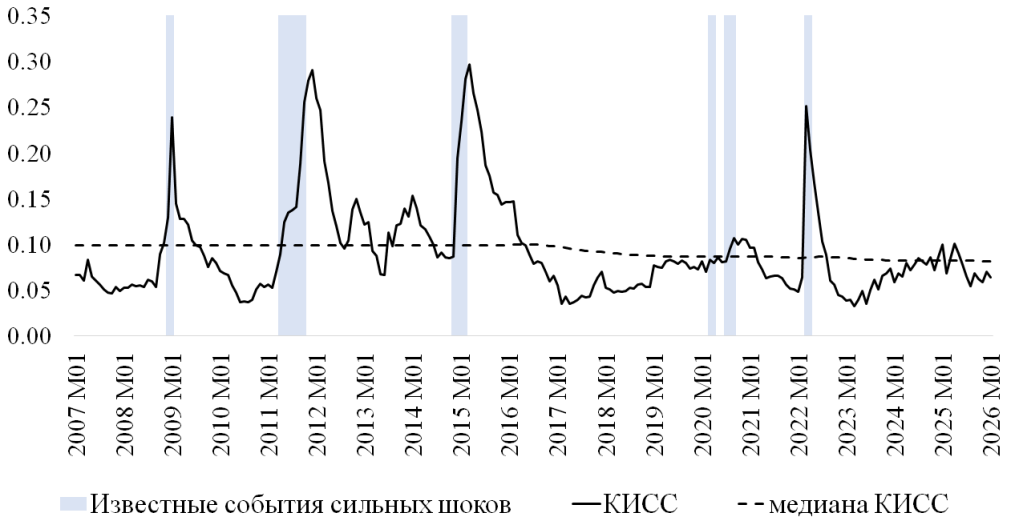


Рисунок 1. Композитный индикатор системного финансового стресса в Республике Беларусь, 2007-2025 гг.

3. VAR-модель: спецификация, данные и результаты. В качестве эмпирического инструмента используется векторная авторегрессионная модель с экзогенными переменными (VAR-X), включающая три эндогенные переменные: композитный индикатор системного финансового стресса, разрыв выпуска и темп инфляции. Разрыв выпуска определяется как процентное отклонение реального ВВП от равновесного уровня и характеризует положение экономики в бизнес-цикле; инфляция измеряется годовым темпом прироста индекса потребительских цен. Для учета реальных финансовых условий функционирования экономики в модель в качестве экзогенного фактора вводится переменная реальной процентной ставки, рассчитанная как разность между номинальной ставкой по всем кредитам банков экономике в белорусских рублях и инфляцией. Модель оценивается на квартальных данных за период с I квартала 2015 г. по IV квартал 2025 г., что соответствует этапу с режимом монетарного таргетирования в рамках денежно-кредитной политики.

Система VAR-X записывается в виде:

$$Yt=A0+A1Yt-1+Brt-1+GDt+t, \quad (3)$$

где Yt – вектор эндогенных переменных, A0 – вектор констант, A1 – матрица коэффициентов при лагах, rt – реальная процентная ставка как экзогенный регрессор, B – вектор коэффициентов при rt, Dt – вектор фиктивных переменных, отражающих ключевые кризисные кварталы (шоки 2020 и 2022 гг.), G – матрица коэффициентов при фиктивных переменных, t – вектор возмущений [5]. Тесты на единичные корни показывают, что переменные модели допустимо рассматривать как стационарные; диагностические проверки на автокорреляцию и гетероскедастичность подтверждают удовлетворительное качество модели, а значения коэффициента детерминации в уравнениях VAR остаются достаточно высокими. Лаг в модели выбран равным 1, чтобы при ограниченном числе наблюдений сохранить как можно больше степеней свободы при оценке. Увеличение количество лагов на основании критериев (AIC/SC/HQ) не дает существенного улучшения модели.

Импульсные отклики используются для отслеживания динамического влияния шоков в каждой переменной на остальные. Положительный шок КИСС (рост системного финансового стресса) приводит к статистически значимому снижению разрыва выпуска и росту инфляции. Отрицательное влияние на разрыв выпуска достигает максимума примерно через 4–8 кварталов и сохраняется на среднесрочном горизонте, что указывает на

депрессивный эффект системного финансового стресса на экономическую активность в Беларуси. Одновременно инфляция растет, особенно заметно в кратко- и среднесрочном периоде (рисунок 2).

Такая динамика разрыва выпуска и цен указывает на стагфляционные эпизоды в периоды повышенного финансового стресса, что в белорусских условиях связано с девальвационными шоками, усилением инфляционных ожиданий при высоком уровне издержек и низкой производительности предприятий, ограничивающих возможности адаптации к шоку без ускорения роста цен [5].

Положительный шок разрыва выпуска вызывает увеличение КИСС и ускорение инфляции. Это означает, что фазы перегрева экономики сопровождаются накоплением финансовых дисбалансов, усиливают инфляционное давление и повышают системную финансовую уязвимость (рисунок 3).

В краткосрочной перспективе инфляционные возмущения оказывают положительное влияние на значение КИСС. Это связано с тем, что внезапное ускорение темпов роста цен усиливает неопределенность в экономике, вызывает колебания финансовых потоков и ведет к расширению процентных спредов, что повышает уровень системного стресса. По мере восстановления ценовой стабильности данный эффект постепенно нивелируется, и на более продолжительном интервале умеренная инфляция может, напротив, способствовать снижению финансовой напряженности.

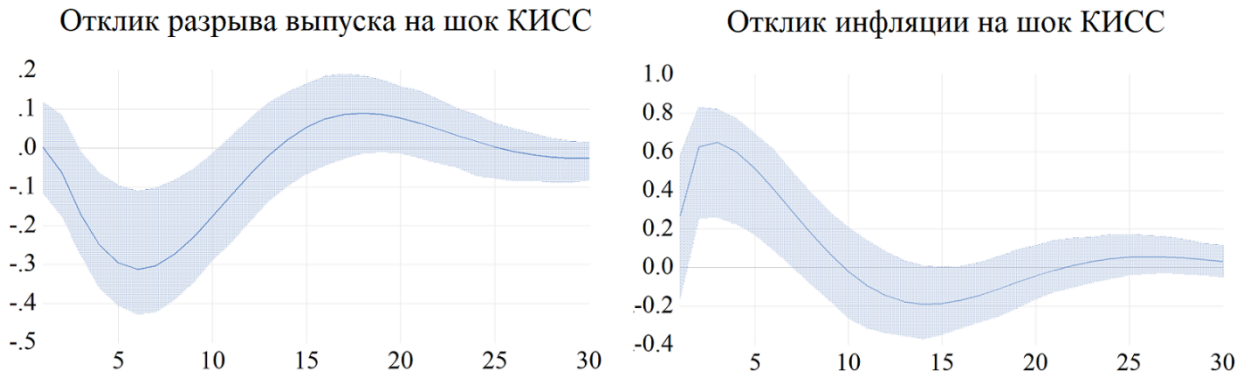


Рисунок 2. Реакция разрыва выпуска и инфляции на шок КИСС

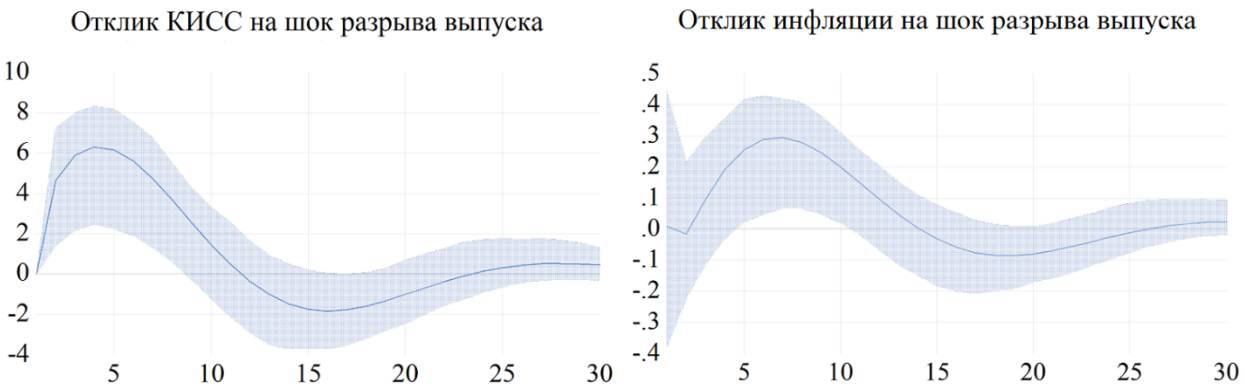


Рисунок 3. Реакция КИСС и инфляции на шок разрыва выпуска

Реакция разрыва выпуска на инфляционные шоки, как правило, имеет отрицательный характер, что отражает снижение реальных доходов и сдерживание экономической активности при росте цен (рисунок 4).

При анализе импульсных откликов следует учитывать, что траектория изменения эндогенных переменных формируется при заданных значениях экзогенной реальной процентной ставки. Поскольку rt-1 включена в модель как экзогенная переменная, влияние денежно-кредитной политики на системный стресс, выпуск и инфляцию частично отфильтровано и не смешивается с шоками, идентифицируемыми в VAR-системе. Это повышает устойчивость полученных оценок взаимосвязей между КИСС, разрывом выпуска и инфляцией к возможным изменениям режима денежно-кредитной политики, влияющих на реальные финансовые условия в экономике.

Декомпозиция дисперсии ошибок прогноза дополняет результаты импульсных откликов, позволяя оценить относительную значимость шоков в каждой переменной для вариации остальных. На среднесрочных горизонтах шоки КИСС объясняют существенную долю дисперсии ошибок прогноза разрыва выпуска и инфляции, что подтверждает важную роль системного финансового стресса как драйвера макроэкономических флуктуаций в Беларуси. Вклад шоков КИСС в дисперсию разрыва выпуска и инфляции увеличивается с горизонтом прогноза и достигает значимых значений уже через 1-2 года.

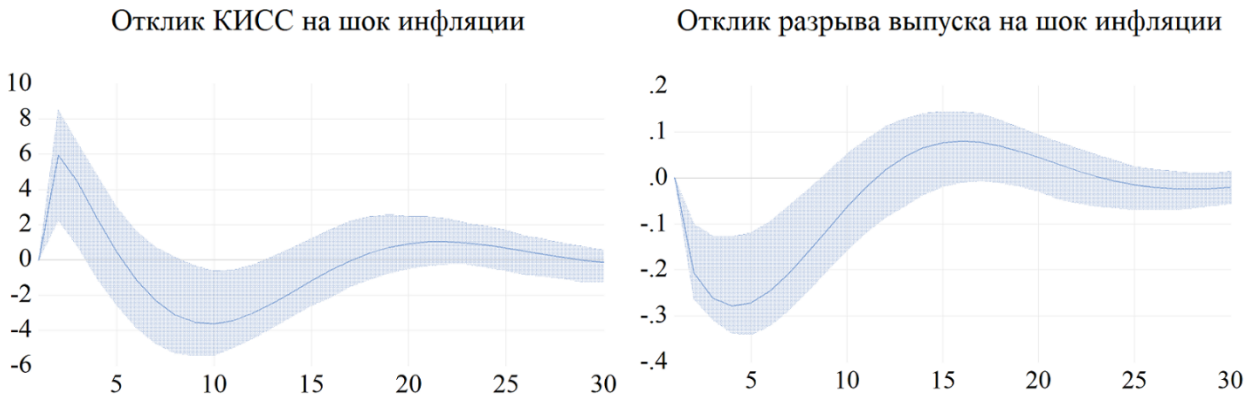


Рисунок 4. Реакция КИСС и разрыва выпуска на шок инфляции

В то же время значительная часть вариации самого КИСС обусловлена собственными шоками и шоками разрыва выпуска, тогда как вклад инфляционных шоков по дисперсии сравнительно невелик. Это, однако, не означает малозначимость инфляции: в краткосрочном периоде ее непредвиденное ускорение повышает финансовый стресс, в том числе через рост процентных ставок, тогда как на более длительном горизонте умеренная и ожидаемая инфляция, снижая реальную долговую нагрузку, может смягчать напряженность в финансовой системе. В совокупности результаты декомпозиции подтверждают выводы импульсного анализа и подчеркивают макрофинансовую значимость системного финансового стресса в Беларуси [4].

4. Обсуждение результатов и выводы для политики. Полученные результаты показывают, что системный финансовый стресс в Беларуси оказывает значимые и многомерные эффекты на экономическую активность и инфляцию. Рост композитного индикатора системного финансового стресса сопровождается снижением разрыва выпуска и одновременным ускорением инфляции, что ухудшает краткосрочный компромисс, стоящий перед денежно-кредитной политикой, и усложняет одновременное достижение целей ценовой и финансовой стабильности. В то же время положительные шоки разрыва выпуска и инфляции вносят вклад в усиление системного стресса, то есть в фазах подъема

и перегрева финансовые уязвимости нарастают, вместо того чтобы сглаживаться. Поскольку в последующем усиление стресса уже контрциклично влияет на выпуск, углубляя спад, взаимодействие реальной активности, инфляции и финансовой стабильности приобретает характер самоподдерживающихся макрофинансовых колебаний, а не простого про- или контрциклического эффекта, что соответствует логике финансового акселератора и современных моделей финансового цикла.

Включение реальной процентной ставки в качестве экзогенного регрессора в VAR-модель позволяет явно захватить влияние жесткости/мягкости реальных финансовых условий на макрофинансовую динамику. Оценки показывают, что ужесточение реальных финансовых условий (рост реальной ставки) ожидаемо оказывает дезинфляционное и сдерживающее воздействие на разрыв выпуска, что соответствует контрциклической реакции процентной политики на фазу бизнес-цикла. В то же время в условиях повышенной уязвимости финансовой системы повышение ставки может сопровождаться увеличением системного финансового стресса, то есть контрциклическая по отношению к разрыву выпуска денежно-кредитная политика может оказывать проциклический эффект на финансовую уязвимость. Это еще раз подчеркивает необходимость координации денежно-кредитной политики с макропруденциальными инструментами, смягчающими возможные побочные эффекты повышения ставки на финансовую стабильность.

Из этого вытекает ряд важных выводов для разработки и реализации денежно-кредитной и макропруденциальной политики в Беларуси. Во-первых, разработанный КИСС может служить операционным индикатором материализованного системного риска, и целесообразно его интегрировать в систему мониторинга и инструменты макрофинансового анализа, включая прогнозные модели и системы оценки рисков. Во-вторых, при формировании процентной политики следует принимать во внимание возможные последствия монетарных решений для состояния финансовой системы: отдельные меры, направленные на сдерживание инфляции, в определенных условиях способны усиливать проявления финансового стресса. В-третьих, макропруденциальные инструменты могут выполнять стабилизирующую функцию, ограничивая рост уязвимостей в периоды положительных разрывов выпуска и ускоренного кредитного расширения, тем самым снижая вероятность и амплитуду будущих эпизодов напряженности на финансовом рынке.

Дальнейшее развитие данного исследования может идти по нескольким направлениям. Во-первых, целесообразно расширить состав индикаторов, включаемых в КИСС, например за счет интеграции показателей, характеризующих состояние рынка ценных бумаг, по мере роста его глубины и ликвидности. Во-вторых, перспективным представляется использование КИСС в рамках сценарных стресс-тестов, учитывающих взаимодействие макроэкономических и финансовых шоков и позволяющих количественно оценить совокупные риски для ценовой и финансовой стабильности. В-третьих, применение нелинейных или структурных VAR-моделей, а также методов с переключением режимов может обеспечить более точный анализ асимметрии воздействий крупных стресс-эпизодов и повысить качество идентификации структурных возмущений.

Заключение. Разработка адаптированного к белорусской экономике композитного индикатора системного финансового стресса и его интеграция в VAR-модель, включающую разрыв выпуска и инфляцию, формируют методологическую основу для углубленного макрофинансового анализа и укрепления согласованности денежно-кредитной и макропруденциальной политики в Беларуси.

Список литературы

[1] Стефанович, Л. Композитный индикатор системного стресса: обзор теоретических и эмпирических подходов / Л. Стефанович, Н. Мирончик // Банкаўскі веснік. – 2025. – № 3 (740). – С. 3–12.

[2] Duprey, T. Canadian Financial Stress and Macroeconomic Conditions / T. Duprey // Bank of Canada Staff Discussion Paper. – 2020. – № 4. – 27 p.

[3] Hollo, D. CISS – a composite indicator of systemic stress in the financial system / D. Hollo, M. Kremer, M. Lo Duca // ECB Working Paper Series – 2012. – № 1426. – 49 p.

[4] Мирончик, Н. Композитный индикатор системного финансового стресса для Республики Беларусь / Н. Мирончик // Банкаўскі веснік. – 2025. – № 5 (742). – С. 3–24.

[5] Мирончик, Н. Анализ взаимосвязи между системным финансовым стрессом, разрывом выпуска и инфляцией с помощью векторной авторегрессии / Н. Мирончик // Банкаўскі веснік. – 2025. – № 11. – С. 3–22.

[6] Мирончик, Н. Л. Системный финансовый стресс / Н. Л. Мирончик // Экономика Беларуси: рост, инновации, безопасность : материалы Международной научно-практической конференции, Минск, 15 мая 2025 г. / Министерство образования Республики Беларусь, Белорусский государственный экономический университет ; [редакционная коллегия: А. А. Быков (ответственный редактор) и др.]. – Минск : БГУИР, 2025. – С. 343.

[7] Hakkio, C. S. Financial stress: what is it, how can it be measured, and why does it matter? / C. S. Hakkio, W. R. Keeton // Economic Review. – 2009. – Vol. 94, № 2. – P. 5–50.

[8] Illing, M. An Index of Financial Stress for Canada / M. Illing, Y. Liu // Bank of Canada Working Paper. – 2003. – № 14. – 52 p.

[9] Van Roye, B. Financial Stress and economic activity in Germany and the Euro Area / B. Van Roye // Kiel Working Paper. – 2011. – № 1743. – 35 p.

[10] Мирончик, Н. Концептуальная основа сопряженности монетарной и макропруденциальной политик / Н. Мирончик // Банковская система: устойчивость и перспективы развития: сборник науч. статей XV Междунар. науч.-практ. конф. по вопросам финансовой и банковской экономики, Пинск, 25 октября 2024 г. / УО «Полесский государственный университет»; редкол.: А В.И. Дунай [и др.] – Минск : ПолесГУ, 2024. – С. 81–85.

[11] Calvo, G.A. Monetary policy challenges in emerging markets: Sudden stop, liability dollarization, and lender of last resort / G.A. Calvo // NBER Working Paper. – 2006. – № 12788. – 31 p.

[12] Cardarelli, R. Financial stress and economic contractions / R. Cardarelli, S. Elekdag, S. Lall // Journal of Financial Stability. – 2011. – Vol. 7, № 2. – P. 78–97.

[13] Mittnik, S. The real consequences of financial stress / S. Mittnik, W. Semmler // Journal of Economic Dynamics & Control. – 2013. – № 37. – P. 1479– 499.

[14] Sahoo, J. Financial stress index, growth and price stability in India: Some recent evidence / J. Sahoo // Theoretical & Applied Economics. – 2020. – Vol. 27, № 1. – P. 105–124.

[15] Minsky, H.P. The Financial Instability Hypothesis / H.P. Minsky // The Jerome Levy Economics Institute Working Paper. – May 1992. – № 74. – 10 p.

[16] Albertazzi, U. The causal effect of inflation on financial stability, evidence from history / U. Albertazzi, J. Hooft, L.T. Steege // ECB Working Paper. – 2025. – № 3108. – 43 p.

Авторский вклад

Мирончик Наталья Леонидовна – постановка цели и задач исследования, формирование понятийного аппарата системного финансового стресса и системного финансового кризиса, разработка методологии построения составного индикатора системного финансового стресса и выбора его компонент, проведение эмпирических расчетов и интерпретация результатов, и формулировка выводов.

MEASURING SYSTEMIC FINANCIAL STRESS AND ANALYSING I TS LINK WITH OUTPUT GAP AND INFLATION: A VAR APPROACH FOR BELARUS

N.L. Mironchik

Head of the Research and Strategic Development Department,
National Bank of the Republic of Belarus, PhD in Economics

Abstract. The article presents a composite indicator of systemic financial stress (CISS) for the Republic of Belarus and evaluates its macroeconomic relevance using a VAR model. The CISS, which aggregates nine market- and bank-based indicators covering key segments of the financial system and the real estate market, is employed in a VAR model with quarterly data for 2015–2025 that also includes the output gap and inflation and controls for major stress episodes via dummy variables. The results show that an increase in the CISS reduces output and accelerates inflation, while a positive output gap and inflation shocks raise financial instability in the Republic of Belarus.

Keywords: systemic financial stress; composite indicator; coordination of monetary and macroprudential policy; vector autoregression.

УДК 004.89

ГИБРИДНАЯ МОДЕЛЬ ОПТИМИЗАЦИИ БИЗНЕС-РЕШЕНИЙ НА ОСНОВЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ



В.И. Борисенко
Ассистент кафедры информационных технологий СПбГУПТД
borisenko.vi.661@suitd.ru

В.И. Борисенко
Окончил Санкт-Петербургский государственный университет промышленных технологий и дизайна. Область научных интересов связана с разработкой методов и алгоритмов интеллектуального анализа данных.

Аннотация. Представлена концепция верифицируемой системы поддержки принятия решений на основе больших языковых моделей. Предложена гибридная архитектура, интегрирующая методы RAG и ReAct, модули предобработки, генерации, многоуровневой верификации и адаптации. Формализовано преобразование проектных переменных в показатели эффективности с использованием аналитических моделей, машинного обучения и имитационного моделирования. Подход обеспечивает достоверность, интерпретируемость выводов и применимость в корпоративной среде.

Ключевые слова: большие языковые модели, система поддержки принятия решений, RAG, ReAct, гибридная модель, неструктурированные данные

Введение. Современные организации функционируют в условиях высокой неопределённости внешней среды и непрерывно растущих объёмов информации, значительная часть которой представлена в неструктурированном виде. Тексты обращений клиентов, протоколы совещаний, сообщения в социальных медиа, журналы событий информационных систем содержат ценные сведения, необходимые для обоснованного принятия управленческих решений. Однако эффективное использование этих данных ограничено их разнородностью, субъективностью и сложностью формализации в рамках традиционных аналитических моделей.

Критическим фактором, сдерживающим применение методов машинного обучения в бизнес-аналитике, выступает дефицит репрезентативных обучающих выборок. Сбор и разметка достаточного объёма реальных неструктурированных данных сопряжены с рядом объективных трудностей. Конфиденциальность коммерческой информации не позволяет передавать данные за пределы организации. Редкие, но значимые события (отказы оборудования, мошеннические действия, кризисные ситуации) представлены в исторических записях недостаточно. Ручная аннотация текстов требует привлечения экспертов и характеризуется высокой трудоёмкостью. Указанные ограничения препятствуют построению точных прогнозных моделей и проведению всестороннего сценарного анализа.

Научная новизна настоящего исследования заключается в разработке метода оптимизации бизнес-решений, основанного на интеграции верифицированных синтетических неструктурированных данных, полученных от большой языковой модели (large language model, LLM), в контур поддержки принятия решений. Предлагаемый подход,

развивающий архитектуру [1], дополняет генеративные способности языковых моделей многоуровневой системой контроля, обеспечивающей семантическую согласованность, статистическую правдоподобность и процессную выполнимость создаваемых данных. Это позволяет расширить обучающие выборки и сценарные пространства без потери достоверности, повышая точность прогнозных моделей и обоснованность управленческих решений.

Целью работы является формализация, реализация метода верифицированной генерации синтетических неструктурированных данных на основе LLM для последующей оптимизации бизнес-показателей. Для достижения поставленной цели решаются задачи по анализу существующих подходов к синтезу данных, выявлению их ограничений, разработке архитектуры генерации с контролируемыми свойствами, построению модели влияния синтетических данных на интегральный показатель эффективности.

Структура гибридной модели

Для системного совершенствования бизнес-процессов организации необходимо формализовать процесс принятия решений как задачу оптимизации, в которой входные проектные переменные преобразуются в показатели эффективности. Предлагаемая гибридная модель объединяет аналитические зависимости, методы машинного обучения, обученные большие языковые модели и имитационное моделирование, что позволяет учитывать, как структурированные, так и неструктурированные данные (тексты документов, сообщения, экспертные заключения) и обеспечивать высокую точность предсказания. В основе модели лежит архитектура верифицируемой системы поддержки принятия решений, описанная в [1], которая обеспечивает многоуровневый контроль достоверности генерируемых гипотез и их соответствие бизнес-онтологии.

Гибридная модель описывает преобразование множества проектных переменных $X = (X_1, X_2, \ldots, X_N)$ в вектор показателей эффективности $Y = (Y_1, Y_2, \ldots, Y_K)$. Такое преобразование реализуется совокупностью функциональных модулей, каждый из которых отвечает за определённый этап обработки данных и формирования оценок. В основу структуры положена архитектура верифицируемой системы поддержки принятия решений, предложенная в работе [1], которая обеспечивает многоуровневый контроль достоверности результатов и их согласованность с онтологией предметной области.

Модуль аналитических моделей и моделей машинного обучения реализует вычисление показателей, для которых могут быть построены явные математические зависимости или обучены статистические модели на основе структурированных исторических данных. К числу таких методов относятся регрессионные модели, аппарат временных рядов, методы классификации и кластеризации. Выходные значения модуля используются как часть вектора Y либо как промежуточные переменные для других компонентов.

Модуль обработки неструктурированной информации на основе LLM выполняет извлечение знаний из текстовых данных, таких как документы, новостные сообщения, отчёты, переписка, и генерацию гипотез о влиянии проектных переменных на показатели эффективности. В его основе лежит когнитивный агент, построенный с применением подходов генерации, дополненной поиском (retrieval-augmented generation, RAG) и парадигмы с синергией размышления и действия (reasoning and acting, ReAct) [2, 3]. Первый из них обеспечивает привязку генерируемых утверждений к внешним источникам, второй отвечает за итеративное уточнение гипотез через чередование этапов рассуждения и запросов к данным. Результатом работы модуля являются не только предсказанные значения показателей, но и структурированное объяснение, содержащее цепочку проверенных рассуждений, ссылки на использованные источники и количественные оценки достоверности.

Модуль имитационного моделирования на основе цифрового двойника проигрывать различные сценарии изменения проектных переменных, оценивая результирующие

значения показателей Y в динамике. Данный модуль особенно важен при проверке логико-процессной выполнимости решений, поскольку он выявляет нарушения регламентов и выход метрик за допустимые границы.

Результаты, полученные из перечисленных выше модулей, поступают в модуль верификации, где проходят три уровня контроля. На первом уровне проверяется синтаксико-семантическая согласованность с онтологией предметной области где выявляются противоречия с формализованными бизнес-правилами и связями сущностей. Второй уровень посвящён фактологической обоснованности. Каждое ключевое утверждение сопоставляется с доверенными источниками из пула контекстов и внешними базами данных (регуляторные документы, рыночные индикаторы). Третий уровень оценивает логико-процессная выполнимость, реализуемый через запуск имитационной модели для оценки реализуемости сценария в рамках существующих процессов. По итогам проверки формируется интегральная оценка достоверности предложенного решения.

На последнем этапе модуль коррекции и адаптации аккумулирует обратную связь от лица, принимающего решения (ЛПР), и результаты верификации. На основе анализа отклонённых решений производится корректировка онтологии, добавление или уточнение правил, настройка порогов срабатывания проверок, обновление тегов доверия источников в пуле контекстов. Таким образом, модель обладает свойством адаптивности к изменениям внешней среды и специфике организации.

Взаимодействие перечисленных компонентов организовано по принципу сквозного контура, происходящего от ввода проектных переменных до выдачи верифицированного решения с полным объяснением. Каждый этап обработки сопровождается фиксацией промежуточных данных, что позволяет в модуле коррекции идентифицировать источник ошибок и целенаправленно обновлять соответствующие параметры. Предложенная структура обеспечивает не только высокую точность предсказаний за счёт интеграции разнородных методов, но и необходимый уровень интерпретируемости, что является критическим условием для применения в корпоративной практике.

Проектные переменные и их влияние на показатели. В рамках гибридной модели центральное место занимает формализация связи между управляемыми параметрами бизнес-решения и результирующими показателями эффективности. Проектные переменные $X = (X_1, X_2, \ldots, X_N)$ представляют собой совокупность факторов, находящихся в зоне ответственности ЛПР, и подлежащих целенаправленному изменению для достижения желаемых значений показателей. Классификация проектных переменных, их размерность и допустимые области значений определяются спецификой решаемой бизнес-задачи и уровнем детализации модели.

Типология проектных переменных

Проектные переменные могут быть разделены на три категории в зависимости от их природы и способа задания.

Непрерывные переменные. К данной категории относятся параметры, принимающие значения из непрерывного интервала. Примерами служат объёмы производства, размер инвестиций, цена единицы продукции, норматив запаса, длительность этапа производственного цикла. Непрерывные переменные обычно поддаются градиентной оптимизации и могут быть выражены в натуральных или стоимостных единицах.

Дискретные переменные. В эту группу входят параметры, принимающие конечное или счётное множество значений. К ним относятся выбор поставщика из утверждённого перечня, конфигурация программного обеспечения, тип логистической схемы, категория клиентского сегмента. Оптимизация по дискретным переменным требует применения комбинаторных методов или перебора.

Качественные (структурные) переменные. Данная категория охватывает параметры, не имеющие естественного числового представления и определяющие структуру или

конфигурацию бизнес-процесса. Примерами являются политика управления запасами, формат взаимодействия с клиентом, организационная структура подразделения. В гибридной модели качественные переменные кодируются с использованием онтологии предметной области и могут обрабатываться как дискретные альтернативы.

Для каждой переменной X_i задаётся область допустимых значений, определяемая технологическими, ресурсными или регламентными ограничениями. Совокупность этих областей формирует пространство проектных решений.

Отображение проектных переменных в показатели эффективности

Зависимость каждого показателя эффективности Y_i от проектных переменных представляется в виде:

$$Y_k = f_k(X, U),\ k = 1, 2, \ldots, K$$

где U – вектор внешних факторов, не контролируемых ЛПР (рыночная конъюнктура, действия конкурентов, изменения регуляторной среды).

Функции f_k в гибридной модели не являются единообразными. Они формируются на основе комбинации различных вычислительных подходов, выбор которых определяется структурой зависимости и доступностью данных.

Аналитические зависимости применяются в случаях, когда связь между переменными и показателями может быть выражена явной формулой, вытекающей из технологических или экономических закономерностей. Например, показатель объёма выручки может быть выражен как произведение цены на объём продаж, где последний в свою очередь зависит от цены через функцию спроса. Аналитические модели обладают преимуществом прозрачности и вычислительной эффективности, однако их применимость ограничена ситуациями, где известна точная функциональная форма.

Модели машинного обучения используются, когда зависимость имеет сложный, нелинейный характер, но имеется достаточный объём структурированных исторических данных для обучения. В рамках гибридной модели применяются регрессионные методы, деревья решений, градиентный бустинг, а также нейросетевые архитектуры. Обученные модели позволяют учитывать взаимодействия между переменными и нелинейные эффекты, которые трудно формализовать аналитически. При этом в модуле верификации осуществляется контроль за тем, чтобы предсказания модели не выходили за границы области применимости обучающей выборки.

Оценки на основе больших языковых моделей применяются для показателей, существенно зависящих от неструктурированной информации. LLM извлекает из текстовых источников семантические признаки, влияющие на прогнозируемые значения. Например, показатель потребительского спроса может корректироваться на основе анализа тональности отзывов и новостных сообщений о конкурентах. Важной особенностью является то, что LLM не заменяет аналитические или статистические модели, а дополняет их, обеспечивая учёт контекстной информации, не представленной в структурированных данных.

Результаты имитационного моделирования используются для показателей, зависящих от динамики процесса, наличия обратных связей и временных лагов. Цифровой двойник бизнес-процесса позволяет при фиксированном наборе проектных переменных XX получить траекторию изменения показателей во времени, что особенно важно для оценки устойчивости решений и выявления эффектов, проявляющихся только в динамике.

Учёт внешних факторов через неструктурированные данные

Вектор внешних факторов U в классических моделях часто либо игнорируется, либо вводится в виде упрощённых сценариев. Предлагаемая архитектура позволяет извлекать информацию о текущем состоянии внешней среды из неструктурированных потоков с

использованием LLM. Извлечённые факторы интегрируются в функции f_k следующим образом:

$$Y_k = f_k(X,\ LLM_{extract}(D_{ext}))$$

где $LLM_{extract}()$ – процедура извлечения количественных и качественных характеристик внешней среды, D_{ext} – совокупность внешних текстовых источников.

Полученные оценки проходят верификацию через сопоставление с несколькими независимыми источниками, что снижает риск ошибок, связанных с недостоверной информацией.

Пример формализации для задачи оптимизации маркетинговой стратегии

Для иллюстрации предложенного подхода рассмотрим задачу оптимизации маркетинговой стратегии. Пусть проектные переменные включают:

- X_1 – бюджет рекламной кампании (непрерывная, тыс. руб.);
- X_2 – набор каналов коммуникации (дискретная, варианты: «социальные сети», «контекстная реклама», «телевидение», «комбинированный»);
- X_3 – выбор определенной группы (качественная, кодируемая через онтологию целевых аудиторий).

Показатели эффективности:

- Y_1 – прирост выручки (тыс. руб.);
- Y_2 – рентабельность инвестиций;
- Y_3 – охват целевой аудитории (тыс. человек).

Функция f_1 для показателя Y_1 формируется комбинированно:

$$Y_1 = f_1^{ML}(X_1,\ X_2) + \Delta Y_1^{LLM}$$

где f_1^{ML} – модель машинного обучения, обученная на исторических данных о кампаниях; ΔY_1^{LLM} – корректировка, полученная на основе анализа текущих новостных потоков (например, информация о выходе нового конкурента на рынок, изменении потребительских предпочтений), извлекаемая LLM и верифицируемая через сопоставление с несколькими источниками.

Показатель Y_2 вычисляется аналитически:

$$Y_2 = \frac{Y_1 - C(X_1,\ X_2)}{X_1}$$

где $C(X_1,\ X_2)$ – функция затрат, учитывающая бюджет и специфику каналов.

Показатель Y_3 определяется имитационной моделью, учитывающей динамику распространения рекламных сообщений и поведенческие реакции аудитории, параметризованной через X_2 и X_3.

Роль формализации в оптимизационном контуре

Введение единообразного представления проектных переменных и функций отображения создаёт основу для постановки и решения задачи оптимизации бизнес-решений. Поскольку функции f_k в гибридной модели могут быть недифференцируемыми, содержать дискретные компоненты и включать вычислительно затратные имитационные блоки, для поиска оптимальных значений X^* применяются методы, не требующие вычисления градиентов. Примерами являются эволюционные алгоритмы, байесовская оптимизация, методы поиска с ограничениями. Каждое вычисление целевой функции

$Y = F(Y_1, \ldots, Y_K)$ при фиксированном X представляет собой последовательный запуск всех компонентов модели с фиксацией промежуточных результатов, что позволяет модулю верификации контролировать корректность на каждом этапе.

Предложенная формализация обеспечивает:

- однозначную интерпретацию проектных переменных как управляемых параметров;
- возможность комбинирования разнородных вычислительных методов в рамках единого оптимизационного цикла;
- учёт неструктурированных данных наравне со структурированными;
- прозрачность влияния каждой переменной на итоговые показатели через фиксируемые промежуточные зависимости.

Интегральный критерий эффективности

Вектор показателей $Y = (Y_1, Y_2, \ldots, Y_K)$ отражает различные, часто противоречивые аспекты эффективности бизнес-решения. Для целенаправленного поиска оптимальных проектных переменных требуется сведение многокритериальной оценки к скалярному значению, позволяющему сравнивать альтернативы и осуществлять направленный поиск. Интегральный критерий эффективности $Y = F(Y_1, Y_2, \ldots, Y_K)$ выполняет функцию агрегирования, при этом выбор функции F и способа её параметризации определяет структуру предпочтений ЛПР.

Линейная свёртка является наиболее распространённым способом агрегирования благодаря простоте интерпретации и вычислительной эффективности [4]:

$$Y = \sum_{k=1}^{K} w_k Y_k$$

где w_k – весовые коэффициенты, отражающие относительную важность показателей.

Недостатком линейной свёртки является предположение о полной компенсируемости. Снижение одного показателя может быть неограниченно скомпенсировано ростом другого, что не всегда соответствует реальным бизнес-приоритетам.

Учёт предпочтений ЛПР и настройка критерия

Весовые коэффициенты и параметры функций агрегирования не могут быть заданы априори в отрыве от конкретной бизнес-ситуации. В предлагаемой архитектуре настройка критерия осуществляется интерактивно с использованием модуля коррекции и адаптации. ЛПР предоставляет возможность ранжировать альтернативные решения, генерируемые системой, либо указывать на неприемлемые компромиссы. На основе этих данных решается обратная задача по подбору параметров F, при которых выбранные ЛПР альтернативы оказываются предпочтительнее отвергнутых. Для этого могут применяться методы обучения предпочтений, в том числе с использованием LLM для извлечения неявных предпочтений из текстовых пояснений ЛПР [5].

Интеграция с оптимизационным контуром

После определения интегрального критерия задача оптимизации проектных переменных принимает вид:

$$X^* = \arg\max Y(X) = \arg\max F(f_1(X, U), \ldots, f_k(X, U))$$

Особенностью гибридной модели является то, что вычисление $Y(X)$ требует последовательного выполнения всех компонентов гибридной модели с фиксацией промежуточных данных для верификации. Каждое вычисление сопровождается формированием объяснения, что позволяет ЛПР контролировать обоснованность получаемых оценок.

Верификация интегрального критерия

Интегральный критерий, как и другие компоненты модели, подлежит верификации. В модуле верификации проверяется:

- согласованность выбранной функции F с онтологией предметной области, например, допустимость линейной компенсации между показателями;
- устойчивость решений к вариациям весовых коэффициентов через анализ чувствительности, позволяющий выявить области параметров, в которых оптимальное решение меняется скачкообразно.

Результаты верификации критерия представляются ЛПР в виде интерактивных графиков (парето-фронты, диаграммы чувствительности), что повышает доверие к рекомендациям системы.

Заключение. Предложенная гибридная модель обеспечивает формализованное преобразование проектных переменных в показатели эффективности с интеграцией больших языковых моделей для обработки неструктурированных данных. Многоуровневая верификация, используя синтаксико-семантическая, фактологическая, процессная проверки, гарантирует достоверность результатов. Введение интегрального критерия с настраиваемой структурой предпочтений и механизмом адаптации на основе обратной связи позволяет решать задачу оптимизации бизнес-решений с требуемой интерпретируемостью. Предложенный подход создаёт основу для построения доверенных систем поддержки принятия решений в корпоративной среде.

Список литературы

[1] Борисенко, В. И. Генерации синтетических наборов данных с использованием больших языковых моделей для решения бизнес-задач / В. И. Борисенко, В. И. Пименов // ПРОМТЕХДИЗАЙН. Естественные и технические науки : Сборник статей всероссийской научной конференции молодых ученых с международным участием. Часть 3 (СПбГУПТД, 1-27 апреля 2025 г.). – 2025. – С. 40-42.

[2] Retrieval-Augmented Generation for Large Language Models: A Survey [Электронный ресурс] / Y. Gao [et al.]. – 2024. – Режим доступа: https://arxiv.org/abs/2312.10997v5

[3] ReAct: Synergizing Reasoning and Acting in Language Models [Электронный ресурс] / S. Yao [et al.]. – 2023. – Режим доступа: https://arxiv.org/abs/2210.03629v3

[4] Nasyuha, A. H. Comparison of WSM and Weight Product Methods With WSM-Score and Vector Approaches / A. H. Nasyuha [et al.] // Sinkron : Jurnal Dan Penelitian Teknik Informatika. – 2025. – №2(9). – PP. 948-956.

[5] Towards a unified view of preference learning for large language models: A survey [Электронный ресурс] / B. Gao [et al.]. – 2024. – Режим доступа: https://arxiv.org/abs/2409.02795v5

Авторский вклад

Борисенко Владислав Игоревич – постановка задачи исследования и определение направления совершенствования бизнес-процессов на основе обработки неструктурированных данных, анализ существующих подходов к обеспечению достоверности и интерпретируемости больших языковых моделей, выявление системных ограничений и пробелов в существующих решениях, разработка архитектуры гибридной модели, описание структуры интегрального критерия эффективности и методов агрегирования показателей.

HYBRID MODEL FOR OPTIMIZING BUSINESS DECISIONS BASED ON LARGE LANGUAGE MODELS

V. I. Borisenko
Assistant of the Department of Information Technology,
SPbSUITD

Abstract. This article presents a concept for a verifiable decision support system based on large language models. A hybrid architecture is proposed that integrates RAG and ReAct methods, along with modules for preprocessing, generation, multi-level verification, and adaptation. The transformation of design variables into performance metrics is formalized using analytical models, machine learning, and imitation modeling. The approach ensures the reliability and interpretability of conclusions, as well as applicability in a business context.

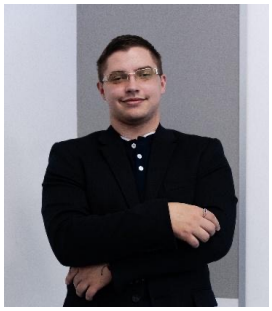
Keywords: large language models, decision support system, RAG, ReAct, hybrid model, unstructured data.

УДК 004.043

ИССЛЕДОВАНИЕ ВЛИЯНИЯ ОРИЕНТАЦИИ ДОКУМЕНТА НА ТОЧНОСТЬ ИЗВЛЕЧЕНИЯ ДАННЫХ МУЛЬТИМОДАЛЬНОЙ МОДЕЛЬЮ PADDLEOCR-VL



Е.А. Курлюк
Студент факультета компьютерного проектирования кафедры электронной техники и технологий БГУИР
kurluke750@gmail.com



Н.А. Ларченко
Студент факультета компьютерного проектирования кафедры электронной техники и технологий БГУИР
nikitadeve@gmail.com



А.В. Бойдич
Студент инженерно-экономического факультета кафедры экономической информатики БГУИР
andrewwboidich@gmail.com



М.В. Давыдов
Первый проректор, доцент, кандидат технических наук, БГИУР
davydov-mv@bsuir.by

Е.А. Курлюк
Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки медицинских данных, оптимизацией извлечения и структурирования информации из медицинских документов, разработкой масштабируемых информационно-компьютерных систем и сервисов.

Н.А. Ларченко
Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки медицинских данных, построением нейросетевых систем анализа физиологических сигналов, разработкой масштабируемых информационно-компьютерных систем и сервисов.

А.В. Бойдич
Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки данных, построением и исследованием нейросетевых систем.

М.В. Давыдов
Первый проректор Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки данных, цифровой обработкой сигналов, электростимуляцией и транскраниальной магнитной стимуляцией.

Аннотация. В работе исследуется влияние ориентации изображения на качество извлечения табличной структуры мультимодальной моделью PaddleOCR-VL. Эксперименты проведены на подмножестве набора данных SynthTabNet с четырьмя фиксированными углами поворота. Показано, что изменение ориентации приводит к существенному снижению качества распознавания, особенно при повороте на 180°. Предложен этап предварительного определения угла поворота на основе классификатора ResNet34. Добавление данного этапа приводит к выравниванию распределений метрик TEDS и TEDS-struct и снижению разброса результатов. Полученные результаты демонстрируют эффективность включения отдельного модуля определения ориентации в конвейер обработки документов.

Ключевые слова: оптическое распознавание текста, мультимодальные модели, анализ структуры таблиц.

Введение. Оптическое распознавание текста (OCR) представляет собой совокупность методов и алгоритмов, направленных на извлечение текстовой информации из изображений и сканированных документов. OCR-системы широко применяются в задачах цифровизации архивов, автоматизации документооборота, обработки медицинских и финансовых записей, а также в системах интеллектуального анализа данных [1-4].

Точность распознавания зависит от качества входных изображений. На результат влияют шум, низкое разрешение, сложный фон и геометрические искажения. Существенным фактором является ориентация документа [5].

На практике изображения документов могут быть повернуты относительно исходной ориентации текста. Это возникает при сканировании, фотографировании документов без фиксации положения камеры, обработке потоков изображений. В таких случаях файл сохраняется без изменения, но текст оказывается повернут относительно горизонтальной оси. Малые углы поворота документа (5-10°) корректируются алгоритмами выравнивания, основанными на анализе текстовых линий [6]. Повороты на 90°, 180° и 270° не относятся к задаче коррекции наклона и требуют отдельного анализа, так как изменяют ориентацию текста на уровне всего изображения и влияют на результат распознавания [7, 8].

Современные подходы к распознаванию текста основаны на мультимодальных моделях, объединяющих визуальные признаки изображения и языковые зависимости текста [9, 10]. Такие модели учитывают локальные графические характеристики символов, контекст на уровне слов, строк и структуры документа. Это позволяет повысить точность извлечения данных в условиях шума, сложного фона и вариативности оформления документов. При этом влияние значительных поворотов изображения на качество работы подобных моделей остаётся недостаточно изученным. В данной работе рассматривается мультимодальная модель PaddleOCR-VL, разработанная в рамках экосистемы PaddleOCR [11]. Модель ориентирована на извлечение структурированной информации из документов и объединяет методы компьютерного зрения и обработки естественного языка. Несмотря на наличие в экосистеме PaddleX инструментов для коррекции ориентации изображений, в рамках настоящего исследования анализируется непосредственно поведение модели PaddleOCR-VL без использования дополнительных модулей выравнивания. В работе рассматривается модель PaddleOCR-VL, разработанная в рамках экосистемы PaddleOCR [11]. Модель предназначена для извлечения текстовой и структурированной информации из документов. В рамках экосистемы PaddleOCR модель PaddleOCR-VL извлекает структурированную текстовую информацию из документов посредством последовательного двухэтапного анализа. На первом этапе алгоритм генерирует визуальные признаки изображения и локализует текстовые области, на втором – выполняет посимвольное распознавание с формированием последовательностей. Итоговая обработка объединяет контекст текстовых данных с их пространственным расположением и визуальными признаками документа, что обеспечивает точное определение логической структуры и семантическую экстракцию данных.

Целью работы является исследование влияния ориентации документа на точность извлечения данных мультимодальной моделью PaddleOCR-VL.

Методология проведения эксперимента. Для исследования влияния ориентации изображения на качество разбора таблиц была выбрана двухэтапная схема эксперимента. На первом этапе оценивалась работа исходной модели без коррекции ориентации входных данных. На втором этапе перед подачей изображения в модель добавлялся отдельный модуль определения угла поворота, после чего выполнялась повторная оценка на том же наборе данных. Такой подход позволил раздельно оценить чувствительность модели к повороту изображения и эффект от предварительного определения и корректировки ориентации. Для проведения эксперимента использовалось готовое окружение в Docker с установленными зависимостями и предварительно загруженными весами PaddleOCR-VL. Все вычисления выполнялись на графическом ускорителе Nvidia RTX 4060.

При запуске PaddleOCR-VL использовалась фиксированная текстовая инструкция «Table Recognition:». Максимальное число генерируемых токенов задавалось равным 512. Для ускорения вычислений использовался Flash Attention 2, при этом расчёты выполнялись в формате FP16. Генерация проводилась в детерминированном режиме (стохастическая выборка была отключена), для каждого изображения формировалась одна выходная последовательность. Такая настройка исключала случайные флуктуации значений между запусками и позволяла получать сопоставимые результаты при сравнении разных углов поворота. Результат работы модели формировался в формате OTSL.

Оценка модели выполнялась на наборе данных SynthTabNet. Для анализа влияния ориентации использовался поднабор из 1000 изображений. Для каждого исходного изображения формировались изображения с четырьмя фиксированными ориентациями: 0°, 90°, 180° и 270°. Это позволяло оценить поведение модели при отсутствии поворота и при трёх дискретных вариантах изменения ориентации. На данном этапе коррекция положения изображения не выполнялась, каждое изображение в заданной ориентации напрямую подавалось в модель, после чего сохранялся результат распознавания таблицы.

Оценка качества выполнялась по метрике TEDS, определяемой как

$$\mathrm{TEDS}=1-\frac{\mathrm{distance}}{\mathrm{nnodes}}, \quad (1)$$

где distance – расстояние редактирования между структурой предсказанной и эталонной таблиц, nnodes – число элементов в эталонной разметке. Также использовалась метрика TEDS-struct, оценивающая только структурное соответствие таблиц без учёта текстового содержимого.

Поскольку выход модели представлен в формате OTSL, было реализовано преобразование в HTML-представление таблицы. Преобразование включало распознавание структуры таблицы, определение ячеек, учёт объединений по строкам и столбцам и формирование корректной HTML-разметки. Полученная разметка использовалась для сравнения с эталонными данными. После завершения первого этапа был добавлен модуль определения угла поворота изображения. Его задача состояла в том, чтобы до запуска PaddleOCR-VL определить, к какому из четырёх классов относится ориентация документа: 0°, 90°, 180° или 270°. На основании этого предсказания изображение приводилось к правильному положению и затем повторно подавалось в PaddleOCR-VL с теми же параметрами, что и на первом этапе.

Для задачи обучения классификатора был сформирован набор данных для обучения модели определения угла поворота. В качестве исходного источника использовалась обучающая часть SynthTabNet. Для каждого изображения выполнялся детерминированный поворот в четыре класса: 0°, 90°, 180° и 270°. Общий объём составил 4000 изображений. Набор был разделён на обучающую, проверочную и тестовую части в соотношении 3200, 400 и 400 изображений соответственно. Для определения угла поворота была использована архитектура ResNet34 с инициализацией весами ImageNet. На вход модели подавалось изображение документа, на выходе предсказывался класс ориентации. Входное изображение приводилось к размеру 224×224. Обучение проводилось в течение 8 эпох с использованием оптимизатора AdamW, скоростью обучения 1e-4 и коэффициентом weight decay 1e-4. В качестве функции потерь использовалась кросс-энтропия. Таким образом, в обеих сериях эксперимента конфигурация основной модели оставалась неизменной, а различие между ними сводилось только к наличию или отсутствию предварительного приведения изображения к правильной ориентации.

Обсуждение результатов. Для оценки поведения модели при различных ориентациях изображения рассматриваются распределения значений метрик TEDS и TEDS-struct по каждому углу поворота. Такой способ представления позволяет сопоставить и средние значения, и характер распределения результатов для разной ориентации изображения. Распределения значений TEDS для PaddleOCR-VL без предварительного определения ориентации приведены на рисунке 1. Для изображений без поворота распределение

сосредоточено в области высоких значений, что соответствует устойчивому распознаванию структуры и содержимого таблиц. При поворотах на 90° и 270° распределения смещаются влево: увеличивается доля результатов со средними значениями метрики, а число примеров с высоким качеством уменьшается. Наиболее выраженное изменение наблюдается при повороте на 180°. В этом случае распределение теряет выраженную концентрацию в правой части шкалы, возрастает доля низких и околонулевых значений, а разброс становится существенно шире.

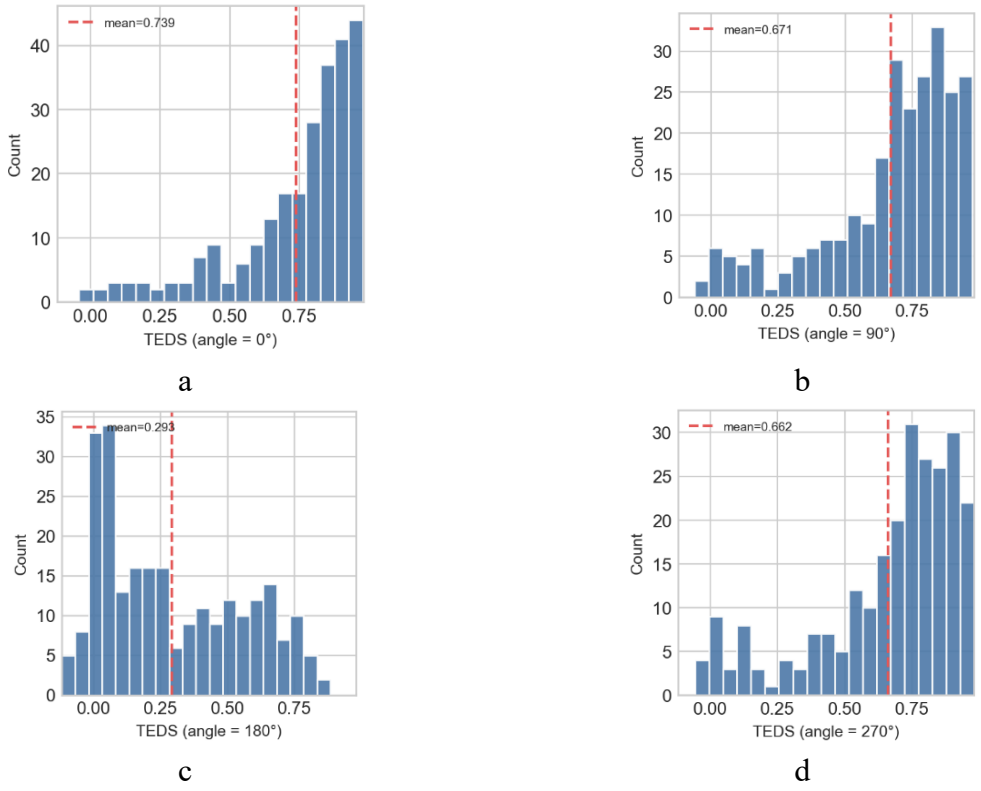


Рисунок 1. Распределение значений TEDS для PaddleOCR-VL без предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

Распределения значений TEDS-struct без предварительного определения ориентации приведены на рисунке 2. Для изображений с ориентацией 0° значения также сосредоточены в правой части шкалы. При поворотах на 90° и 270° сохраняется смещение в сторону меньших значений, однако изменение выражено слабее, чем для полной метрики TEDS. Для поворота на 180° снижение качества также фиксируется, но структура распределения остаётся более устойчивой, чем в случае TEDS. Это означает, что при перевороте изображения модель чаще сохраняет общий каркас таблицы, однако некорректно восстанавливает полное представление таблицы вместе с содержимым и внутренними связями между ячейками. Таким образом, влияние угла поворота сильнее проявляется на уровне полного структурно-текстового соответствия, чем на уровне одной только структуры. Для определения угла поворота обучен классификатор на основе ResNet34. Кривые обучения приведены на рисунке 3. Функция потерь на обучающей и валидационной выборках быстро снижается в первые эпохи и далее практически не изменяется. Значения точности и macro-F1 на валидационной выборке достигают максимальных значений на раннем этапе обучения и остаются стабильными. Точность на валидационной и тестовой выборках составила 1. При применении классификатора к набору данных, использованному в основном эксперименте, точность составила 0.997.

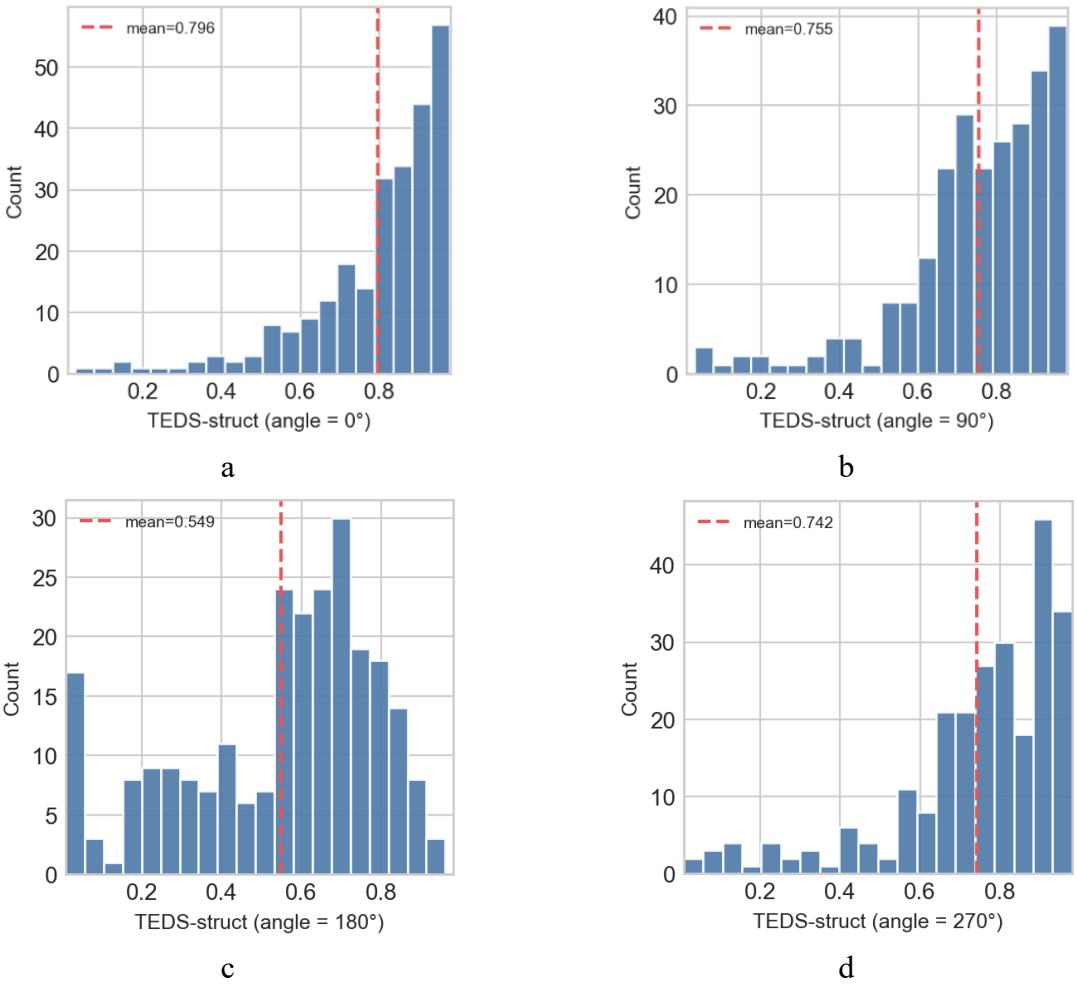


Рисунок 2. Распределение значений TEDS-struct для PaddleOCR-VL без предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

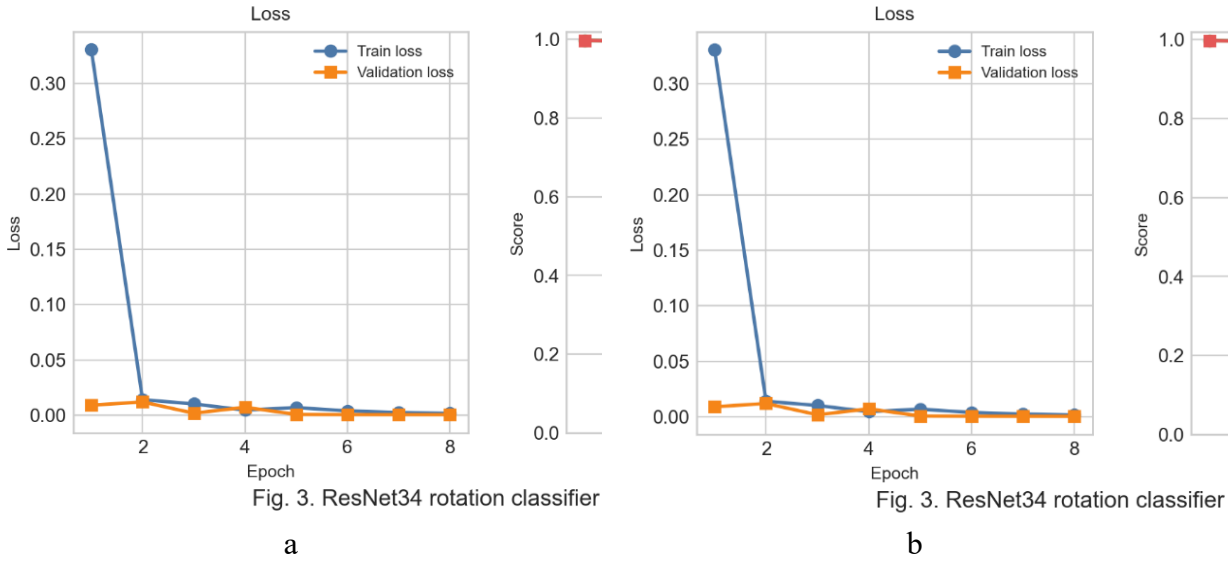


Рисунок 3. Кривые обучения классификатора ResNet34: а) функция потерь на обучающей и валидационной выборках; б) точность и macro-F1 на валидационной выборке.

После добавления классификатора ориентации перед запуском PaddleOCR-VL распределения TEDS изменяются существенным образом. Результаты приведены на рисунке 4.

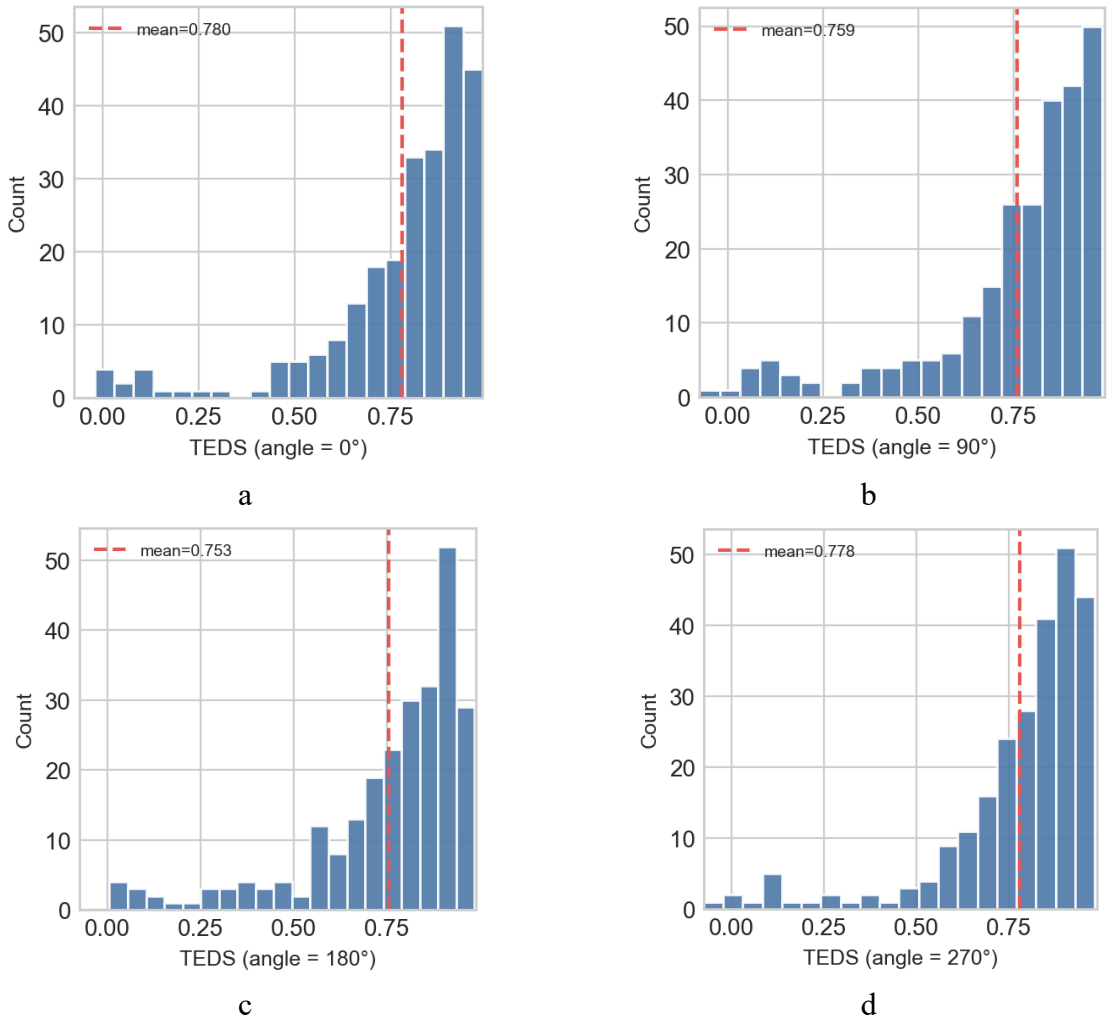


Рисунок 4. Распределение значений TEDS для PaddleOCR-VL после предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

Для всех четырёх углов распределения сосредоточены в области высоких значений и имеют близкую форму. Различие между поднаборами по углам поворота становится значительно менее выраженным. Наибольшее изменение относится к изображениям, повернутым на 180°: поднабор, ранее характеризовавшийся широким разбросом и заметной долей низких значений, после исправления ориентации демонстрирует распределение, сопоставимое с остальными углами. Это указывает на то, что существенная часть ошибок на первом этапе была связана не со сложностью таблиц как таковой, а с неправильным положением изображения относительно ожидаемой ориентации текста. Сопоставление результатов первого и второго этапов показывает, что предварительное определение угла поворота влияет не только на средние значения метрик, но и на форму распределений. На первом этапе качество распознавания зависело от угла поворота, причём наиболее выраженное снижение наблюдалось для изображений, перевёрнутых на 180°. После добавления классификатора ориентации распределения для всех четырёх углов становятся близкими как по положению, так и по форме. Это означает, что влияние ориентации изображения на итоговое качество распознавания в значительной степени снимается ещё

до запуска PaddleOCR-VL. В результате модель работает на более однородном входном наборе, а разброс качества между различными углами поворота существенно сокращается.

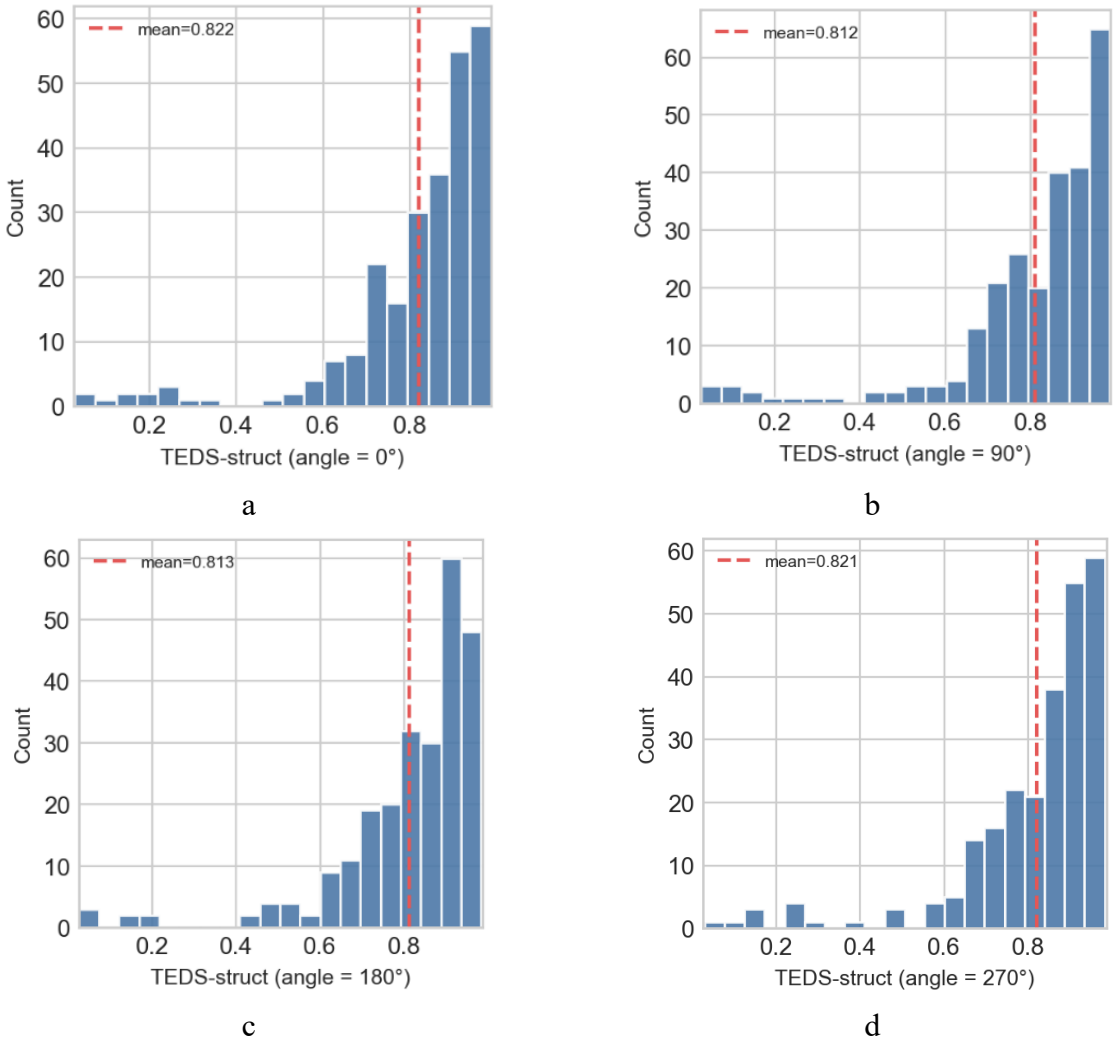


Рисунок 5. Распределение значений TEDS-struct для PaddleOCR-VL после предварительного определения ориентации: а) 0°; б) 90°; в) 180°; г) 270°

Выводы. Полученные результаты подтверждают, что ориентация изображения является значимым фактором, влияющим на качество извлечения табличной структуры. Изменение положения документа приводит к перераспределению значений метрик и увеличению доли некорректных предсказаний, что указывает на зависимость модели от исходной ориентации входных данных.

Введение этапа определения угла поворота изменяет характер работы системы. После приведения изображений к единой ориентации значения метрик стабилизируются, а различия между поднаборами, сформированными по углам поворота, практически исчезают. Это указывает на то, что часть ошибок обусловлена не сложностью самих таблиц, а некорректным положением текста на изображении относительно ожидаемой структуры. Отдельно следует отметить, что классификация ориентации решается с высокой точностью и не требует сложной модели или длительного обучения. Это делает возможным включение данного этапа в общий конвейер без существенного увеличения вычислительных затрат. Полученные результаты показывают, что использование специализированного предварительного этапа обработки позволяет повысить устойчивость мультимодальной модели без изменения её архитектуры.

Список литературы

[1] Fleischhacker, D.; Kern, R.; Göderle, W. Enhancing OCR in historical documents with complex layouts through machine learning // International Journal on Digital Libraries. – 2025. – Vol. 26.

[2] Liu, Y.; Li, Z.; Huang, M.; Zheng, Y.; Zhang, X.; Chen, L. OCRBench: on the hidden mystery of OCR in large multimodal models // Science China Information Sciences. – 2024. – Vol. 67. – Art. 220102.

[3] Ullah, H.; Tanveer, M.; Jan, A. Enhancing handwritten prescription recognition with AI-driven OCR // Journal of Computing & Biomedical Informatics. – 2025.

[4] Nagalaxmi, B.; Sujatha, P.; Sandeep, G.; Lakshmi Varun, G.; Harshavardhan Teja, B. Optimized OCR approaches for accurate text extraction in legal and financial document automation // Journal of Science Engineering Technology and Management Science. – 2025. – Vol. 2, No. 7. – P. 486–495.

[5] Курлюк, Е. А. Применение компьютерного зрения для автоматизированной обработки медицинских документов / Е. А. Курлюк, Н. А. Ларченко, М. В. Давыдов, Е. К. Курлянская // Цифровая трансформация. – 2025. – Т. 31, № 4. – С. 55–64.

[6] Soumya, B. J.; Vasudev, T. Enhancing document image processing: correcting skew in printed documents using deep learning // Journal of Information Systems Engineering and Management. – 2025. – Vol. 10, No. 25s.

[7] Tuggener, L.; Stadelmann, T.; Schmidhuber, J. Efficient rotation invariance in deep neural networks through artificial mental rotation // Frontiers in Computer Science. – 2025. – Vol. 7. – Art. 1644044.

[8] Goswami, S.; Ravi, A.; Kolla, R.; Faraz, A.; et al. Seeing straight: document orientation detection for efficient OCR // arXiv. – 2025. – DOI: 10.48550/arXiv.2511.04161.

[9] Ding, Y.; Han, S. C.; Lee, J.; et al. Deep learning based visually rich document content understanding: a survey // Artificial Intelligence Review. – 2026. – Vol. 59. – Art. 114.

[10] Ke, W.; Zheng, Y.; Li, Y.; Xu, H.; et al. Large language models in document intelligence: a comprehensive survey, recent advances, challenges, and future trends // ACM Transactions on Information Systems. – 2026. – Vol. 44, No. 1. – Art. 18.

[11] Cui, C.; Sun, T.; Liang, S.; Gao, T.; et al. PaddleOCR-VL: Boosting multilingual document parsing via a 0.9B ultra-compact vision-language model // arXiv. – 2025. – DOI: 10.48550/arXiv.2510.14528.

Авторский вклад

Курлюк Евгений Александрович – постановка задачи исследования, разработка методологии оценки эффективности предварительной обработки изображений при извлечении текстовой информации; проектирование и обучение классификатора на базе ResNet34; анализ влияния геометрических искажений на метрики TEDS и TEDS-struct.

Ларченко Никита Александрович – постановка задачи исследования, подготовка выборок данных на основе датасета SynthTabNet; проведение испытаний PaddleOCR-VL на всех этапах исследования, участие в сборе статистических данных, формулировка выводов и направлений для дальнейших исследований.

Бойдич Андрей Викторович – постановка задачи исследования; сравнительный анализ распределений метрик до и после внедрения модуля коррекции; интерпретация результатов распознавания табличных структур и подготовка графических материалов исследования.

Давыдов Максим Викторович – постановка задачи исследования; руководство исследованием по оценке устойчивости мультимодальных моделей к изменению ориентации входных данных; формулировка выводов и направлений для дальнейших исследований.

STUDY OF THE IMPACT OF DOCUMENT ORIENTATION ON DATA EXTRACTION ACCURACY USING THE PADDLEOCR-VL MULTIMODAL MODEL

E.A. Kurlyuk
Student of the Faculty of Computer-Aided Design of the Department of Electronic Engineering and Technologies BSUIR

N.A. Larchenko
Student of the Faculty of Computer-Aided Design of the Department of Electronic Engineering and Technologies BSUIR

A.V. Boidzich
Student of the Faculty of Engineering and Economics, Department of Economic Informatics BSUIR

M.V. Davydov
First Vice-Rector, PhD, BSUIR

Abstract. This paper investigates the impact of image orientation on table structure extraction quality using the PaddleOCR-VL multimodal model. Experiments are conducted on a subset of the SynthTabNet dataset with four fixed rotation angles. The results show that changes in orientation lead to a significant degradation in recognition quality, particularly for 180° rotations. A preprocessing stage based on a ResNet34 rotation classifier is introduced to determine the image orientation prior to inference. Incorporating this stage results in more consistent TEDS and TEDS-struct score distributions and reduces performance variability. The findings demonstrate the effectiveness of integrating a dedicated orientation detection module into the document processing pipeline.

Keywords: OCR, multimodal models, table structure analysis.

УДК 004.8

РАЗРАБОТКА НЕЙРОСЕТЕВОГО КОНСУЛЬТАНТА ДЛЯ СТУДЕНТОВ НА БАЗЕ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ



Т.А. Васяева
Декан факультета информационных систем и технологий, доцент кафедры автоматизированных систем управления ДонНТУ, кандидат технических наук, доцент
tanechka.vasyaeva@yandex.ru



М.В. Душа
студент IV-го курса бакалавриата кафедры автоматизированных систем управления ДонНТУ
dushamihail@yandex.ru

Т.А. Васяева

Окончила Донецкий национальный технический университет. Область научных интересов связана с разработкой методов и алгоритмов искусственного интеллекта, организацией учебного и научно-исследовательского процессов в техническом университете.

М.В. Душа

Студент Донецкого национального технического университета. Область научных интересов связана с разработкой систем машинного обучения

Аннотация. Статья посвящена разработке нейросетевого консультанта информационной поддержки студентов на базе технологии Retrieval-Augmented Generation (RAG) с использованием российских больших языковых моделей. Проведен сравнительный анализ эмбеддинг моделей, методов индексации документов инструментами фреймворка LlamaIndex, генеративных моделей GigaChat-2-Lite, GigaChat-2-Pro и DeepSeek-V3.2. Продемонстрирован оптимальный набор средств для построения RAG системы на основе внутривузовских нормативных актов: ru-en-RoSBERTa, векторное индексирование, GigaChat

Ключевые слова: большие языковые модели, RAG, GigaChat, LlamaIndex, эмбеддинг-модели, векторная индексация, LLM as a Judge, нейросетевой консультант, информационная поддержка студентов.

Введение. Деятельность современного студента состоит из множества зачастую разрозненных направлений, включая учебную, научную, административную и внеучебную сферы. В связи с этим критически важно получать актуальную информацию своевременно. Однако структура информационной поддержки высшего учебного заведения не может в полной мере закрыть эту потребность. Работа в строго ограниченное время и необходимость в физическом присутствии сотрудника деканата для получения консультации способствуют замедлению учебного процесса и повышению стресса студентов. Автоматизированные цифровые консультанты могут стать практичным решением, благодаря доступу к информации 24/7. Применение генеративных нейронных сетей позволит получать простые для понимания ответы, а формат чат-бота – удобный и интуитивно понятный интерфейс взаимодействия с консультантом.

В связи с нынешними санкционными ограничениями использование популярных западных нейронный сетей ограничено. Поиск качественных отечественных больших языковых моделей позволит разрабатывать современные продукты на базе искусственного интеллекта поддерживая политику цифрового суверенитета.

Цель исследования: решение проблемы ограниченного доступа к актуальной информации путём обеспечения информационной поддержки студентов 24/7 за счет разработки нейросетевого консультанта на базе технологии RAG с использованием российских больших языковых моделей.Задача исследования: провести сравнительный анализ и выбор эмбеддинг моделей, методов индексации, больших языковых моделей наиболее оптимальных для разработки нейросетевого консультанта.

Разработка нейросетевого ассистента с технологией RAG. Технология генерации с дополнительной выборкой (Retrieval Augmented Generation, RAG) – механизм, расширяющий генеративные модели техникой извлечения информации [1]. Приложения, использующие технологию RAG в общем виде, состоят из двух основных компонентов: поискового модуля (retrieval, ретривер) и модуля генерации.

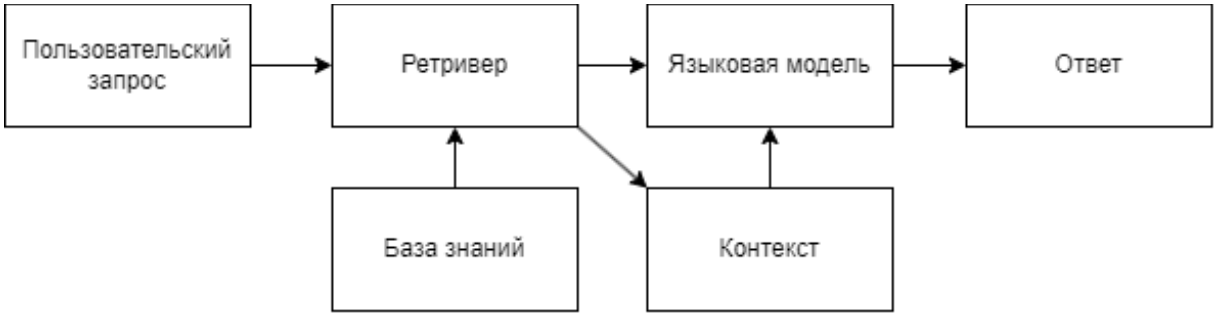


Рисунок 1. Схема пайплайна RAG системы

Рисунок 1 иллюстрирует типовой процесс работы Retrieval-Augmented Generation (RAG) – архитектуры, которая сочетает в себе поиск информации (Retrieval) и генерацию текста (Generation). Основная цель системы – дать языковой модели (LLM) доступ к своим данным, которые изначально не входили в её обучающую выборку.

Описание работы:

1. Запрос пользователя (Question). Взаимодействие с системой начинается с того, что пользователь задает вопрос (Prompt) через интерфейс чата.

2. Ретривер. Полученный текстовый вопрос передается в поисковый модуль, задачей которого является нахождение наиболее релевантных документов в базе знаний.

3. База знаний. В общем виде содержит информацию разных форматов, отвечающей специализированной области.

4. Получение релевантного контекста (Relevant Chunks). В результате поиска извлекаются фрагменты текста (чанки), которые наиболее близки по смыслу к заданному вопросу. Эти фрагменты содержат актуальную информацию, необходимую для ответа.

5. Формирование расширенного промпта. Исходный вопрос пользователя и найденные релевантные фрагменты объединяются в один расширенный промпт. Таким образом, языковой модели передается не только вопрос, но и информацию, на основе которой необходимо дать ответ.

6. Генерация финального ответа (LLM Answer). Большая языковая модель (LLM) получает этот обогащенный промпт и генерирует на его основе итоговый ответ пользователю. Благодаря добавленному контексту, ответ получается основанным на предоставленных данных.

На основании доступных инструментов была выбрана следующая последовательность оценки системы: анализ и оптимизация контекста, поиск наилучшей эмбеддинг модели, поиск лучшего способа индексирования документов, сравнительный анализ результатов генерации языковых моделей.

Компоненты системы и данные. Информационным контекстом системы являются нормативные акты ДонНТУ, поставляются как не оцифрованные файлы формата pdf. Для извлечения информации из таких файлов используется технология оптического распознавания символов (OCR), позволяющая преобразовать текст с изображений в машиночитаемый формат. В первой реализации использовалась библиотека pyteseract, представляющим собой обертку для Google's Tesseract-OCR Engine [2]. В ходе вычисления метрики accuracy, означающую долю правильно найденных документов на общее число запросов, для эмбеддинг модели, значение метрики в среднем принимала значение 0.1, что указывает на крайне низкие поисковые возможности ретривера. Причиной низкого значения является наличие сложных таблиц в документах и перекрытие некоторой информации подписями и печатями, которые pyteseract не смог обработать правильно. Во второй реализации был использован инструмент фрейморка LlamaParse [3], режим сохранения текста в формате языка разметки markdown. По результатам оценки accuracy в среднем увеличилась до 0.8. Основным инструментом анализа компонентов RAG системы в исследовании выступает фреймворк с открытым исходным кодом LlamaIndex [4], доступный в РФ, предоставляющий множество инструментов обработки и извлечения информации, имеет множество интеграций с популярными большими языковыми моделями.

В ходе исследования были протестированы 3 механизма индексации LlamaIndex: Vector Index, List Index и Keyword Table Index. Vector Index – общепринятый стандарт основанный на семантическом поиске. List Index представляет все документы в виде списка. Для поисков релевантных документов передает пользовательский запрос и каждый документ в промпт языковой модели для оценки релевантности. Keyword Table Index формирует таблицу ключевых слов из документов. Поиск релевантных документов осуществляется по наибольшему числу совпавших ключевых слов. Для оценки методов эффективность был использован инструмент фреймворка RetrieverEvaluator, автоматический генерирующий вопросы на основе загруженных документов и вычисляя метрики ранжирования. Тестирование проводилось на 208 вопросах, включающих в себя, например, «На основании каких документов и в каком порядке стипендиальная комиссия определяет список обучающихся, рекомендованных к назначению ПГАС?», «Рассчитайте общее количество баллов, которое получит студент, если он получил только оценки «отлично» в течение трёх семестров подряд, предшествующих назначению ПГАС. Объясните, как вы пришли к этому расчёту.». Vector Index использовался с эмбеддинг модель ru-en-RoSBERTa со следующими параметрами: размер батча – 10; максимальное количество используемых токенов – 512, нормализация векторов включена. List Index и Keyword Table Index использовался с большой языковой моделью DeepSeek-3.2-chat со значением температуры 0.1.

По результатам тестирования (табл. 1) List Index завершил обработку значительно быстрее, получив оценку по метрикам Hit Rate и Recall максимальное значение. Полученные результаты объясняются особенностью работы списочной индексации – List Index оценивает релевантность для всех документов, что гарантировано находит подходящие к запросу, однако полностью исключает ранжирование по семантике, в результате этого значения остальных метрик находятся на низком уровне.

Оценки метрик для Vector Index и Keyword Table Index практически равны, однако скорость обработки запросов Vector Index выше, что является важным фактором для высоконагруженных систем. Исходя из вышеперечисленных выводов, для ретривера нейросетевого консультанта был выбран векторный метод индексирования на основе Vector Index.

Таблица 1. Средние результаты оценок методов индексирования

Метрика	Vector Index	List Index	Keyword Table Index
hit_rate	0,620192	1,000000	0,623188
mrr	0,415385	0,050254	0,352243
precision	0,124038	0,009615	0,070911
recall	0,620192	1,000000	0,623188
ap	0,415385	0,050254	0,352243
ndcg	0,466158	0,207080	0,417133
Общее время обработки, с	47,21	2,63	445,58

Выбор эмбеддинг модели для векторного индексирования прямо влияет на скорость обработки документов и выдачу наиболее релевантной к запросу информации. Для систем RAG используются модели на базе трансформеров. При загрузке документов модель разбивает информацию на чанки, и преобразует их в вектор. Пользовательские запросы также преобразуются в вектор и при помощи метрики косинусного сходства модель извлекает наиболее близкие по значениям вектора документов.

В оценке использовалось 4 open source модели:

- multilingual-e5-small. Мультиязычная модель на 100 млн параметров;
- all-MiniLM-L6-v2. Мультиязычная модель семейства Sentence transformers на 22.7 млн параметров;
- rubert-tiny2. Модель, обученная на русскоязычных текстах на 29.4 млн параметров;
- ru-en-RoSBERTa: Обученная на русскоязычных и англоязычных текстах модель на 400 млн параметров

Оценка проводилась при помощи инструмента фрейморка LlamaIndex InformationRetrievalEvaluator. Оценка работы системы предполагает создание валидационного датасета, строки которого содержат три компонента: запрос пользователя, корпус документов и указание идентификаторов документов, соответствующих данному запросу. Оценка проводилась по основным метрикам ранжирования для 8 пользовательских запросов (табл. 2) [5]. Приведем перечень запросов:

- Что такое повышенная государственная академическая стипендия?
- Что из себя представляет процедура рейтинговой оценки студентов?
- Кто входит в комиссию по предоставлению академического отпуска?
- Какая стипендия положена аспирантам?
- Кто имеет право получить академический отпуск?
- Размер стипендии призедента РФ
- Влияет ли выход на академический отпуск на получение повышенной стипендии?
- Что включает в себя материальное обеспечение детей сирот?

Для вычисления метрик использовались различные значений параметра k, означающего количество релевантных к контексту чанков. При значении k=10 значение метрики recall для всех моделей принимало значение 1 и приводило к падению метрики precision с 0,6250–0,4750 до 0,3750. Такое поведение обусловлено ситуацией, когда найденные документы имеют высокое семантическое сходство, но низкую релевантность к контексту. Наилучшее значение показала модель ru-en-RoSBERTa, сохранив высокие значения precision для k в диапазоне от 1 до 10 и показав лучшие значения для метрик интегрального ранжирования. Для дальнейшего анализа выбрано значение k=5, что соответствует высоким значениям метрики экономит входные токены. По результатам оценок, значительных статистических отличий между мультиязычными и русскоязычными моделями нет, однако русскоязычные модели предпочтительны в выборе с точки зрения построения более структурно правильных ответов для русскоязычных запросов. В качестве эмбеддинг модели была выбрана ru-en-RoSBERTa, поскольку модель показала наилучшие значения и обучена на преимущественно русскоязычных текстах.

Таблица 2. Средние результаты оценок эмбеддинг моделей

Метрика	multilingual-e5-small	all-MiniLM-L6-v2	rubert-tiny2	ru-en-RoSBERTa
accuracy	0,843750	0,843750	0,781250	0,843750
map	0,649681	0,631548	0,605506	0,687798
mrr	0,681250	0,531250	0,503906	0,691406
ndcg	0,638635	0,562953	0,525936	0,657915
precision	0,483333	0,450000	0,433333	0,516667
recall	0,552083	0,581845	0,519345	0,550595

Оценка генеративных моделей. Модулем генерации в целом может выступать любая большая языковая модель, однако в современных реалиях критическими факторами становиться доступность и возможность качественной генерации русскоязычного текста. Немаловажными являются характеристики, формирующие общий показатель качества модели: высокие значения при оценках на популярных бенчмарках, использование современной архитектуры, наличие application programming interface (API), подробная документация. Приведем описание особенностей современных больших языковых моделей.

GigaChat – семейство языковых моделей, разработанных компанией Сбер в России в 2023 году [6]. Взаимодействие с моделью доступно через чат-интерфейс и API. Через API доступны следующие модели: GigaChat-2-Lite для быстрого решения рутинных задач без высоких ресурсных затрат; GigaChat-2-Pro для сложных запросов и прикладных областей; GigaChat-2-Max для работы в сложных научных областях и длинного контекста. Бесплатно предоставляется ограниченное количество токенов для работы с API [7]. Часть моделей выложены в открытый доступ (open source) [8].

DeepSeek – семейство языковых моделей, разработанных компанией DeepSeek в Китае в 2023 году [9]. DeepSeek были выбраны для анализа как представители моделей, разработанных в дружественных странах. Доступ к чату-интерфейсу свободно доступен в России. Использование API платное, однако пополнить баланс возможно только через сторонние сервисы. Через API доступны следующие модели: DeepSeek-v3.2-chat с меньшим количеством выходных токенов и DeepSeek-V3.2-reasoner с большим количеством выходных токенов и дополнительными рассуждениями при обработке запроса [10]. Модели выложены в открытый доступ [11]. GPT – семейство языковых моделей, разработанных компанией Open AI в США [12], модель приведена как представитель западных решений. Доступ к чату-интерфейсу в России заблокирован. Платное API, пополнение доступно только через сторонние сервисы. Доступно множество моделей, включающие новейшие реализации GPT-5.3, так и не поддерживаемые, например, GPT-4.

Таблица 3. Сравнительный анализ моделей языковых моделей

Характеристика	DeepSeek-V3.2	Chat GPT-4o	GigaChat-2-Lite	GigaChat-2-Pro
Прямое пополнение API из РФ	Нет	Нет	Да	Да
Поддержка русского языка	Да	Да	Да	Да
Открытый исходный код	Да	Нет	Да	Да
Архитектура	MoE	-	MoE	MoE
Размер контекстного окна	128 тыс токенов	128 тыс токенов	128 тыс токенов	128 тыс токенов
MMLU	87%	88%	72%	82%
MATH	61%	76%	64%	78%
HumanEval	65%	90%	74%	91%

Перефразированный и уточнённый вариант:

Обобщим обзорную информацию, дополнив её результатами оценки моделей на бенчмарках из статей [11–15] (таблица 3).

Значения в таблице приведены в процентах правильно решённых задач.

Описание бенчмарков:

- MMLU. мультимодальный бенчмарк, оценивающий знания и способность модели решать задачи в 57 различных предметных областях (гуманитарные науки, социальные науки, STEM и др.) с помощью вопросов с выбором ответа.
- MATH. набор из 12 500 сложных математических задач разной сложности, предназначенный для оценки математического мышления и навыков решения задач.
- HumanEval. бенчмарк для оценки способности модели генерировать функционально корректный код.

На основании выбранных технологий ретривера проведена оценка генерации ответов больших языковых моделей.

Для оценки использовались метрики предоставляемые Llama index, использующие метод оценки llm-as-judge, автоматизирующий экспертную оценку путем использования другой большой языковой модели в качестве оценщика.

В проведенном анализе арбитром выступает модель DeepSeek-V3.2-reasoner.

Корпус пользовательских запросов состоял из 8 вопросов, использованных при оценке эмбеддингов.

Параметры генерации для всех моделей были одинаковые: контекстное окно – 128 тыс. токенов, температура – 0.1, k – 5.

Таблица 4. Средние результаты оценок генерации моделей

Модель	Faithfulness score	Answer Relevancy score	Correctness score	Relavency score
DeepSeek-V3.2-chat	1,000	0,928571	4,5625	1,000
GigaChat-2-Lite	0,625	0,500000	3,3125	0,500
GigaChat-2-Pro	0,625	0,562500	3,0625	0,625

Для оценки каждая модель запускалась в формате чата, позволяя использовать системный промпт для регулирования ответов моделей. На основании исследования [16], инструкции, требующие исключить генерацию ответов, если переданный контекст пуст или не релевантен запросу, записаны прописными буквами.

Однако модели GigaChat при таком подходе перестали генерировать ответы на пользовательские запросы, отправляя ответ-заглушку, в общем виде означающий, что модель «не может обсуждать чувствительные темы» или «не обладает собственным мнением». Формирование системного промпта без заглавных букв и общее «смягчение» требований к ответам уменьшило частоту появления ответов-заглушек.

Итоговый системный промпт представлен ниже:

«Ты консультант ДонНТУ. Отвечай строго по документам ниже.

Если в контексте нет ответа – напиши "Информация в документах отсутствует или недостаточна".

Контекст из документов:

{context_str}

Вопрос пользователя:

{query_str}

Ответ:»

По результатам оценок (табл. 4) DeepSeek-V3.2-chat показал значительное преимущество по метрикам. Низкие оценки средних метрик моделей GigaChat обусловлены ответами-заглушками, которые все ещё встречаются в ответах вследствие чего в среднюю

оценку добавляются нули. Ответы такого типа встречаются на такие запросы как «Кто имеет право получить академический отпуск?» или «Что включает в себя материальное обеспечение детей сирот?». Модель DeepSeek-V3.2-chat для тех же запросов генерирует ответы, метрики которых близки к максимальным.

На Рисунке 2 представлены результаты оценок метрик для каждого вопроса для каждой языковой модели. По оси x обозначен номер вопроса, по оси y – значение метрики. Оценивая значения метрик без учета ответов-заглушек, метрики моделей GigaChat-2-lite и GigaChat-2-pro близки к ответам DeepSeek-V3.2-chat.

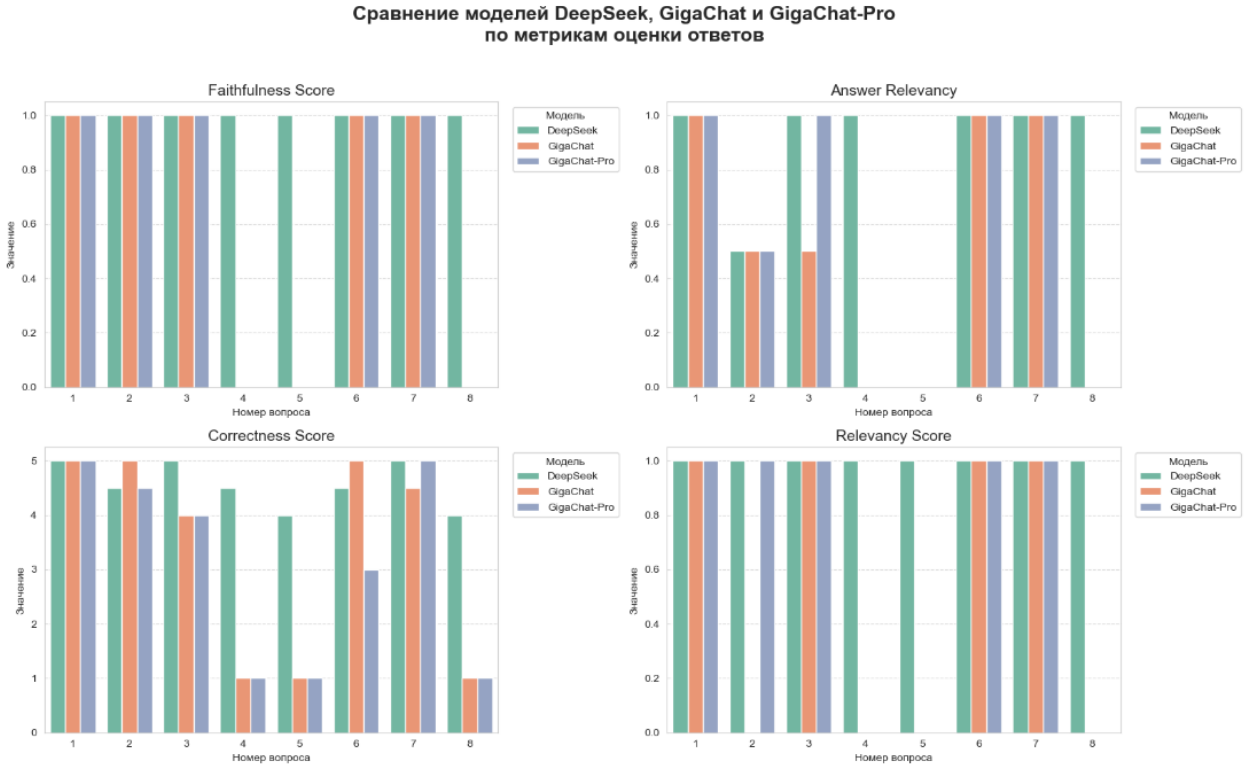


Рисунок 2. Сравнение моделей по метрикам оценки ответов

Таким образом, уровень генерации ответов моделей GigaChat-2-lite и GigaChat-2-pro сопоставим с уровнем генерации DeepSeek-V3.2-chat, что в совокупности с доступностью моделей GigaChat на территории РФ, склоняет к выбору моделей от компании Сбер для построения консультирующей RAG системы.

Заключение. В результате исследования подобрана оптимальная конфигурация технологий для построения нейросетевого консультанта: извлечение информации из неоцифрованных документов при помощи llama parse, векторная индексация с эмбендинг моделью ru-en-RoSBERTa, модуль генерации ответов на запросы на основе GigaChat-2-Lite и GigaChat-2-Pro.

Использование Open Source технологий и отечественных больших языковых моделей позволяет исключить влияние санкционных ограничений на стабильность работы системы и использовать современные подходы построения RAG систем.

В качестве дальнейшего исследование планируется разработка дружественного пользовательского интерфейса для студента и для администратора, который занимается поддержкой системы (управление хранилищем, актуализация нормативной базы и тестирование консультанта).

Список литературы

[1] Ротман Д. RAG и генеративный ИИ. Создаем собственные RAG-пайплайны с помощью LlamaIndex, Deep Lake и Pinecon. – СПб. : Питер, 2026. – 320 с.

[2] Tesseract OCR [Электронный ресурс] // GitHub. – URL: https://github.com/UB-Mannheim/tesseract/wiki.

[3] LlamaParse [Электронный ресурс] // GitHub. – URL: https://github.com/maxgoff/llama_parse.

[4] LlamaIndex [Электронный ресурс]. – URL: https://www.llamaindex.ai/.

[5] Задача ранжирования [Электронный ресурс] // Яндекс Образование. – URL: https://education.yandex.ru/handbook/ml/article/zadacha-ranzhirovaniya.

[6] Зачем нужен GigaChat? [Электронный ресурс] // SberDevices. – URL: https://sberdevices.ru/press/blog/detail/zachem-nuzhen-giga-chat/.

[7] Обновления моделей GigaChat [Электронный ресурс]. – URL: https://developers.sber.ru/docs/ru/gigachat/models/updates.

[8] GigaChat3 [Электронный ресурс] // Hugging Face. – URL: https://huggingface.co/collections/ai-sage/gigachat3.

[9] DeepSeek [Электронный ресурс] // Википедия. – URL: https://ru.wikipedia.org/wiki/DeepSeek.

[10] DeepSeek API Docs [Электронный ресурс]. – URL: https://api-docs.deepseek.com.

[11] deepseek-ai/DeepSeek-V3.2-Exp [Электронный ресурс] // Hugging Face. – URL: https://huggingface.co/deepseek-ai/DeepSeek-V3.2-Exp.

[12] GPT-4o [Электронный ресурс]. – URL: https://openai.com/ru-RU/index/hello-gpt-4o/. – (Дата обращения: 31.03.2026).

[13] GigaChat Lite [Электронный ресурс]. – URL: https://developers.sber.ru/docs/ru/gigachat/models/gigachat-2-lite.

[14] GigaChat Pro [Электронный ресурс]. – URL: https://developers.sber.ru/docs/ru/gigachat/models/gigachat-2-pro.

[15] Оценка производительности DeepSeek-V3 [Электронный ресурс] // Habr. – URL: https://habr.com/ru/articles/955282/.

[16] CAPS for prompt engineering [Электронный ресурс]. – URL: https://kth.diva-portal.org/smash/record.jsf?dswid=-5361&pid=diva2%3A1954037.

Авторский вклад

Васяева Татьяна Александровна – руководство исследованием, участие в проведении анализа полученных данных.

Душа Михаил Валерьевич – реализация RAG системы, подбор метрик и параметров системы, анализ полученных данных.

DEVELOPMENT OF A NEURAL NETWORK CONSULTANT FOR STUDENTS BASED ON LARGE LANGUAGE MODELS

T.A. Vasyaeva

Dean of the Faculty of Information Systems and Technologies, Associate Professor of the Department of Automated Control Systems, Donetsk National Technical University, Candidate of Technical Sciences, Associate Professor

tanechka.vasyaeva@yandex.ru

M.V. Dusha

4th-year Bachelor Student, Department of Automated Control Systems, Donetsk National Technical University

dushamihail@yandex.ru

Abstract. The article is devoted to the development of a neural network consultant for student information support based on Retrieval-Augmented Generation (RAG) technology using Russian large language models. A comparative analysis of embedding models, document indexing methods using the LlamaIndex framework, and generative models GigaChat-2-Lite, GigaChat-2-Pro, and DeepSeek-V3.2 was conducted. The optimal set of tools is demonstrated: ru-en-RoSBERTa, vector indexing, and GigaChat**.**

Keywords. large language models, RAG, GigaChat, LlamaIndex, embedding models, vector indexing, LLM as a Judge, neural network consultant, student information support.

УДК 004.912:004.8

TRANSFORMER-АРХИТЕКТУРА: КАК ИЗМЕНИЛА ПОДХОД К РАЗРАБОТКЕ СИСТЕМ ВЫЯВЛЕНИЯ ТЕКСТОВЫХ ЗАИМСТВОВАНИЙ



К.С. Крез
Ассистент кафедры проектирования информационно-компьютерных систем БГУИР,
магистр технических наук
k.krez@bsuir.by



М.А. Кривоносова
Студентка 3 курса БГУИР Факультета компьютерного проектирования по специальности Информационные системы и технологии (в бизнес-менеджменте)
krvria@gmail.com



А.Р. Шипуль
Студентка 3 курса БГУИР Факультета компьютерного проектирования по специальности Информационные системы и технологии (в бизнес-менеджменте)
angelinashipul@gmail.com



А.С. Гугалев
Студент 3 курса БГУИР Факультета компьютерного проектирования по специальности Информационные системы и технологии (в бизнес-менеджменте)
gugalevandrei@gmail.com

К.С. Крез
Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов: корреляционный анализ цифровых следов пользователей, нейронные сети, разработка и анализ структуры хранения данных.

М.А. Кривоносова
Студентка 3 курса Белорусского государственного университета информатики и радиоэлектроники Факультета компьютерного проектирования по специальности Информационные системы и технологии (в бизнес-менеджменте)

А.Р. Шипуль
Студентка 3 курса Белорусского государственного университета информатики и радиоэлектроники Факультета компьютерного проектирования по специальности Информационные системы и технологии (в бизнес-менеджменте)

А.С. Гугалев

Студент 3 курса Белорусского государственного университета информатики и радиоэлектроники Факультета компьютерного проектирования по специальности Информационные системы и технологии (в бизнес-менеджменте)

Аннотация. В статье рассматривается влияние архитектуры Transformer на развитие методов автоматического выявления текстовых заимствований в современных системах антиплагиата. Показано, что переход от лексико-статистических и строковых методов сравнения к контекстно-зависимому семантическому анализу способствует повышению качества обнаружения перефразированных и частично модифицированных заимствований. Проанализированы ограничения традиционных подходов, основанных на шинглах, n-граммах и статических эмбеддингах, а также раскрыта роль механизма self-attention в формировании контекстуальных представлений текстовых фрагментов. Особое внимание уделено применению архитектурных стратегий bi-encoder и cross-encoder в составе гибридного конвейера антиплагиатной проверки, обеспечивающего баланс между производительностью и точностью. Представлен упрощённый пример семантического сопоставления сегментов документа с локальным хранилищем и расчёта коэффициента семантических совпадений. Сделан вывод о целесообразности интеграции трансформерных моделей в многоуровневые системы антиплагиата, сочетающие точные лексические алгоритмы и семантический анализ.

Ключевые слова: семантическое сходство текстов, STS, Transformer, self-attention, BERT, SBERT, sentence embeddings, семантический поиск.

Введение. В условиях стремительного роста объёмов цифровых текстов задача автоматического выявления заимствований перестала сводиться к поиску буквальных совпадений. Современная система антиплагиата должна обнаруживать не только дословное копирование, но и перефразирование, замену слов синонимами, перестановку фрагментов, а также частичное изменение формулировок при сохранении исходного смысла [1]. Именно поэтому при разработке таких систем всё большую роль играет семантическое сравнение текстов Semantic Textual Similarity (далее – STS), позволяющее оценивать близость фрагментов на уровне смысла, а не только на уровне совпадения слов.

Существовавшие ранее системы обнаружения заимствований на протяжении длительного времени основывались преимущественно на лексических методах, шингловом анализе и строковых алгоритмах сопоставления текстов. Эти подходы остаются полезными для выявления прямых заимствований, однако их эффективность заметно снижается в случаях смыслового переписывания текста. Появление архитектуры Transformer стало важным этапом развития интеллектуальных систем проверки текстовых работ, поскольку позволило перейти от поверхностного сопоставления к контекстно-зависимому анализу содержания. В результате современные антиплагиатные системы приобретают способность учитывать семантическую близость текстовых фрагментов, что существенно расширяет их функциональные возможности в образовательной и научной сфере.

Основная часть. Развитие систем автоматической проверки текстов на заимствования приоритет отдавался методам, ориентированным на поиск точных или частично изменённых совпадений текстовых последовательностей. Для этого применялись алгоритмы строкового сравнения, n-граммный анализ, шингли и лексико-статистические методы, в том числе Term Frequency – Inverse Document Frequency (далее – TF-IDF). Такие решения достаточно эффективно выявляли прямое копирование и заимствования с минимальными изменениями. Однако при перефразировании, перестановке слов или замене отдельных лексических единиц их точность заметно снижалась. По этой причине тексты, выражающие одно и то же содержание разными языковыми средствами, часто оценивались системой как различные.

Использование векторных представлений слов, таких как Word2Vec, GloVe и FastText [2], стало значимым шагом в развитии методов анализа текстового сходства. Эти модели позволили лучше учитывать семантическую близость слов и терминов, благодаря чему повысилась эффективность обработки текстов, связанных общей тематикой. Однако их ключевым ограничением оставался статический характер представлений: каждому слову соответствовал один и тот же вектор независимо от контекста. В задачах выявления заимствований это создавало риск неточной интерпретации многозначных слов, особенно в

научно-технических текстах, где смысл языковой единицы определяется не только самим словом, но и его окружением.

Дальнейшее развитие методов обработки текста сопровождалось применением рекуррентных нейронных сетей и моделей Long Short-Term Memory (далее – LSTM), позволивших в большей степени учитывать порядок слов и структуру последовательности. Однако высокая вычислительная сложность и ограниченная эффективность при обработке длинных текстов затрудняли их использование в высоконагруженных антиплагиатных системах, ориентированных на анализ больших массивов документов.

В системах автоматической проверки текстов стали появлятся архитектуры Transformer, в основе которой лежит механизм self-attention. Его особенность состоит в том, что при обработке текста каждый токен рассматривается не изолированно, а с учётом связи со всеми остальными элементами последовательности [3]. Для задач обнаружения заимствований это особенно важно, так как смысл фрагмента нередко сохраняется даже после изменения структуры предложения, замены отдельных слов и частичного перефразирования. Благодаря этому self-attention позволяет глубже учитывать контекст, точнее выявлять смысловые соответствия между фрагментами и более надёжно отличать тематическое сходство от фактического смыслового заимствования.

Появление моделей семейства Bidirectional Encoder Representations from Transformers (далее – BERT) и их производных сделало возможным использование контекстуальных эмбеддингов текстовых фрагментов в системах обнаружения заимствований.

В результате задача антиплагиата постепенно трансформируется из задачи поиска текстовых совпадений в задачу многоуровневого анализа, включающего лексическое, структурное и семантическое сравнение. Трансофрмер-модели выступают в этом процессе как ядро семантического уровня, позволяя обнаруживать скрытые и перефразированные заимствования.

При разработке системы антиплагиата на базе трансформеров ключевым становится выбор архитектурной стратегии сравнения фрагментов текста. В практических системах применяются два базовых подхода – bi-encoder и cross-encoder, каждый из которых решает свою часть задачи [4].

В схеме bi-encoder каждый сегмент проверяемого документа и каждый сегмент фрагмента источников кодируются независимо в фиксированный вектор. Далее сравнение выполняется в векторном пространстве, как правило, с использованием косинусной меры. Такой подход особенно эффективен на этапе предварительного поиска подозрительных совпадений, поскольку эмбеддинги фрагмента можно вычислить заранее и сохранить в индексе. Это позволяет оперативно находить семантически близкие фрагменты даже при работе с крупными локальными хранилищами текстовых работ.

В архитектуре cross-encoder сегмент проверяемого документа и найденный текст-кандидат подаются в модель как единая пара. Такой подход даёт возможность механизму внимания напрямую учитывать связи между двумя фрагментами и за счёт этого точнее оценивать степень их сходства. Хотя данный подход требует больших вычислительных затрат по сравнению с bi-encoder, он особенно полезен на этапе уточняющей проверки, где необходимо более точно оценить ограниченное число отобранных кандидатов и уменьшить количество ложноположительных результатов.

С точки зрения практической реализации в системах обнаружения заимствований наиболее оправданным является применение двухэтапной архитектуры. На первом этапе модель типа bi-encoder обеспечивает быстрый отбор потенциально релевантных кандидатов для каждого сегмента документа. На втором этапе модель типа cross-encoder выполняет уточнённую оценку степени семантической близости между сопоставляемыми фрагментами. Подобная организация процесса позволяет совместить высокую производительность с повышенной точностью анализа, что имеет принципиальное значение при массовой проверке студенческих и научных работ.

Практическая часть. В рамках разработки антиплагиатной системы семантическое сравнение целесообразно рассматривать как один из этапов общего конвейера анализа текстового документа. На начальной стадии выполняется предварительная обработка, включающая извлечение текстового содержимого, удаление служебных элементов, нормализацию и сегментацию документа на смысловые фрагменты. Далее каждый выделенный сегмент преобразуется в векторное представление с использованием модели типа bi-encoder, после чего осуществляется поиск семантически близких фрагментов в локальном хранилище или в базе ранее проверенных работ [5].

Отобранные кандидаты могут передаваться в модуль уточняющего сравнения, реализованный, например, на основе архитектуры cross-encoder, где степень релевантности оценивается с более высокой точностью. На основании полученных оценок формируется карта совпадений по сегментам проверяемого документа. На следующем этапе модуль может вычислять интегральный показатель потенциальных заимствований с учётом количества подтверждённых совпадений, длины соответствующих фрагментов и характера их распределения по структуре документа. Такой подход обеспечивает возможность не только выявления отдельных потенциально заимствованных участков, но и формирования интерпретируемого итогового отчёта, предназначенного для преподавателя, эксперта или иного уполномоченного пользователя.

Ниже может быть приведён упрощённый пример программной реализации семантического модуля антиплагиатной системы. В рамках данного примера сегменты проверяемого документа и сегменты локального хранилища кодируются в эмбеддинги, после чего для каждого сегмента документа определяется наиболее близкий по смыслу сегмент из хранилища. Если значение меры сходства превышает заданный порог, соответствующее совпадение интерпретируется как потенциально релевантное семантическое совпадение. На этой основе может быть рассчитан упрощённый коэффициент семантических совпадений, определяемый как доля сегментов документа, для которых были обнаружены семантически близкие соответствия.

```
    import torch
    import torch.nn.functional as F
    from transformers import AutoTokenizer, AutoModel
    model_name = "google-bert/bert-base-multilingual-cased"
    tokenizer = AutoTokenizer.from_pretrained(model_name)
    model = AutoModel.from_pretrained(model_name)
    def mean_pooling(model_output, attention_mask):
        token_embeddings = model_output.last_hidden_state
        mask = attention_mask.unsqueeze(-
1).expand(token_embeddings.size()).float()
        return torch.sum(token_embeddings * mask, dim=1) /
torch.clamp(mask.sum(dim=1), min=1e-9)
    def encode_texts(texts):
        encoded = tokenizer(
            texts,
            padding=True,
            truncation=True,
            return_tensors="pt"
        )
        with torch.no_grad():
            output = model(**encoded)
        embeddings = mean_pooling(output,
encoded["attention_mask"])
        embeddings = F.normalize(embeddings, p=2, dim=1)
```

```
        return embeddings
    document_segments = [
        "Transformer использует механизм self-attention для анализа
контекста.",
    "Система должна находить не только дословные, но и
перефразированные совпадения."
    ]
    reference_segments = [
        "Механизм self-attention в Transformer позволяет учитывать
связи между токенами.",
    "Система обязана выявлять перефразирование и скрытые
совпадения."
    ]
    doc_emb = encode_texts(document_segments)
    ref_emb = encode_texts(reference_segments)
    threshold = 0.70
    matched = 0
    for i, text in enumerate(document_segments):
        scores = torch.matmul(doc_emb[i], ref_emb.T)
        best_idx = torch.argmax(scores).item()
        score = scores[best_idx].item()
        if score >= threshold:
            matched += 1
        status = "релевантно" if score >= threshold else "не
подтверждено"
        print(f"Исходный сегмент: {text}")
        print(f"Найденный                                    сегмент:
{reference_segments[best_idx]}")
        print(f"Сходство: {score:.4f}")
        print(f"Статус: {status}")
        print("-" * 80)
    print(f"\nДоля    релевантных    совпадений:    {matched    /
len(document_segments):.2%}")
```

С целью повышения точности анализа и снижения количества ложноположительных результатов в системе обнаружения заимствований целесообразно предусмотреть второй этап проверки, реализованный на основе архитектуры cross-encoder. В отличие от bi-encoder, где сегменты сравниваются через заранее вычисленные векторы, cross-encoder получает пару фрагментов одновременно и оценивает их релевантность с учётом прямого межтекстового взаимодействия токенов. Такой подход вычислительно дороже, поэтому он применяется только к ограниченному числу кандидатов, найденных на первом этапе семантического поиска. В результате система получает более точную фильтрацию подозрительных совпадений и более надёжную оценку степени семантической близости сравниваемых фрагментов.

```
    from sentence_transformers import CrossEncoder
    cross_encoder = CrossEncoder("cross-encoder/ms-marco-MiniLM-L-
6-v2")
    pairs = [[m["segment"], m["matched_segment"]] for m in matches]
    cross_scores = cross_encoder.predict(pairs)
    for m, s in zip(matches, cross_scores):
    m["cross_score"] = float(s)
```

```
print(matches[0]["segment"], "->", matches[0]["cross_score"])
```

После этого итоговое решение о подтверждении совпадения может приниматься не только по порогу bi-encoder, но и с учётом cross_score, что особенно полезно при анализе перефразированных фрагментов, близких по теме, но не являющихся фактическим заимствованием.

Ограничения и особенности применения трансформеров в антиплагиате. Несмотря на высокую эффективность, применение моделей класса Transformer в системах обнаружения заимствований сопряжено с рядом ограничений. Во-первых, наличие семантической близости между фрагментами не всегда свидетельствует о некорректном заимствовании. В научных и учебных текстах закономерно встречаются типовые определения, общеупотребимые формулировки и устойчивые терминологические конструкции, совпадение которых по смыслу не может рассматриваться как достаточное основание для вывода о наличии заимствования. В связи с этим итоговое решение должно приниматься на основе совокупности признаков, а не исключительно на значении семантической меры.

Во-вторых, качество обнаружения в значительной степени определяется предметной областью и характеристиками фрагмента, в рамках которого функционирует система. Модель, демонстрирующая высокую результативность на текстах общего назначения, может менее точно интерпретировать специализированные термины и контексты в технических, юридических или медицинских документах. Это обусловливает актуальность задачи доменной адаптации моделей, а также настройки порогов релевантности с учётом специфики проверяемых текстов.

В-третьих, существенное влияние на результат оказывает стратегия сегментации документа. Использование чрезмерно коротких фрагментов может приводить к потере значимого контекста, тогда как чрезмерно длинные сегменты затрудняют точную локализацию потенциального заимствования. Следовательно, эффективность системы обнаружения заимствований определяется не только выбором нейросетевой модели, но и качеством всей методики обработки текста, включая сегментацию, индексацию, переранжирование кандидатов и расчёт итоговых показателей.

Заключение. Архитектура Transformer оказала существенное влияние на развитие систем обнаружения заимствований, обеспечив переход от выявления буквальных совпадений к обнаружению семантически близких и перефразированных фрагментов, требующих дополнительной экспертной интерпретации. Использование механизма self-attention и контекстуальных эмбеддингов позволило расширить аналитические возможности обработки текста и повысить точность выявления смысловых совпадений в тех случаях, когда традиционные строковые и шингловые методы оказываются недостаточно эффективными. Это имеет особое значение при проверке учебных и научных работ, в которых заимствование нередко маскируется посредством замены лексических единиц, перестановки синтаксических конструкций и частичной переработки исходного текста.

С практической точки зрения наиболее целесообразным решением для реальных систем обнаружения заимствований является применение гибридной архитектуры, в рамках которой лексические методы, bi-encoder и cross-encoder функционируют как взаимодополняющие компоненты единого аналитического конвейера. Подобная организация обеспечивает высокую скорость первичного поиска потенциально подозрительных совпадений, более точную семантическую верификацию отобранных кандидатов и возможность формирования интерпретируемого итогового отчёта по результатам анализа. Вместе с тем следует учитывать, что сама по себе семантическая близость не может рассматриваться как достаточный признак некорректного заимствования. В связи с этим итоговая оценка должна формироваться на основе совокупности признаков, включая характер совпадения, тип текстового фрагмента и контекст его использования.

Список литературы

[1] Wahle, J. P. Identifying Machine-Paraphrased Plagiarism / J. P. Wahle [et al.] // Information for a Better World: Shaping the Global Future: 17th International Conference. – Springer, 2022. – С. 393-413.

[2] Krez, K. S. From words to vectors: text vectorization techniques in natural language processing / K. S. Krez // 2. – Минск: БГУИР, 2025. – С. 60-62.

[3] Rogers, A. A Primer in BERTology: What We Know About How BERT Works / A. Rogers, O. Kovaleva, A. Rumshisky // Transactions of the Association for Computational Linguistics. – 2020. – Т. 8. – С. 842-866.

[4] 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. – С. 3982-3992.

[5] Karpukhin, V. Dense Passage Retrieval for Open-Domain Question Answering / V. Karpukhin [et al.] // Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2020. – С. 6769-6781.

Авторский вклад

Крез Карина Сергеевна – научное руководство, проверка методологии, проведение литературного обзора, пересмотр и редактирование содержания статьи, утверждение окончательного варианта рукописи для публикации.

Шипуль Ангелина Робертовна, **Кривоносова Мария Анатольевна**, **Гугалев Андрей Сергеевич**, – анализ алгоритмических подходов, участие в написании отдельных разделов рукописи, оформление текста и списка литературы.

TRANSFORMER ARCHITECTURE: HOW IT TRANSFORMED THE APPROACH TO DEVELOPING TEXT PLAGIARISM DETECTION SYSTEMS

K.S. Krez
Assistant of the Department of Information and Computer Systems Design, BSUIR, Master of Technical Sciences

M.A. Krivonosova
3rd year student of the BSUIR Faculty of Computer Engineering, specializing in Information Systems and Technologies (in business management)

A.R. Shipul
3rd year student of the BSUIR Faculty of Computer Engineering, specializing in Information Systems and Technologies (in business management)

A.S. Gugalev
3rd year student of the BSUIR Faculty of Computer Engineering, specializing in Information Systems and Technologies (in business management)

Abstract. This article examines the impact of the Transformer architecture on the development of automated text plagiarism detection methods within modern anti-plagiarism systems. It demonstrates that the transition from lexico-statistical and string-based comparison methods to context-aware semantic analysis contributes to improved detection accuracy for paraphrased and partially modified instances of plagiarism. The limitations of traditional approaches – based on shingles, n-grams, and static embeddings – are analyzed, and the role of the self-attention mechanism in generating contextual representations of text fragments is elucidated. Particular attention is devoted to the application of bi-encoder and cross-encoder architectural strategies within a hybrid anti-plagiarism verification pipeline, designed to strike a balance between performance and accuracy. A simplified example is presented illustrating the semantic matching of document segments against a local repository and the subsequent calculation of a semantic similarity score. The article concludes that integrating Transformer-based models into multi-layered anti-plagiarism systems – which combine precise lexical algorithms with semantic analysis – is a highly advisable strategy.

Keywords: semantic similarity of texts, STS, Transformer, self-attention, BERT, SBERT, sentence embeddings, semantic search.

УДК 004.912:004.8

АЛГОРИТМЫ СРАВНЕНИЯ ТЕКСТОВОЙ ИНФОРМАЦИИ



К.С. Крез
Ассистент и аспирант кафедры проектирования информационно-компьютерных систем БГУИР
k.krez@bsuir.by



Е.Н. Шнейдеров
Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент, проректор по учебной работе
alexvikt.minsk@gmail.com



В.И. Голушко
Студент специальности информационные системы и технологии (в бизнес-менеджменте) кафедры проектирования информационно-компьютерных систем БГУИР
vadimgolushko2004@gmail.com

К.С. Крез
Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов: корреляционный анализ цифровых следов пользователей, нейронные сети, разработка и анализ структуры хранения данных.

Е.Н. Шнейдеров
Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с разработкой феноменологических моделей и научно-методических основ компьютерного проектирования радиоэлектронных средств, технического обеспечения безопасности и создания электронных систем безопасности.

В.И. Голушко
Является студентом Белорусского государственного университета информатики и радиоэлектроники по специальности «Информационные системы и технологии (в бизнес-менеджменте)», дневная форма обучения. Профессиональные интересы связаны с разработкой программного обеспечения и изучением различных направлений в сфере информационных технологий.

Аннотация. В статье рассматриваются алгоритмы сравнения текстовой информации, применяемые в задачах обработки естественного языка. Выполнен обзор подходов, ориентированных на анализ лексического, структурного и семантического сходства текстов, включая статистические методы, расстояние Левенштейна, алгоритмы локально-чувствительного хеширования и фонетические подходы. В статье рассмотрены архитектуры, принципы работы и области применения моделей Word2Vec, GloVe, FastText, BERT и Doc2Vec. На основании проведённого анализа установлено, что выбор оптимального алгоритма зависит от требований к точности, вычислительной эффективности и специфики обрабатываемых данных. Особо отмечается высокая эффективность модели BERT в задачах сравнения текстов.

Ключевые слова: алгоритмы сравнения текстов; BERT; TF-IDF; семантический анализ

Введение. Определение степени сходства между текстами представляет собой одну из ключевых задач в области автоматической обработки естественного языка. Эта задача возникает в различных контекстах, таких как подготовка рефератов, дипломных работ, отчетов, машинный перевод, установление авторства, выявление академического плагиата, и других, где требуется оценить уровень подобия между двумя текстовыми документами. В подобных сценариях требуется определение меры близости между двумя текстовыми документами на основе их лексических, структурных и семантических характеристик. Цель

статьи заключается в проведении обзора существующих алгоритмов сравнения текстовой информации и анализе принципов их работы. При анализе текстового сходства в современных алгоритмах обычно учитываются следующие группы признаков:

– формат файла текстового процессора Microsoft Word DOCX;

– взаимное расположение слов и устойчивых сочетаний, включая порядок следования и совместную встречаемость [1];

– синтаксические и семантические связи, существующие между словами [2];

– предметная область и тематическая направленность текстов [3].

Основная часть. Для сравнения текстов применяются различные методы, отличающиеся уровнем анализа и типом используемых признаков. К наиболее распространённым из них относятся модели векторного пространства, включая TF-IDF и методы распределённых представлений; редакционные расстояния, оценивающие количество операций, необходимых для преобразования одной последовательности в другую; методы локально-чувствительного хеширования, обеспечивающие эффективное сопоставление текстов с учётом их структурных характеристик; а также алгоритм Soundex, основанный на сопоставлении текстовых единиц по фонетическому сходству.

Особое распространение получили алгоритмы, использующие векторные представления слов, предложений и текстов. Их основная идея заключается в отображении текста в пространство признаков, в котором степень сходства между объектами определяется на основе расстояния или меры близости между соответствующими векторами [4]. При данном подходе компоненты векторного представления фиксируют существенные характеристики текста, включая частотные показатели употребления терминов: каждому слову ставится в соответствие отдельное измерение, а сами документы описываются как векторы в многомерном признаковом пространстве. При этом близкое расположение точек в векторном пространстве интерпретируется как свидетельство высокой степени сходства между текстами. Существенным различием между методами данного класса является выбор метрики, определяющей способ вычисления расстояния или близости между векторами.

Одним из ключевых подходов в рамках векторного представления текста является назначение каждому слову определённого веса, отражающего его значимость в документе и во всём тексте.

При данном подходе расчёт метрики опирается на предварительное определение информационного веса каждого слова в тексте. Наиболее известным методом, реализующим такой принцип, является Term Frequency – Inverse Document Frequency (далее – TF-IDF). В его основе лежат два показателя: Term frequency (далее – TF), отражающий частоту появления термина в конкретном тексте, и Inverse document frequency (далее – IDF), характеризующий обратную частоту его встречаемости в коллекции документов. Для более точного расчета весов значения TF и IDF перемножаются, что называется взвешиванием TF-IDF. Логика метода TF-IDF основана на том, что тексты, в которых определенное слово встречается с близкой частотой, считаются схожими [5]. Однако TF-IDF не учитывает контекст употребления слов [6]. Векторное представление текстов строится на основе гипотезы «мешка слов». Мера TF-IDF вычисляется как произведение частоты слова в тексте и обратной частоты слова во всём наборе документов и определяется по формуле 1.

Формула для определения показателя имеет следующий вид:

$$TF_IDF = TF \cdot IDF, \tag{1}$$

где TF – частота слова в конкретной категории / документе / коллекции (в зависимости от того, какие данные анализируются), IDF – обратная частота документа (популярность слова).

Частота слова в категории определяется по формуле 2:

$$TF = \frac{n_t}{\sum_{i=1}^{k} n_i}, \tag{2}$$

где n_t – количество отдельных слов в категории / документе / коллекции, $\sum_{i=1}^{k} n_i$ – общее количество всех слов в категории / документе / коллекции.

Обратная частота документа (также часто называют инверсией частоты) определяется по формуле 3:

$$IDF = \ln \frac{n_c}{\sum_{j=1}^{m} n_j}, \tag{3}$$

где n_c – количество категорий / документов / коллекций всего, $\sum_{j=1}^{m} n_j$ – количество категорий / документов / коллекций, в которых содержится интересующее слово.

TF-IDF не учитывает семантическую близость текстов в различных документах. Иными словами, при совпадении смыслового содержания, выраженного различными лексическими средствами, данный подход демонстрирует эффективность преимущественно на уровне лексико-статистических совпадений. Вместе с тем в задачах попарного сравнения документов, где существенное значение имеют точные формулировки и прямые текстовые совпадения, указанная особенность не всегда должна рассматриваться как недостаток метода. Дополнительно следует отметить, что TF-IDF не позволяет локализовать конкретные изменения и различия между сравниваемыми документами, а формирует лишь обобщённую количественную оценку степени их сходства.

Помимо TF-IDF в задачах анализа текстовых данных находят применение методы векторного представления, ориентированные на более полное моделирование семантических характеристик текста. К числу таких методов относятся Word2Vec, GloVe, FastText, BERT, Doc2Vec и Sense2Vec.

Word2Vec – модель распределённого представления слов, разработанная Google, предназначенная для построения векторных эмбеддингов слов. Получаемые векторы отражают семантические и контекстные связи между лексическими единицами, вследствие чего слова с близким значением имеют близкое расположение в векторном пространстве.

Global Vectors for Word Representation (далее – GloVe) представляет собой модель построения векторных представлений слов, основанную на анализе глобальной статистики их совместной встречаемости во фрагментах текстов. В отличие от Word2Vec, которая в большей степени ориентирована на использование локального контекста, модель GloVe объединяет преимущества статистического подхода и распределённого представления слов.

FastText является модификацией модели Word2Vec, разработанной компанией Facebook, в которой слово рассматривается как совокупность символьных n-грамм. Такой подход позволяет учитывать морфологические особенности слов и формировать векторные представления не только для слов, присутствующих в обучающем фрагменте, но и для ранее не встречавшихся лексических единиц.

Bidirectional Encoder Representations from Transformers (далее – BERT) представляет собой контекстно-зависимую языковую модель, основанную на архитектуре Transformer и разработанную компанией Google. За счёт двунаправленного механизма обучения модель способна учитывать как левый, так и правый контекст слова, что обеспечивает более полное выявление смысловых связей в тексте и высокую результативность при решении широкого круга задач обработки естественного языка.

Doc2Vec – расширение Word2Vec, предназначенное для построения векторных представлений документов, абзацев и предложений. Использование данного подхода позволяет выполнять сравнение текстовых фрагментов на уровне семантической близости, а не только по совпадению отдельных слов.

Sense2Vec – модель векторного представления, ориентированная на учёт многозначности слов. В рамках данного подхода для различных значений одной и той же лексемы формируются отдельные векторы на основе контекста её употребления, что способствует более точному моделированию семантики.

В обзоре [7] рассмотрены применения векторных представлений текстов.

Современное развитие данного подхода связано с несколькими направлениями совершенствования моделей представления текста.

Во-первых, осуществляется расширение векторных представлений слов путём интеграции дополнительных типов лингвистической информации, включая именованные сущности, морфосинтаксические характеристики и языковые модели.

Во-вторых, ведётся разработка методов, учитывающих многозначность слов и словосочетаний, что повышает точность семантической интерпретации.

В-третьих, наблюдается переход от дискретных векторных представлений к непрерывным вероятностным моделям, описываемым функциями распределения плотности. Наряду с этим разрабатываются специализированные представления, адаптированные к отдельным предметным областям, а также многоязычные модели, обеспечивающие обработку текстов на нескольких языках в рамках единого признакового пространства.

Метод редакционного расстояния, наиболее известной реализацией которого является расстояние Левенштейна, представляет собой способ количественной оценки различий между двумя символьными последовательностями. Данный метод основан на вычислении минимального числа элементарных преобразований, необходимых для преобразования одной последовательности в другую. В качестве таких преобразований рассматриваются вставка, удаление и замена символов.

К элементарным операциям в рамках данного подхода относятся:

– вставка/удаление символа;

– замена одного символа на другой.

Алгоритм производит подсчет операций, необходимых для преобразования одного набора символов (текста, документа) в другой. Данный метод назван в честь советского математика Владимира Левенштейна, который рассматривал это расстояние в 1965 году [8].

Расстояние Левенштейна применяется в следующих областях:

– сравнение текстовых данных – выявление различий между файлами и документами;

– коррекция ошибок – исправление опечаток в поисковых запросах, базах данных, системах ввода текста, а также при OCR и обработке речи;

– биоинформатика – анализ схожести генетических последовательностей (генов, хромосом, белков) [9].

К числу недостатков расстояния Левенштейна относится высокая чувствительность к перестановке слов, даже если она не изменяет смысл текста, а также зависимость значения метрики от длины сравниваемых единиц. В результате короткие, но несвязанные слова могут демонстрировать малое редакционное расстояние, тогда как длинные и частично совпадающие слова – большее, что снижает точность метода при оценке смысловой близости.

Локально-чувствительное хеширование (далее – TLSH) – уникальный набор символов, рассчитанный по определенному алгоритму и соответствующий определенному набору данных. Существует ряд алгоритмов (хеш-функций), используемых для расчета хеш-суммы: CRC32, SHA256, MD5. В случае, если файл подвергается изменению,

изменяется и хеш-сумма этого файла. Для полностью идентичных файлов хеш-сумма будет равной.

Для сравнения документов большого объёма, имеющих большое количество разделов, может применяться локально-чувствительное хеширование. Алгоритмы локально-чувствительного хеширования помещают схожие данные в одну корзину с высокой вероятностью [10]. Вместе с тем эффективность таких методов зависит от способа разбиения текста на признаки и выбора параметров хеширования поэтому два раздела могут быть определены как разные, если в одном из них есть пробел, а в другом его нет, даже если оба файла содержат одинаковый текст.

SoundEx – алгоритм, который позволяет сравнивать набор слов по их звучанию [11]. Изначально алгоритм был ориентирован на американский вариант английского языка, однако впоследствии появились его модификации для других языков, включая русский.

Иногда этот алгоритм не способен обнаружить сходство между очень близкими фамилиями: например, «Levinson» получит код L152, а «Lewinson» – код L525. Кроме того, Soundex плохо работает в ситуациях, когда произношение сильно расходится с написанием, что в английском языке бывает нередко. По этой причине его прямое применение к русскоязычным текстам сопровождается значительными погрешностями и обычно требует предварительной адаптации или замены более подходящими фонетическими алгоритмами.

1 Тип данных – категория текстовых данных, для обработки которых предназначен алгоритм. Выделяются следующие типы:

– короткие строки – текстовые последовательности малой длины, для которых важно точное посимвольное сравнение; такие методы эффективны на коротких данных, но менее пригодны для длинных текстов;

– тексты произвольной длины – слова, предложения и абзацы, обрабатываемые на основе токенов или n-грамм; данные методы универсальны по длине текста, однако ограниченно учитывают семантику;

– документы – объёмные структурированные тексты, анализ которых требует учёта не только лексических совпадений, но и смыслового содержания;

– слова и контексты – отдельные слова или их окружение в тексте, анализируемые на основе статистики совместной встречаемости;

– слова (онтологии) – слова, связанные через иерархии понятий (например, WordNet);

– длинные тексты – тексты из множества предложений, где сравнение требует агрегации результатов на уровне абзацев или предложений (а не символов / слов).

2 Учет контекста – свойство алгоритма учитывать окружение слова или фрагмента, но не обязательно его смысловую связь.

3 Чувствительность к опечаткам – степень, в которой алгоритм реагирует на ошибки в написании слов (опечатки, перестановки символов, лишние / пропущенные буквы). Чувствительность к опечаткам делится на:

– высокая (алгоритм реагирует на малейшие изменения символов);

– умеренная (частично игнорирует перестановки или замены);

– низкая (ориентирован на семантику, а не на символы).

4 Семантическое понимание – способность алгоритма учитывать сходство значений, а не только поверхностное совпадение символов или слов.

5 Внешние ресурсы – дополнительные данные или инструменты, требуемые для работы алгоритма.

Корпус – большой структурированный набор текстов, используемый для статистического анализа языка (частотность слов, контексты, семантика). WordNet – лексическая база данных английского языка, где слова связаны в иерархии (синонимы, гиперонимы, гипонимы).

Таблица 1. Сравнение алгоритмов

Алгоритм	Тип данных	Учет контекста	Чувстви-тельность к опечаткам	Семанти-ческое понимание	Внешние ресурсы	Примеры использования
Расстояние Левенштейна	Короткие строки	Нет	Высокая	Нет	Нет	Проверка орфографии (MS Word), системы исправления ошибок в базах данных
Jaro-Winkler	Короткие строки	Нет	Умеренная	Нет	Нет	Сопоставление записей в CRM-системах (Salesforce), идентификация пользователей.
N-граммы	Текст любой длины	Частично	Зависит от n	Нет	Нет	Поиск плагиата (Turnitin), кластеризация коротких текстов
LSA	Документы	Да	Низкая	Да	Требует	Тематическое моделирование (Gensim), рекомендательные системы (Amazon)
ESA	Высокая	Да	Низкая	Да	Корпус	Семантический поиск (IBM Watson), анализ научных статей
PMI	Слова/ контексты	Да	Низкая	Да	Требует	Определение коллокаций (Google N-gram Viewer), анализ тональности отзывов
Wu-Palmer	Слова (онтологии)	Нет	Высокая	Да	Корпус	Оценка эссе (ETS), классификация медицинских терминов (UMLS)
Lin	Слова (онтологии)	Нет	Средняя	Да	WordNet	Дизъюнкция слов в NLP-библиотеках (NLTK), анализ юридических документов
Resnik	Слова (онтологии)	Нет	Низкая	Да	WordNet + корпус	Семантическая кластеризация (Apache OpenNLP), поиск синонимов
SoftTFIDF	Текст	Частично	Умеренная	Нет	Нет (или корпус для TF-IDF)	Обработка заявок в банках (сопоставление клиентских данных), очистка дубликатов в CRM
Monge-Elkan	Длинные тексты	Нет	Низкая	Нет	Нет	Сравнение резюме с вакансиями (LinkedIn), анализ текстовых шаблонов
BERT	Текст (любой длины)	Да	Низкая	Да	Предобучен	Поисковые системы (Google), чат-боты (ChatGPT), анализ тональности в соцсетях (Twitter)

Заключение. Проведённый анализ показал, что алгоритмы сравнения текстовой информации существенно различаются по принципам работы, уровню представления текста и области практического применения. Такие алгоритмы как расстояние Левенштейна, SoundEx и методы локально-чувствительного хеширования, ориентированы преимущественно на анализ формальных характеристик текста и позволяют эффективно выявлять совпадения на уровне символов, слов или структурных элементов документа. Однако данные методы обладают ограниченной способностью учитывать семантические связи между словами и устойчивость к изменениям формулировок.

Список литературы

[1] Errecalde M., Ingaramo D., Rosso P. A new AntTree-based Algorithm for Clustering Short-text Corpora // Journal of Comput-er Science and Technology. – 2010. – V. 10. – № 1. – P. 1–7.

[2] Hayes R., Pisano G., Wheelwright S. Operations, Strategy, and Technical Knowledge. Hoboken, NJ: Wiley, 2007. Ferrґandez и др. Deep vs. Shallow Semantic Analysis Applied to Textual Entailment Recognition // Advances in Natural Language Pro-cessing 5th International Conference on NLP, FinTAL 2006 Turku, Finland, August 23-25, 2006 Proceedings

[3] Yue Wang, Hongsong Li, Haixun Wang, Kenny Q. Zhu. Toward Topic Search on the Web // JMIR Publications, 2012, P. 28. 20. Zampieri M., A

[4] Joaquin Perez-Iglesias – Integrating the Probabilistic Model BM25: BM25F into Lucene // Cornell University free distribu-tion platform[arXiv.org], 2009, P. 7.

[5] Christopher D. Manning, Prabhakar Raghavan & Hinrich Schütze. An Introduction to Information Retrieval // Cambridge Uni-versity Press, 2009, P. 19.

[6] Automatic Keyword Extraction from Individual Documents [Электронный ресурс]. – Режим доступа: https://www.researchgate.net/publication/227988510Automatic_Keyword_Extraction-dfrom_Individual_Docments.

[7] Харламов А. А., Гордеев Д. И. Дистрибутивная VS сетевая семантика в диалоговых системах // Проблемы искусственного интеллекта, 2019, №2 (13), С. 93. 6. Ч

[8] Левенштейн В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов // Докл. АН СССР, 1965, Т. 163, №4, С. 845.

[9] Shama Rani, Jaiteg Singh – Enhancing Levenshtein's Edit Distance Algorithm for Evaluating Document Similarity // Communi-cations in Computer and Information Science, 2017, P. 75.

[10] Loïc Paulevé, Hervé Jégou, Laurent Amsaleg – Locality sensitive hashing: A comparison of hash function types and querying mechanisms // HAL, 2011, P. 11.

[11] Wagner R.A., Fischer M.J. The string-to-string correction problem / Journal of the ACM, 1974. Vol. 21, № 1, pp. 168–173.

[12] Zeeshan Bhatti, Ahmad Waqas, Imdad Ali Ismaili, Dil Nawaz Hakro, Waseem Javaid Soomr // Phonetic based SoundEx & ShapeEx algorithm for Sindhi Spell Checker System, 2014, P. 7.

Авторский вклад

Крез Карина Сергеевна – постановка цели и задач исследования, проведение литературного обзора, анализ и систематизация алгоритмов сравнения текстовой информации, написание черновика рукописи, подготовка сравнительной таблицы, общая координация проекта.

Евгений Николаевич Шнейдеров – научное руководство, проверка методологии, пересмотр и редактирование содержания статьи, утверждение окончательного варианта рукописи для публикации.

Вадим Иванович Голушко – сбор и обработка данных, анализ алгоритмических подходов, участие в написании отдельных разделов рукописи, оформление текста и списка литературы.

ALGORITHMS FOR COMPARING TEXT INFORMATION

K.C. Krez
Assistant and graduate student in the Department of Information and Computer Systems Design at BSUIR

E.N. Shneiderov
Associate Professor of the Department of Design of Information and Computer Systems of BSU-IR, Candidate of Technical Sciences, Associate Professor, Vice-Rector for Academic Affairs

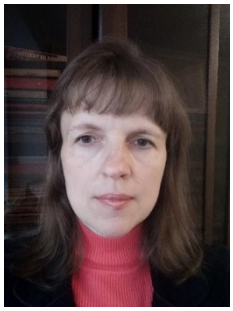
V.I. Golushko
Student of the information systems and technologies (in business management) specialty of the Department of Information and Computer Systems Design at BSUIR

Abstract. This article examines algorithms for comparing textual information used in natural language processing tasks. It provides an overview of approaches focused on analyzing the lexical, structural, and semantic similarity of texts, including statistical methods, Levenshtein distance, locality-sensitive hashing algorithms, and phonetic approaches. The article discusses the architectures, operating principles, and application areas of the Word2Vec, GloVe, FastText, BERT, and Doc2Vec models. Based on the conducted analysis, it is established that the selection of the optimal algorithm depends on requirements regarding accuracy, computational efficiency, and the specific characteristics of the data being processed. Particular emphasis is placed on the high effectiveness of the BERT model in text comparison tasks.

Keywords text comparison algorithms; BERT; TF-IDF; semantic analysis.

УДК 159.9.072

ПРИМЕНЕНИЕ СИСТЕМЫ ЭЛЕКТРОННОГО ОБУЧЕНИЯ ДЛЯ ИЗУЧЕНИЯ ПСИХОЛОГИЧЕСКИХ ХАРАКТЕРИСТИК СТУДЕНТОВ (НА ПРИМЕРЕ СКЛОННОСТИ К РИСКУ)



А.Ю. Яцкевич
Старший преподаватель кафедры инженерной психологии и эргономики БГУИР,
магистр педагогических наук
yatskevich@bsuir.by

А.Ю. Яцкевич

Окончила Белорусский государственный университет имени М. Танка по специальности «Практическая психология и иностранный язык». В 2001 году защитила магистерскую диссертацию на соискание степени магистра педагогических наук. Сфера научных интересов: психология безопасности, инженерно-психологическое проектирование информационных технологий.

Аннотация. В статье представлены результаты исследования склонности к риску студентов 2-4 курсов БГУИР в 2020-2025 гг. с использованием возможностей СЭО БГУИР для организации тестирования и сбора первичных эмпирических данных. Указаны рискованные действия с максимальной и минимальной вероятностью совершения студентами. Описана взаимосвязь готовности совершить различные опасные действия с уровнем общей склонности к риску.

Ключевые слова: склонность к риску, опасность, риск, психодиагностика.

Введение. Цифровизация системы образования –устойчивый современный тренд. На протяжении последних десятилетий информационные технологии активно используются для предъявления и объяснения нового материала, наглядной демонстрации примеров его применения, сбора и хранения результатов выполнения студентами практических заданий, контроля усвоения знаний и т.п. Рассматривая преимущества автоматизированной проверки практических работ студентов, в частности, тестирования посредством различных систем электронного обучения, традиционно указывают на рост возможности одновременного охвата большой аудитории, отсутствие необходимости готовить персональные раздаточные материалы, быстроту обработки данных и получения итогового результата. Однако при этом незаслуженно игнорируется еще одна возможность – более глубокий анализ ответов тестируемых, позволяющий выявить не только интегральный уровень того или иного показателя, но и получить подробную информации о его компонентах, выявить сильные и слабые стороны, достигая более глубокого понимания факторов, определяющих итоговый уровень изучаемого показателя. И эти возможности могут быть применимы как по отношению к компетенциям, так и личностным характеристикам студентов. Рассмотрим данный тезис на примере изучения склонности к риску.

Склонность к риску студентов БГУИР. Склонность к риску – это устойчивая личностная характеристика, связанная с наличием внутренней мотивации к выполнению

деятельности, характеризуемой значительной вероятностью опасности, в ситуации, когда выбор более рискованной стратегии поведения не связан с получением выгоды [1].

В 2010-2012 и 2020-2025 гг. в рамках образовательного процесса по дисциплине «Психология профессиональной деятельности» нами было проведено исследование склонности к риску студентов БГУИР специальности «Инженерно-психологическое обеспечение информационных технологий». Для диагностики применялся опросник оценки склонности к риску Г. Шуберта [2]. Ранее нами было доказано, что склонность к риску студентов 2020-2023 гг снизилась по сравнению с показателями студентов 2010-2012 гг. [1]. Использование СЭО Moodle в качестве средства организации тестирования в 2020-2025 гг. позволило собрать первичные эмпирические данные для дифференцированного анализа ответов студентов на отдельные вопросы методики Г. Шуберта и, таким образом, провести более глубокий анализ склонности к риску. Рассмотрим полученные результаты.

В соответствии с инструкцией к опроснику, студентам предлагалось оценить по 5-ти балльной шкале (от «+2» – полностью согласен до «–2» – совершенно не согласен) свою готовность выполнить определенные рискованные действия. Всего методика Г. Шуберта включает 25 таких действий [2]. В 2020-2025 гг. в исследовании приняли участие 284 студента. Среднее значение итогового показателя составило – 4,26 при, что соответствует среднему уровню склонности к риску. Нами также были проанализированы средние значения готовности выполнить каждое из указанных в опроснике рискованных действий. Результаты представлены в таблице 1.

Таблица 1. Средние значения готовности студентов совершить различные рискованные действия

№ во-проса	Средний балл	№ во-проса	Средний балл	№ во-проса	Средний балл	№ во-проса	Средний балл	№ во-проса	Средний балл
1	1,39	6	-0,86	11	-1,27	16	0,66	21	-0,42
2	-0,13	7	-0,53	12	0,11	17	0,84	22	0,25
3	-0,13	8	-0,07	13	1,04	18	-0,95	23	-0,87
4	-0,82	9	-0,71	14	-0,16	19	-0,01	24	-0,51
5	0,4	10	-0,67	15	0,19	20	-0,69	25	-0,34

Представим полученные данные графически.

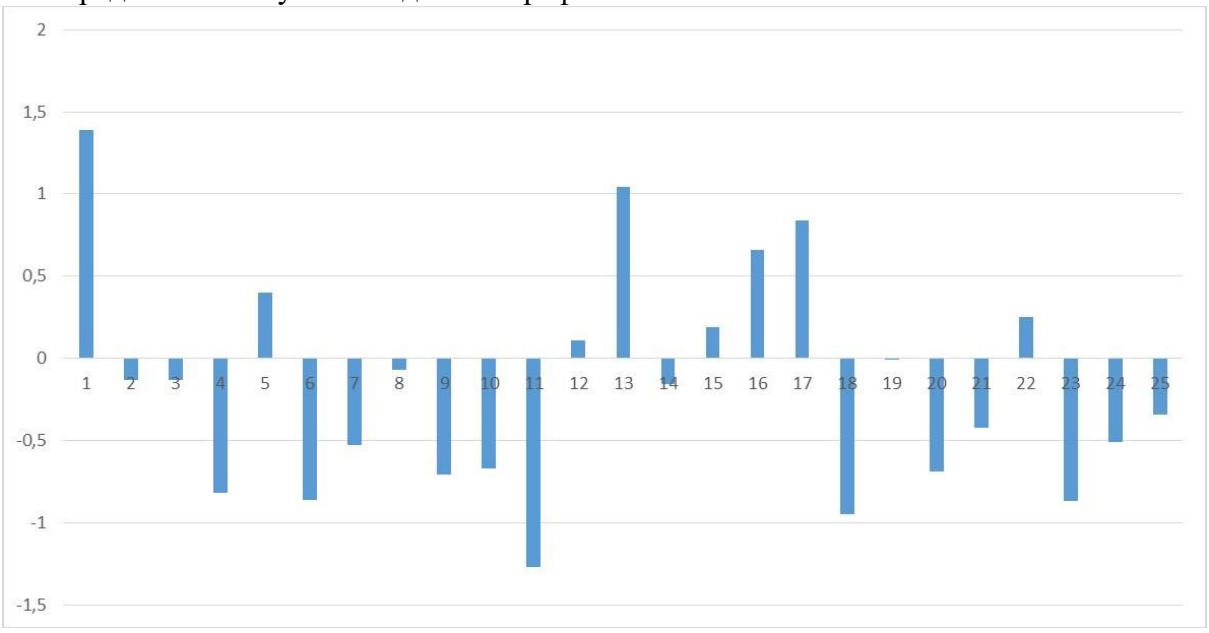


Рисунок 1. Средние значения готовности студентов совершить различные рискованные действия

В соответствии с полученными данными, студенты выражают большую готовность выполнить следующие рискованные действия:

№ 1 превысить установленную скорость, чтобы быстрее оказать необходимую помощь тяжелобольному человеку;

№ 13 совершить прыжок с парашютом;

№ 17 пойти на сложную, опасную для жизни операцию, чтобы избавиться от затяжной болезни с постельным режимом.

Высокие показатели по пунктам 1 и 17, вероятно, обусловлены влиянием эпидемии Covid, осознанием связанных с ней рисков, взаимосвязи индивидуальных и коллективных действий для сохранения физического здоровья.

Высокая готовность совершить прыжок с парашютом, возможно, объясняется доступностью и рекламой данного экстремального развлечения как не требующего дополнительной подготовки.

Наименьшая готовность отмечается по таким действиям как

№ 11 схватить за уздечку бегущую лошадь;

№ 18 прыгнуть с подножки товарного вагона, идущего со скоростью 50 км в час;

№ 23 по указанию начальника взяться за высоковольтный провод, если бы он заверил, что провод обесточен;

№ 6 первым переходить вброд незнакомую очень холодную реку;

№ 4 ехать на подножке товарного вагона при скорости поезда более 100 км в час.

Низкая готовность совершить эти рискованные действия предположительно обусловлены их непривычностью, отсутствием информации о них в личном опыте (прежде всего № 11, 6, 23) а также проводимыми в Беларуси рекламными кампаниями против экстремальных увлечений, в частности, зацепинга [3].

Нами также исследовалась взаимосвязь готовности совершать отдельные рискованные действие с интегральным показателем склонности к риску на основе вычисления коэффициентов ранговой корреляции Спирмена. Результаты представлены в таблице 2.

Таблица 2. Коэффициенты корреляции (КК) готовности совершить различные рискованные действия с интегральным показателем склонности к риску

№ вопроса	КК	№ вопроса	КК	№ вопроса	КК	№ вопроса	КК	№ вопроса	КК
1	0,401	6	0,466	11	0,326	16	0,414	21	0,560
2	0,595	7	0,243	12	0,513	17	0,219	22	0,451
3	0,387	8	0,487	13	0,499	18	0,400	23	0,363
4	0,572	9	0,591	14	0,428	19	0,439	24	0,328
5	0,349	10	0,434	15	0,547	20	0,425	25	0,422

Все пункты опросника статистически значимо коррелируют с итоговым результатом (вероятность ошибки $p \leq 0.01$), однако сила взаимосвязей различается.

Наиболее сильная взаимосвязь наблюдается между склонность к риску и готовностью студентов совершить следующие действия:

№ 2 участвовать в опасной и длительной экспедиции ради хорошего заработка;

№ 9 под руководством извне, по подсказкам со стороны залезть на высокую фабричную трубу;

№ 4 ехать на подножке товарного вагона при скорости поезда более 100 км в час;

№ 21 выполнять опасную для жизни работу, если за нее хорошо платят.

Наименьшая корреляция – с готовностью совершить действия

№ 7 одолжить вы другу большую сумму денег, будучи не совсем уверенным, что он сможет вернуть эти деньги;

№ 17 пойти на сложную, опасную для жизни операцию, чтобы избавиться от затяжной болезни с постельным режимом.

Таким образом, склонность к риску у студентов проявляется преимущественно в готовности подвергаться физической опасности, а готовность нести финансовый риск незначительно связана с рассматриваемым интегральным показателем.

Малая корреляция склонности к риску и готовности пойти на опасную операцию ради выздоровления предположительно обусловлена тем, что после эпидемии Covid сама по себе затяжная болезнь рассматривается как опасное событие и хирургическая операция для избавления от нее может расцениваться как средство снижения риска.

Заключение. Применение современных систем электронного обучения при проведении диагностики склонности к риску студентов БГУИР позволило не только определить интегральные значения исследуемой характеристики, но и выполнить более глубокий анализ ее структуры.

Полученные результаты о ранжировании видов рискованных действий по степени готовности студентов их совершить, о взаимосвязи отдельных опасных действий и общей склонности к риску могут быть использованы для разработки профилактических мероприятий, в частности социальных рекламных кампаний по повышению безопасности молодежи.

Список литературы

[1] Яцкевич, А. Ю. Сравнительный анализ склонности к риску студентов = Comparative analysis of students' risk-taking / А. Ю. Яцкевич // BIG DATA и анализ высокого уровня = BIG DATA and Advanced Analytics : сборник научных статей X Международной научно-практической конференции, Минск, 13 марта 2024 г. : в 2 ч. Ч. 1 / Белорусский государственный университет информатики и радиоэлектроники ; редкол.: В. А. Богуш [и др.]. – Минск, 2024. – С. 155–157.

[2] Практикум по дифференциальной психодиагностике профессиональной пригодности. Учебное пособие / Под общ. ред. В.А. Бодрова. – М.: ПЕР СЭ, 2003 – С. 581 - 584.

[3] БЖД с 1 по 12 сентября проведет акцию «Дети и безопасность» – URL: https://belta.by/society/view/bzhd-s-1-po-12-sentjabrja-provedet-aktsiju-deti-i-bezopasnost-734320-2025/ (дата обращения: 10.03.2026).

USING OF AN E-LEARNING SYSTEM TO STUDY STUDENTS' PSYCHOLOGICAL CHARACTERISTICS (ON THE EXAMPLE OF RISK-TAKING)

A.Y. Yatskevich

Senior lecturer at the Department of Engineering Psychology and Ergonomics of BSUIR, MA

Abstract. The article presents the results of a study of BSUIR students' risk-taking. The study was conducted in 2020-2025 using the capabilities of BSUIR e-learning system to organize testing and collect primary empirical data. Risky actions with the maximum and minimum probability of being committed by students are indicated. The interrelation of willingness to commit various risky actions with the level of students' general risk risk-taking is described.

Key words: risk-taking, danger, risk, psychodiagnostics.

УДК 37:004.6

АРХИТЕКТУРА ПОЗНАНИЯ: ТРАНСФОРМАЦИЯ ОБРАЗОВАТЕЛЬНЫХ СИСТЕМ В ЭПОХУ ИНТЕЛЛЕКТУАЛЬНОЙ АНАЛИТИКИ И БОЛЬШИХ ДАННЫХ



С.Т.Веллиева
Преподаватель кафедры «Теория и методика спорта, физкультуры» Туркменского государственного института физкультуры и спорта
sonashtw@gmail.com



Н.М.Реджепов
Преподаватель кафедры «Общие технические предмеды» института Телекоммуникаций и информатики Туркменистана



А.Гурбанов
Преподаватель кафедры «Общие технические предмеды» института Телекоммуникаций и информатики Туркменистана

С.Т.Веллиева
Окончил Туркменский государственный институт физкультуры и спорта. Работает в государственном институте физкультуры и спорта.

Н.М.Реджепов
Окончил Государственный сельскохозяйственный университет. Работает в институте Телекоммуникаций и информатики Туркменистана.

А.Гурбанов
Окончил Государственный сельскохозяйственный университет. Работает в институте Телекоммуникаций и информатики Туркменистана.

Аннотация. На протяжении веков образование строилось вокруг фигуры учителя и печатного станка. Это была эпоха стандартизации, продиктованная индустриальным веком: одна программа для всех, один темп освоения материала и единые критерии успеха. Однако сегодня мы становимся свидетелями заката этой модели. В мире, где каждое действие учащегося в цифровой среде оставляет след, образование превращается из массового производства в высокое искусство персонализации. Большая аналитика (Big Data) и продвинутые методы обработки данных (Advanced Analytics) становятся теми инструментами, которые позволяют увидеть не группу студентов, а каждого отдельного человека с его уникальным ритмом, талантами и трудностями.

Ключевые слова: Big Data, AI, повышение эффективности, Google A, IBM Watson.

Введение. Цифровизация образования – это не просто замена бумажного учебника планшетом. Это создание непрерывного потока данных о процессе мышления.

Когда студент работает в электронной среде, система фиксирует не только конечный ответ, но и путь к нему: сколько раз он перечитывал условие задачи, на каком этапе возникла пауза, к каким справочным материалам он обратился. Эти данные формируют так называемый «цифровой след».

В классической педагогике эти нюансы были невидимы для учителя, особенно в больших аудиториях. Теперь же, благодаря Advanced Analytics, мы получаем возможность проводить «рентгенографию» процесса обучения в реальном времени. Мы видим не просто результат теста, а динамику когнитивных усилий.

Это позволяет переосмыслить само понятие «оценки»: вместо статичного балла в конце четверти мы получаем живой график развития компетенций, где важен не факт ошибки, а алгоритм её преодоления.

Одной из центральных идей новой образовательной парадигмы является адаптивное обучение.

Представьте себе учебник, который меняется в зависимости от того, кто его читает. Если система видит, что учащийся быстрее воспринимает визуальную информацию, она предлагает видеоролик или инфографику. Если же обнаруживается пробел в базовых знаниях, алгоритм бесшовно встраивает короткий повторительный модуль, прежде чем двигаться дальше.

Основная часть. Оптимизация образовательного процесса здесь достигает своего пика: время учащегося больше не тратится на повторение того, что он уже знает, и он не оказывается в ситуации когнитивного перегруза, когда материал слишком сложен. Advanced Analytics выстраивает «зону ближайшего развития» для каждого, делая обучение максимально эффективным и менее стрессовым.

Это превращает образование из «полосы препятствий» в «индивидуальный маршрут восхождения». Одной из самых ценных функций больших данных в образовании является их способность предсказывать будущее. Анализируя массивы данных прошлых лет, нейронные сети выявляют тонкие паттерны поведения, которые предшествуют потере интереса или полному отсеву студента. Это может быть изменение времени входа в систему, специфические ошибки в тестах или снижение активности на форумах. Предиктивная аналитика позволяет образовательным учреждениям действовать на опережение.

Система подает сигнал куратору или преподавателю: «этому студенту нужна поддержка». Важно понимать, что это не «приговор» машины, а инструмент для своевременного человеческого вмешательства. Оптимизация здесь заключается в том, что ресурсы поддержки направляются именно туда, где они критически необходимы в данный момент, спасая потенциал тех, кто мог бы просто исчезнуть из системы.

Существует опасение, что алгоритмы вытеснят живого педагога. Однако истинная цель аналитики – вернуть учителю его истинное призвание. Освобождая педагога от рутинной проверки работ и механического сбора статистики, Big Data дает ему нечто более ценное – объективную картину происходящего в классе.

На утреннем дашборде учитель видит «тепловую карту» знаний своих учеников. Он знает, что 70% класса не поняли тему вчерашней лекции, и может изменить план сегодняшнего занятия, не тратя время на опрос. Это превращает преподавание в доказательную практику.

Учитель становится не просто источником информации, а наставником и аналитиком, который управляет развитием каждого ученика, опираясь на твердую почву данных, а не на субъективные ощущения. На уровне управления университетом или школой использование Advanced Analytics позволяет оптимизировать саму структуру образовательной организации. Анализ востребованности курсов, эффективности различных методик преподавания и даже загруженности аудиторий помогает принимать взвешенные стратегические решения. Данные позволяют видеть связь между учебным планом и успехами выпускников на рынке труда. Анализируя запросы работодателей и сопоставляя их с результатами обучения, вузы могут оперативно обновлять программы, сокращая разрыв между теорией и практикой.

Образование становится гибким и отзывчивым к изменениям внешнего мира, что является ключевым фактором успеха в быстро меняющейся экономике знаний. Сбор и анализ колоссальных объемов личных данных в образовании ставит перед нами острые этические вопросы.

Право на приватность и защиту цифрового профиля учащегося должно быть незыблемым. Но есть и другой аспект: право на ошибку. Если система «помнит» каждый промах ученика в начальной школе, не станет ли это клеймом на его будущем? Философия использования Big Data в образовании должна строиться на принципе «поддержки, а не контроля». Данные должны использоваться для расширения возможностей, а не для создания цифровых барьеров.

Развитие аналитических систем должно идти рука об руку с развитием законодательства и этических кодексов, гарантирующих, что искусственный интеллект в образовании будет служить гуманистическим целям.

Заключение: Образование как путешествие длиною в жизнь. Мы стоим на пороге создания глобальной экосистемы обучения, где Big Data и Advanced Analytics связывают воедино формальное образование, саморазвитие и профессиональный рост.

Это мир, где обучение не заканчивается с получением диплома, а продолжается всю жизнь, поддерживаемое интеллектуальными помощниками, которые знают наши сильные стороны и зоны роста.

Оптимизация образования через данные – это не поиск самого дешевого пути обучения, а поиск самого эффективного пути раскрытия человеческого потенциала.

В конечном счете, главная победа больших данных в образовании будет заключаться в том, что каждый человек, вне зависимости от обстоятельств, получит доступ к знаниям, представленным в той форме и в тот момент, когда они ему нужнее всего.

Список литературы

[1] Чарыев А.Б., Шайымов С.С., Аннабердиев И.А., «Важность использования BIG DATA для оптимизации управления роботами-манипуляторами». Международный научный журнал «ИННОВАЦИОННАЯ НАУКА» №2-1/2025.

[2] Шайымов С.С., Гочов А.С., Веллиева С.Т., «Значение BIG DATA в цифровую эпоху». Международный научный журнал «ИННОВАЦИОННАЯ НАУКА» №2-2/2025.

Авторский вклад

Авторы внести равнозначный вклад.

ARCHITECTURE OF COGNITION: TRANSFORMING EDUCATIONAL SYSTEMS IN THE ERA OF INTELLIGENCE AND BIG DATA

S.T.Vellieva
Lecturer at the Department of Theory and Methodology of Sports, Physical Education, Turkmen State Institute of Physical Education and Sports

N.M.Rejepov
Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan

A. Gurbanov
Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan

Abstract. For centuries, education was built around the teacher and the printing press. This was the era of standardization, dictated by the industrial age: one curriculum for everyone, one pace of learning, and uniform criteria for success. However, today we are witnessing the decline of this model. In a world where every student's action in the digital environment leaves a trace, education is evolving from mass production into the fine art of personalization. Big Data and advanced data processing methods are becoming tools that allow us to see not just a group of students, but each individual with their unique rhythm, talents, and challenges.

Keywords: Big Data, AI, efficiency gains, Google A, IBM Watson.

УДК 004.738;615.47

ЦИФРОВЫЕ ДВОЙНИКИ НА НЕЙРОННЫХ СЕТЯХ В ЭЛЕКТРОННОЙ ТЕРАПИИ ПАЦИЕНТОВ С НЕВРОЛОГИЧЕСКИМИ ЗАБОЛЕВАНИЯМИ



В.А. Вишняков
Доктор технических наук, профессор кафедры инфокоммуникационных технологий БГУИР
vish@bsuir.by

В.А. Вишняков
Область профессиональных интересов и направление исследований: информационное управление и безопасность в инфокоммуникациях, Интернет вещей, блокчейн, электронный маркетинг, ИТ в образовании и медицине, интеллектуальные системы управления.

Аннотация. В докладе рассматривается технология цифровых двойников, основанных на искусственном интеллекте, для персонализированного ведения пациентов с нейродегенеративными заболеваниями (Альцгеймера и Паркинсона). Цифровой двойник представляет собой динамическую виртуальную модель индивидуальных когнитивных и физиологических систем пациента, создаваемую с использованием мультимодальных данных и нейросетевых алгоритмов. Анализируются архитектура таких систем, ключевые типы данных для их построения, принципы работы, лежащих в их основе нейронных сетей, а также перспективы их интеграции в клиническую практику для ранней диагностики, прогнозирования течения заболевания и подбора оптимальной терапии. Представлена структура и описание такого цифрового двойника.

Ключевые слова: цифровой двойник, болезни Альцгеймера**,** Паркинсона, нейронные сети, искусственный интеллект, персонализированная медицина, электронная терапия.

Введение. Глобальное бремя нейродегенеративных заболеваний, таких как болезнь Альцгеймера и болезнь Паркинсона, продолжает расти, создавая значительное давление на системы здравоохранения по всему миру. Традиционные подходы к диагностике и лечению часто носят реактивный характер и основаны на популяционных усредненных данных, что не учитывает уникальные особенности каждого пациента. В этой связи возникает острая необходимость в переходе к предиктивной и персонализированной медицине.

Технология цифровых двойников (Digital Twins), пришедшая в здравоохранение из промышленности, открывает новые горизонты для управления здоровьем [1]. Цифровой двойник пациента – это не статичная модель, а динамическая, постоянно обновляемая виртуальная копия, отражающая текущее состояние его организма на основе потока данных из различных источников [2]. В контексте нейродегенеративных заболеваний эта технология, усиленная возможностями искусственного интеллекта и нейронных сетей, позволяет создавать комплексные модели, способные симулировать течение болезни и реакцию на терапию в виртуальной среде, до применения вмешательств к реальному пациенту [2, 3]. Цель доклада – проанализировать архитектуру, принципы функционирования и перспективы внедрения цифровых двойников на нейронных сетях для электронной терапии пациентов с болезнями Альцгеймера и Паркинсона.

Модель нейронных сетей в основе цифрового двойника. Нейронные сети являются вычислительным ядром, обеспечивающим способность цифрового двойника к

обучению, прогнозированию и адаптации. По своей сути, искусственная нейронная сеть – это модель, приближенная к сетям нейронов головного мозга, способная к обучению на данных [1]. В контексте диагностики и моделирования заболеваний используются следующие подходы. Обработка мультимодальных данных: ключевым преимуществом является интеграция разнородных данных [4]. Нейросети обучаются на совокупности нейровизуализации, генетической информации, цифровых биомаркеров (например, траектория движения глаз, голос, походка) и клинических шкал. Исследования показывают, что такие мультимодальные подходы существенно превосходят по точности оценку, основанную на одном типе данных [2].

Ученые Сеченовского университета разработали нейросеть, которая по данным электроэнцефалограммы (ЭЭГ) выявляет болезнь Паркинсона на ранней стадии с точностью до 97 % [5]. Другие разработки, такие как израильский стартап Neuralight, используют ИИ для анализа около 100 параметров движения глаз, включая скорость расширения зрачка, что сильно коррелирует с болезнью Паркинсона и позволяет оценивать эффективность терапии [6].

Структура данных пациента. Построение точного цифрового двойника требует интеграции данных из множества источников. Эти данные можно структурировать по следующим модулям (таблица 1).

Таблица 1. Модули и типы данных для построения цифрового двойника

Модуль данных	Типы данных	Значение для модели
Нейровизуализация и физиология	МРТ, фМРТ, ЭЭГ, МЭГ, ПЭТ, диффузионная томография (для построения коннектома)	Оценка структурных и функциональных изменений мозга, выявление областей атрофии, нарушений нейронных связей [2, 3].
Генетика и биомаркеры	Геномные данные, биомаркеры крови (например, бета-амилоид, тау-белок), гистологические данные	Оценка генетической предрасполо-женности, подтверждение патологии на молекулярном уровне [2, 7]
Цифровые биомаркеры	Данные с носимых устройств (активность, сон), характеристики голоса, моторики (тремор), походки, когнитивные тесты в приложении, траектория движения глаз [2, 6].	Непрерывный мониторинг в реальных условиях, выявление subtle changes, объективная оценка эффективности лечения.
Клинические и анамнестические данные	История болезни, демография, назначенная лекарственная терапия, субъективные жалобы, психометрические шкалы (MMSE, UPDRS)	Контекстуализация объективных данных, учет клинического состояния и проводимого лечения [1].

Структура цифрового двойника. Архитектура цифрового двойника представляет собой многоуровневую систему, где данные пациента питают вычислительные модели, которые, генерируют виртуальные симуляции для поддержки клинических решений.

Виртуальный двойник мозга (Virtual Brain Twin): Как отмечается в проекте EBRAINS, такая модель создается путем комбинации структурных данных мозга (МРТ) с функциональными данными (ЭЭГ, фМРТ) для построения вычислительной модели нейронной сети пациента [3]. Эта модель симулирует, как различные регионы мозга соединены и взаимодействуют через нейронные пути (коннектом). Для эпилепсии и рассеянного склероза такие модели уже помогают планировать хирургические вмешательства и реабилитацию, и их применение для болезни Паркинсона и Альцгеймера является логическим продолжением [3].

Динамическая симуляция и сценарное моделирование: ядро двойника – это способность запускать симуляции «что-если». Врач может виртуально «назначить» определенный препарат или процедуру (нейростимуляцию) и наблюдать в модели предсказанный отклик систем организма – изменение паттернов ЭЭГ, динамику двигательных функций или когнитивных показателей [1, 3]. Это позволяет подобрать оптимальную терапию, минимизируя риски неэффективности или побочных эффектов у реального пациента.

Интерфейс взаимодействия: для практического использования цифровой двойник должен быть интегрирован в удобный для врача интерфейс, который предоставляет визуализацию симуляций, аналитические отчеты и рекомендации, сформированные ИИ [7].

Таблица 2. Уровни архитектуры цифрового двойника и их функции

Уровень архитектуры	Основные функции	Технологии
Уровень данных	Сбор, хранение, верификация и стандартизация мультимодальных данных пациента в режиме, близком к реальному времени.	Облачные хранилища, IoT-платформы, сенсоры.
Модельный уровень	Создание и обновление вычислительных моделей органов и систем (мозг, сердечно-сосудистая система). Нейросетевые алгоритмы прогнозируют динамику заболевания.	Глубокое обучение, байесовский вывод, механистическое моделирование [2, 3].
Сервисный уровень	Предоставление врачу инструментов для сценарного анализа, визуализации данных и результатов симуляций, формирования терапевтических рекомендаций.	Веб-интерфейсы, VR/AR для хирургического планирования [1, 7].

Программирование и техническая реализация. Разработка системы цифрового двойника – это междисциплинарная задача, требующая команду врачей, инженеров и программистов [1]. Первый этап – создание инфраструктуры для сбора и предобработки (очистка, аугментация, разметка) больших объемов данных. Для обучения медицинских нейросетей необходимы большие размеченные датасеты, сбор которых является нетривиальной задачей [8].

Обучение моделей: процесс обучения нейронной сети включает подачу на входной слой данных (например, голос, движение), их обработку через скрытые слои (где каждый нейрон имеет свои «веса» – коэффициенты значимости) и получение выходного результата (например, вероятности наличия заболевания) [9]. Модель последовательно проходит множество «эпох» обучения, постепенно минимизируя ошибку прогноза [8]. Для задач цифрового двойника часто используются гибридные подходы, сочетающие глубокое обучение для автоматического извлечения признаков и более интерпретируемые методы машинного обучения (например, случайные леса) там, где это необходимо клиницистам [2].

Валидация и интеграция: обученная модель должна пройти строгую валидацию на независимых наборах данных. Важно отметить, что заявления о высокой точности (85–95 %), полученные на небольших однородных когортах, могут не воспроизводиться в реальных клинических условиях, где их точность, вероятно, окажется на 10–15% ниже [2]. После успешной валидации модель интегрируется в клинические рабочие процессы и системы (например, EHR – Electronic Health Records).

Блок электронной терапии (модули пациента и врача, принятие решений, цифровой двойник на НС), который дополняет диагностическую систему динамической поддержкой терапевтических решений. Блок реализован в виде гибридной системы, которая сочетает в себе: рекуррентную нейронную сеть на основе GRU (изучает временные и последовательные взаимосвязи при прогрессировании симптомов), компонент семантической памяти системы OSTIS, который хранит варианты решений, реакции на лечение и контекстно-зависимые правила. Эти два компонента

соединены через интерфейсный уровень, который позволяет нейронной модели считывать данные из семантического графа знаний и записывать их в него, тем самым постоянно обогащая цифрового двойника как обучением, основанным на данных, так и знаниями, полученными под руководством человека. Представленная структура описывает поток данных взаимодействия врача с пациентом. Поток записывается в цифровом двойнике в режиме реального времени, происходит поиск и обработка временных данных, моделирование прогнозов развития (БА, БП), формирование новых данных для лечения с поддержкой врача. Цифровой двойник работает в трех шаговом режиме, обеспечивая замкнутый цикл обратной связи для интеллектуальной терапевтической поддержки.

1. Поиск предыдущего состояния и синхронизация данных. При каждой регистрации в клинике или при получении данных цифровой двойник получает доступ к предыдущим записям пациента из Sc-памяти. Это включает в себя: оценку речи и двигательных функций с отметками времени; предыдущие процедуры (лекарства, дозировки, планы физиотерапевтических процедур); прошлые предсказания модели и одобренные врачом вмешательства.

2. Прогностическое моделирование. Используя последние наблюдения в качестве входных данных, модель, основанная на GRU, моделирует: ожидаемую динамику симптомов в течение следующих 1–4 недель; прогнозируемую реакцию пациента на различные сценарии лечения (коррекция медикаментозного лечения, введение логопеда); оценка вероятности двигательных или когнитивных нарушений.

3. Выходные данные и семантическое объяснение. Лучшие рекомендации по лечению представляются через семантический веб-интерфейс OSTIS с сопроводительными пояснениями: какие симптомы, что ожидается или улучшатся; какие предыдущие решения повлияли на рекомендации; как модель оценивала потенциальную пользу.

Заключение. В докладе рассмотрена технология цифровых двойников на нейронных сетях, которая представляет собой новый подход в поддержке терапии больных с болезнями Альцгеймера и Паркинсона. Технология осуществляет переход от реактивной, обобщенной медицины к проактивной и персонализированной. Представлен блок электронной терапии для неврологических заболеваний (модули пациента и врача, принятие решений, цифровой двойник на НС), который дополняет диагностическую систему динамической поддержкой терапевтических решений. Блок реализован в виде гибридной системы, которая включает: рекуррентную нейронную сеть на основе GRU (временные и последовательные взаимосвязи при прогрессировании симптомов), компонент семантической памяти системы OSTIS, который хранит варианты решений, реакции на лечение и контекстно-зависимые правила.

Список литературы

[1] Виртуальные клоны: как цифровые двойники пациентов помогают врачам // Scientific Russia. 2023 [Электронный ресурс]. – Код доступа :https://scientificrussia.ru/articles/virtualnye-klony-kak-cifrovye-dvojniki-pacientov-pomogaut-vracam. – Дата доступа : 7.11.2025.

[2] Gkintoni, E., Halkiopoulos C. Digital Twin Cognition: AI-Biomarker Integration in Biomimetic Neuropsychological Frameworks // Biomimetics. – 2025. 10 (10): 640.

[3] Hashemi, M., Depannemaecker D., Saggio M., Triebkorn P., Rabuffo G., Fousek J. Principles and Operation of Virtual Brain Twins. IEEE Reviews in Biomedical Engineering. 2025, № 4:1-29.

[4] Вишняков В.А., Ся И. Распознавание признаков болезни Паркинсона на основе анализа голосовых маркеров и двигательной активности. Информатика. – 2023. − Т. 20, № 3. – С. 106–114.

[5] В Сеченовском университете разработали нейросеть для выявления болезни Паркинсона по данным ЭЭГ-исследования. 2025 [Электронный ресурс]. – Код доступа : https://www.sechenov.ru/pressroom/news/v-sechenovskom-universitete-razrabotali-neyroset-dlya-vyyavleniya-bolezni-parkinsona-po-dannym-eeg-i/. – Дата досту-па : 7.11.2025.

[6] Болезнь Паркинсона: как нейросети помогают врачам поставить правильный диагноз, 2022 [Электронный ресурс]. – Код доступа : https://sbermed.ai/neyroseti-pomogayut-vracham-postavit-pravilnyy-diagnoz. – Дата доступа : 7.11.2025.

[7] Виртуальные органы и испытание лекарств: цифровые двойники в медицине // РБК Тренды. – 2021 [Электронный ресурс]. – Код доступа : https://trends.rbc.ru/trends/industry/613b35369a7947506473665e. – Дата доступа : 7.11.2025.

[8] Нейросети в медицине: что это и как работает // Celsus. – 2020. [Электронный ресурс]. – Код доступа : https://celsus.ai/blog/kak-rabotaet-mediczinskaya-nejroset/. – Дата доступа : 7.11.2025.

[9] Вишняков В. А. Модель и структура сети Интернета вещей для ИТ-диагностики пациентов / В. А. Вишняков, Юй Чуюэ // Доклады БГУИР. – 2024. – Т. 22, № 5. – С. 104–112.

DIGITAL TWINS ON NEURAL NETWORKS IN ELECTRONIC THERAPY OF PATIENTS WITH ALZHEIMER'S AND PARKINSON'S DISEASES

U.A. Vishniakou
Doctor of Technical Sciences, Professor of the Department of Infocommunication Technologies, BSUIR
vish@bsuir.by

Abstract. The report examines the potential of artificial intelligence-based digital twin technology for the personalized management of patients with neurodegenerative diseases (Alzheimer and Parkinson). The digital twin is a dynamic virtual model of the patient's individual cognitive and physiological systems, created using multimodal data and neural network algorithms. The architecture of such systems, the key types of data for their construction, the principles of the underlying neural networks, as well as the prospects for their integration into clinical practice for early diagnosis, prediction of the course of the disease and the selection of optimal therapy are analyzed.

Keywords: digital twin, Alzheimer's disease, Parkinson's disease, neural networks, artificial intelligence, personalized medicine, electronic therapy.

УДК 004.8:004.4

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОГО МОБИЛЬНОГО ПРИЛОЖЕНИЯ ДЛЯ ИЗУЧЕНИЯ ИНОСТРАННЫХ ЯЗЫКОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ API



Д.А. Молдасейтов
Студент 4 курса института бизнеса и цифровых технологий, НАО «Казахский агротехнический исследовательский университет имени С.Сейфуллина»
diasmoldasseitov@gmail.com



Б.Е. Таныкпаева
Старший преподаватель, магистр технических наук, института бизнеса и цифровых технологий, НАО «Казахский агротехнический исследовательский университет имени С.Сейфуллина»
balausa1.80@mail.ru

Д.А. Молдасейтов

Студент выпускного курса по направлению информационных технологий НАО «КАТИУ им. С. Сейфуллина». Область научных интересов связана с мобильной разработкой под Android, проектированием архитектуры программного обеспечения (MVVM) и внедрением технологий искусственного интеллекта в образовательные продукты.

Б.Е. Таныкпаева

Старший преподаватель, магистр технических наук НАО «КАТИУ им. С. Сейфуллина. Научные интересы сосредоточены в области разработки мобильных приложений для платформы Android, проектирования архитектуры программного обеспечения с использованием паттерна MVVM, а также интеграции технологий искусственного интеллекта в образовательные системы.

Аннотация. В статье рассматривается процесс разработки нативного мобильного приложения для платформы Android, предназначенного для эффективного изучения иностранных слов.

Ключевые слова: мобильное приложение, искусственный интеллект, Android, архитектура MVVM, машинный перевод, интервальное повторение, образовательные технологии, DeepL API.

Введение. Освоение новой лингвистической системы исторически сопряжено с высоким когнитивным трением: острый дефицит времени, быстрое угасание мотивации и потребность в строго индивидуальном ритме обучения всегда оставались главными барьерами для студентов. Сегодня интеграция нейросетевых алгоритмов способна радикально перестроить этот процесс, переведя его из плоскости механической зубрежки в формат интерактивного диалога с машиной. Государственный вектор, обозначенный в Послании Президента РК от 8 сентября 2025 года [1], прямо указывает на необходимость глубокой цифровизации ключевых сфер общества. Разработанный интеллектуальный продукт выступает прикладным ответом на этот стратегический запрос, иллюстрируя, как машинное обучение переходит из категории абстрактных концепций в статус повседневного образовательного ассистента для каждого жителя Казахстана [2, 4]. Архитектура представленного мобильного решения базируется на синергии продуманного пользовательского опыта (UX) и вычислительной мощности сторонних языковых моделей. Глобальная задача проекта – демистифицировать сложные IT-технологии, упаковав продвинутый контекстный перевод в интуитивно понятный и доступный инструмент.

Фундаментальная архитектура и эволюция машинного обучения. Современный этап развития интеллектуальных систем характеризуется окончательным отказом от жестко детерминированных алгоритмов. Фундаментальная идея машинного обучения (ML) кроется в способности системы выявлять скрытые закономерности без написания прямых инструкций для каждого сценария. Если классическое программирование требовало описать объект через сотни условий («наличие острых ушей и усов классифицирует объект как кошку»), то алгоритмы машинного зрения и обработки текстов обучаются вероятностно. Опираясь на подходы, заложенные пионерами нейросетевых исследований, современный ИИ анализирует колоссальные массивы размеченных данных. Информация проходит через каскад скрытых слоев: начальные узлы сети распознают лишь базовые примитивы, но по мере продвижения вглубь архитектуры уровень абстракции возрастает, и на выходе формируется целостный семантический образ (рисунок 1) [10].

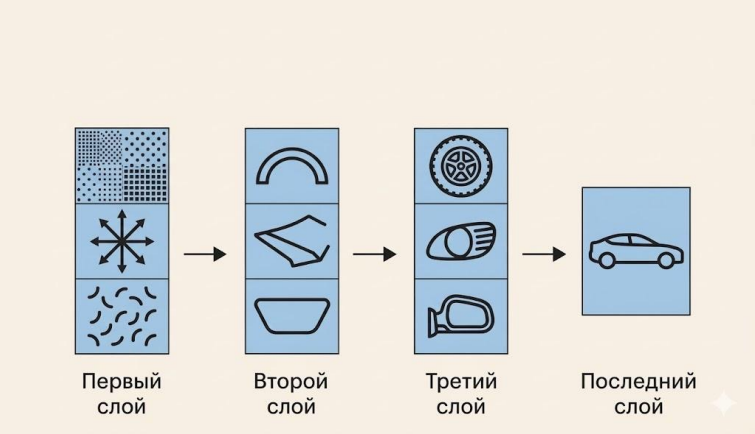


Рисунок 1. Иерархическое обучение признаков в глубокой нейронной сети

Долгое время непреодолимой преградой для алгоритмов оставалась естественная человеческая речь, изобилующая омонимами, сарказмом и культурными отсылками. Классические переводчики спотыкались о многозначность: фразы «замок на двери» и «рыцарский замок» интерпретировались ошибочно вне контекста. Революционным сдвигом стало появление архитектуры «Трансформер». Ее главное новшество заключается во внедрении механизма внутреннего внимания (Self-Attention). Прежде чем сгенерировать токен перевода, модель математически взвешивает ценность каждого слова относительно других элементов фразы. Алгоритм улавливает семантическую связь между словами «король», «трон» и «замок», безошибочно выдавая перевод «castle», а не «lock», что наглядно иллюстрирует схема работы механизма внимания (рисунок 2).

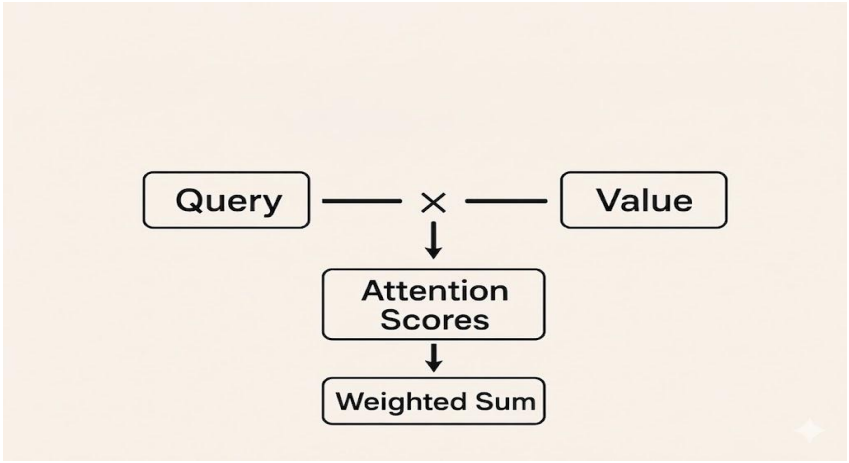


Рисунок 2. Принцип работы механизма внимания в архитектуре Трансформер

Интеграция сторонних вычислений и специфика API. На базе подобных сверхсложных трансформеров функционирует ядро платформы DeepL [5]. Обучение столь массивных языковых моделей требует серверных кластеров индустриального масштаба и петабайтов текстовых корпусов. В результате нейросеть начинает «чувствовать» стилистику, грамматику и культурный код десятков языков, генерируя текст, неотличимый от работы живого носителя.

Попытка воспроизвести подобную инфраструктуру в рамках локального продукта технически нецелесообразна. На практике это противоречие элегантно решается через API-шлюзы. Мобильный клиент не занимается тяжелыми вычислениями, а лишь формирует структурированный запрос к серверам DeepL. Подобный инжиниринговый подход позволяет сфокусироваться на образовательных механиках, доверяя ресурсоемкую аналитику облачным дата-центрам. Как показывает индустриальная аналитика, интеграция готовых ML-интерфейсов сокращает время вывода цифрового продукта на рынок (Time-to-Market) в среднем на 40–50%, что особенно актуально для темпов цифровизации Казахстана.

Проектирование платформы и клиент-серверное взаимодействие. Выбор ниши образовательных технологий (EdTech) для апробации прототипа обусловлен взрывным ростом спроса на микрообучение. Разработанный программный продукт «Сөздік» представляет собой нативное Android-приложение, написанное на языках Kotlin и Java [6, 7]. Использование архитектурного паттерна MVVM (Model-View-ViewModel) гарантирует, что бизнес-логика жестко отделена от пользовательского интерфейса.

Отказ от статичных предустановленных словарей стал принципиальным архитектурным решением. Платформа предлагает полностью персонализированную, динамическую среду. Главный экран функционирует как центральный хаб (рисунок 3):

здесь агрегируются все сохраненные лексические паттерны с возможностью мгновенной фильтрации, надежно закешированные в памяти устройства через библиотеку Room.

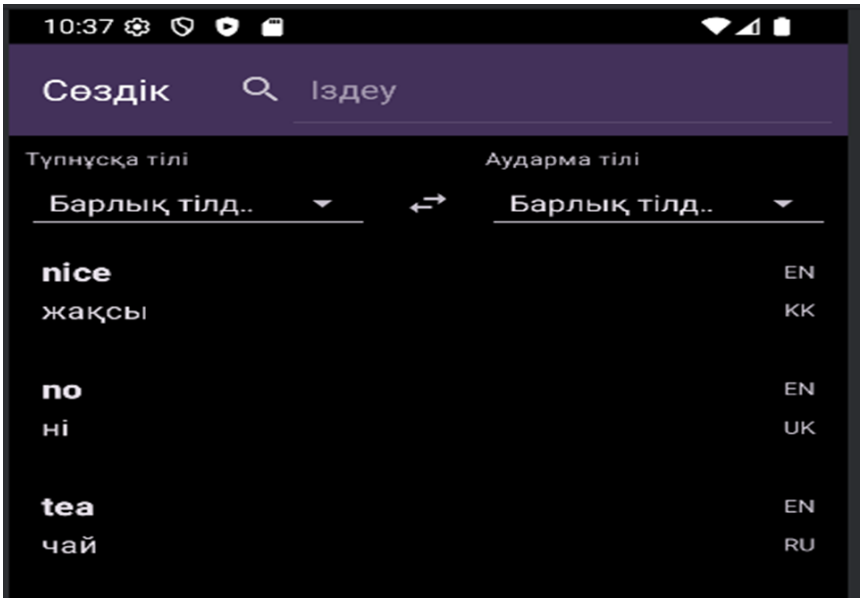


Рисунок 3. Главный экран приложения и структура персонального словаря

Переход платформы из категории базового хранилища в статус интеллектуального инструмента происходит в момент добавления новой лексики. Для конечного пользователя этот процесс редуцирован до ввода слова и выбора языковой пары, как продемонстрировано на интерфейсе добавления (рисунок 4).

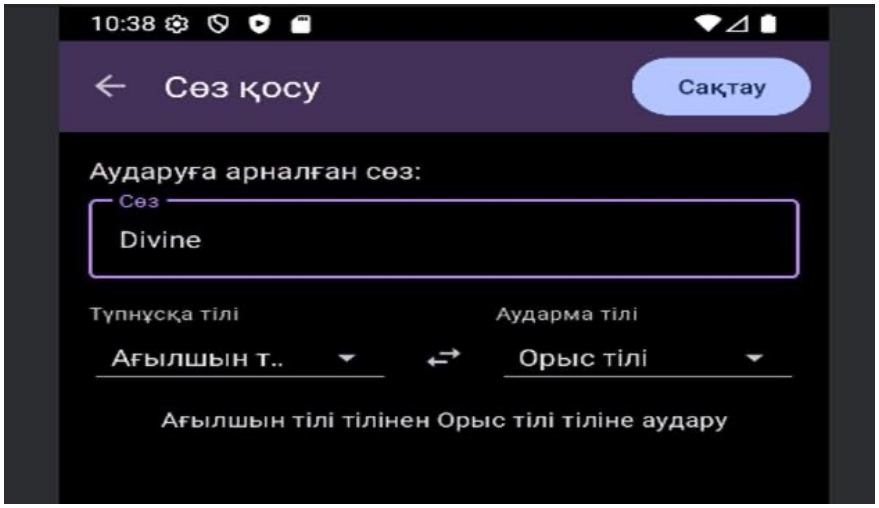


Рисунок 4. Интерфейс ввода лексики для нейросетевой обработки

Под капотом в этот момент инициируется сложный асинхронный процесс клиент-серверного обмена. Приложение с помощью библиотеки Retrofit формирует криптографически защищенный HTTPS-запрос к RESTful API DeepL. На стороне серверов проприетарная глубокая сеть анализирует контекст и возвращает ответ в формате JSON. Мобильный клиент парсит этот пакет и совершает транзакцию записи в локальную базу. Вся цепочка занимает доли секунды, создавая у пользователя визуальное ощущение бесшовного интеллектуального перевода (рисунок 5).

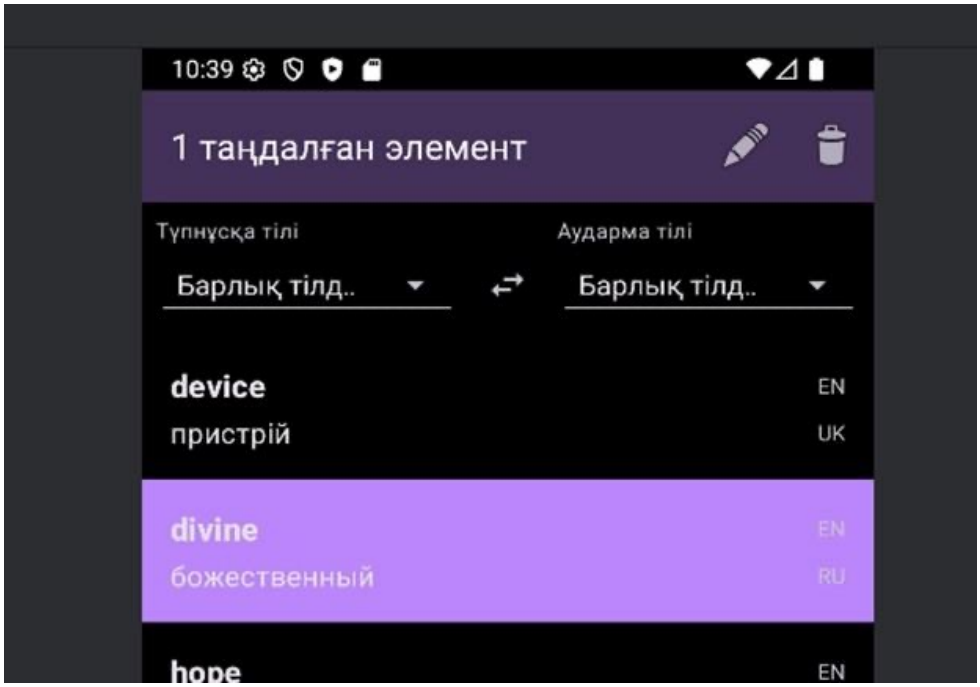


Рисунок 5. Визуальное подтверждение интеграции новой записи

Несмотря на автоматизацию, архитектура системы оставляет за человеком право финальной редактуры. Пользователь может обогатить карточку сленговым значением или ассоциативным рядом, превращая машинный перевод в детализированный обучающий артефакт (рисунок 6).

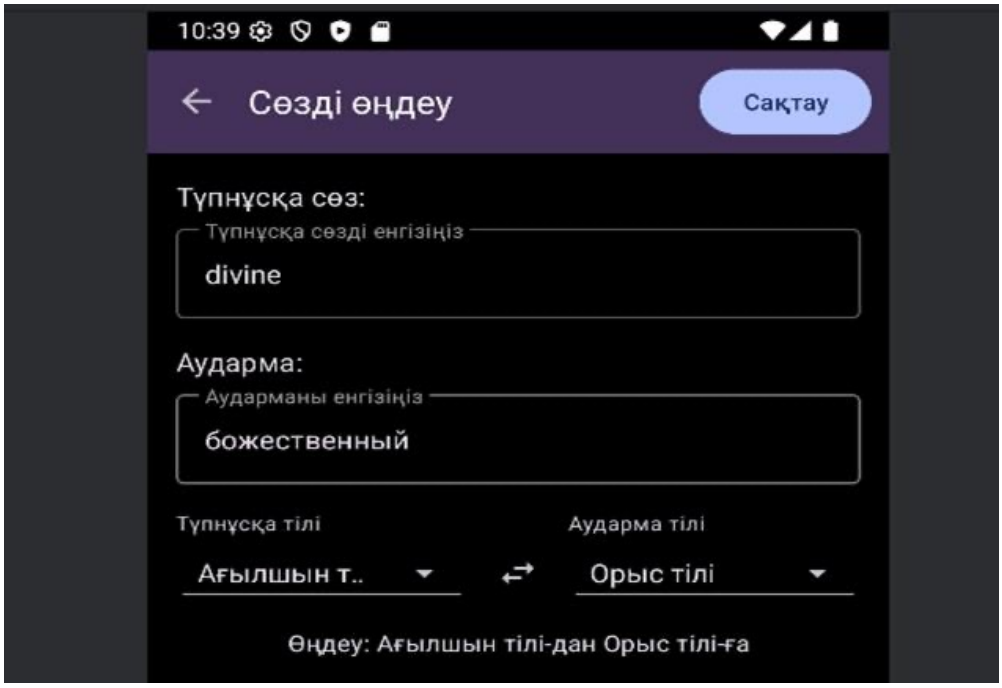


Рисунок 6. Интерфейс ручного контроля и адаптации учебного контента

Геймификация и когнитивные механики удержания. Логическим продолжением формирования базы данных является активное усвоение материала через арсенал интервальных тренировок. Первичный контакт обеспечивается режимом «Флеш-карты».

Изолированная демонстрация стимула на иностранном языке (рисунок 7) заставляет мозг активировать процессы извлечения информации из памяти.

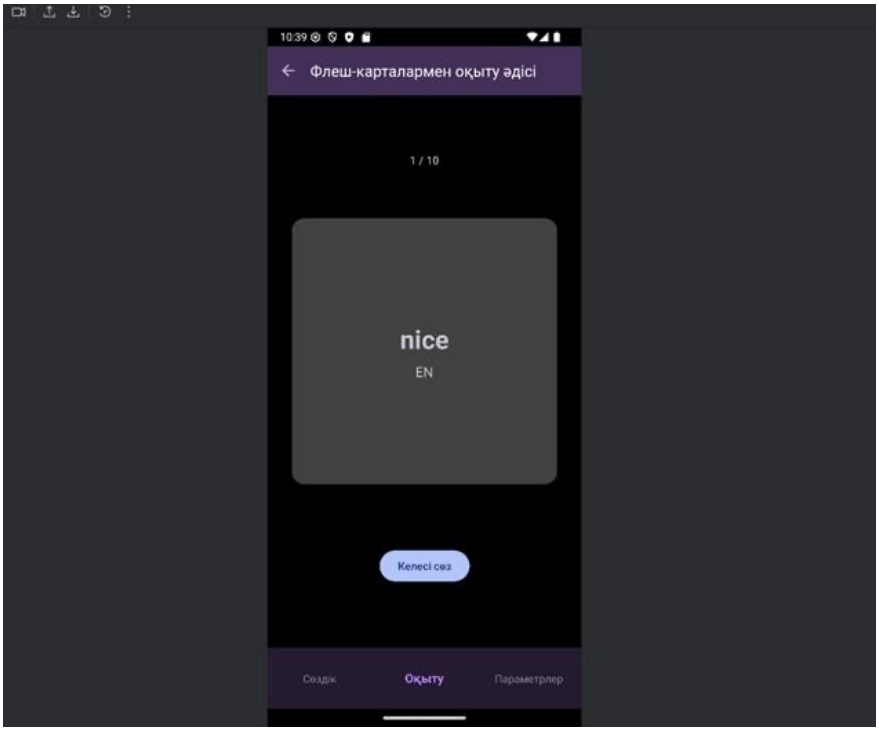


Рисунок 7. Изолированная демонстрация стимула в режиме «Флеш-карты»

Моментальный показ оборотной стороны карточки с правильным ответом (рисунок 8) замыкает петлю обратной связи (Feedback Loop). Этот механизм позволяет мгновенно закрепить верную ассоциацию или скорректировать когнитивную ошибку

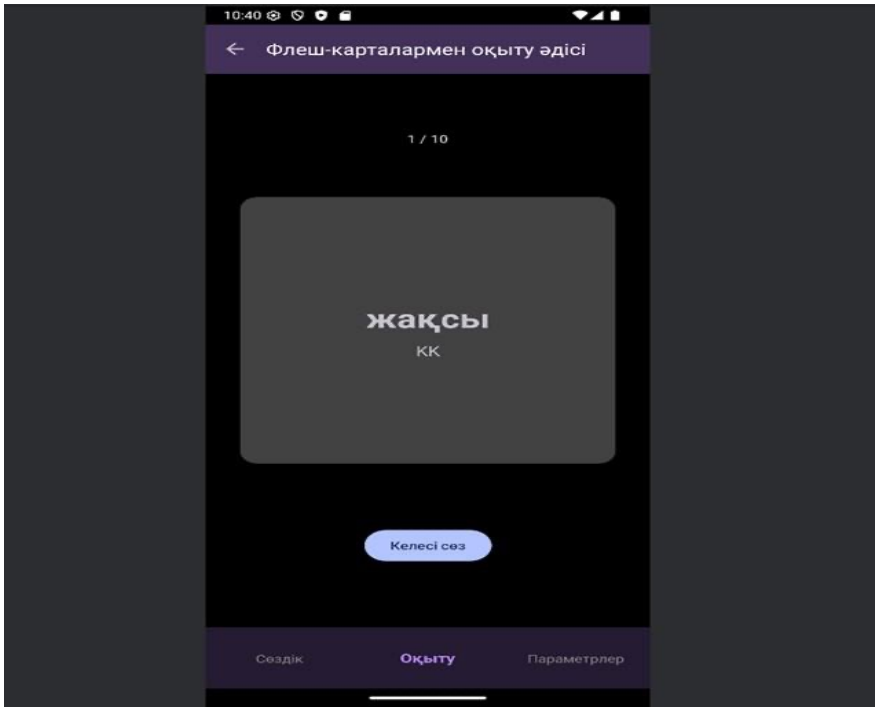


Рисунок 8. Мгновенная верификация перевода для закрепления нейронных связей

Для более глубокого погружения предусмотрен режим прямого ввода текста. Это упражнение форсирует переход лексики из пассивного словаря в активный, требуя от студента самостоятельной генерации ответа (рисунок 9).

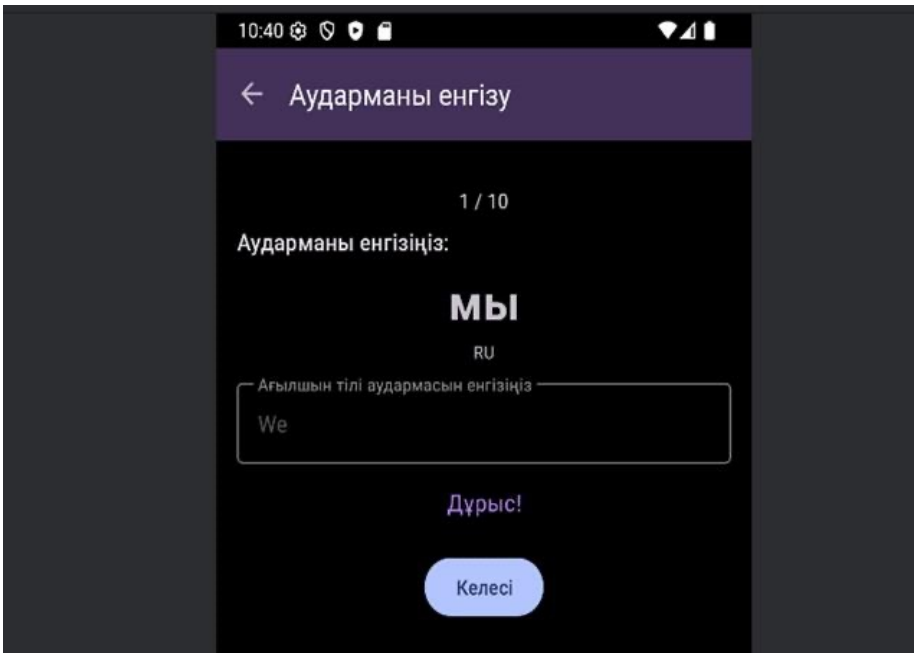


Рисунок 9. Тренажер активного воспроизведения лексики

Снизить когнитивную перегрузку и повысить показатель удержания (Retention Rate) помогает геймифицированный модуль викторины. Выбор правильного перевода среди алгоритмически сгенерированных дистракторов (рисунок 10) выступает эффективным стресс-тестом на внимательность, где каждая ошибка становится точкой роста [8].

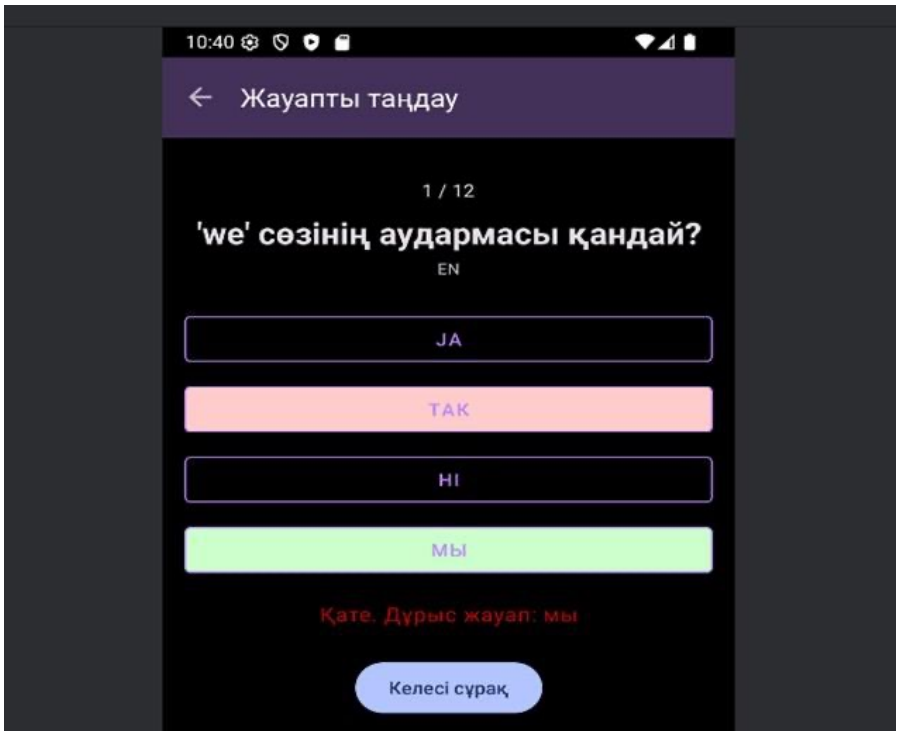


Рисунок 10. Геймифицированный модуль для комплексной проверки знаний

Векторы стратегического развития EdTech-экосистемы. Реализованный продукт доказывает жизнеспособность связки «пользовательский интерфейс плюс облачный ИИ». Однако национальный курс на цифровую трансформацию требует масштабирования подобных подходов. Развитие отечественного сектора образовательных технологий целесообразно выстраивать по трем взаимосвязанным векторам.

Во-первых, назрел переход к адаптивному обучению. Платформы будущего обязаны непрерывно анализировать цифровой след студента, выявляя паттерны ошибок и автоматически перестраивая учебную траекторию. Фундаментом для такой гиперперсонализации должно стать создание R&D центров на базе ведущих технических вузов.

Во-вторых, открытие государственных анонимизированных датасетов дает мощнейший толчок разработчикам. На фоне того, что с 2020 по 2024 год объем глобальных инвестиций в ИИ-стартапы вырос более чем на 37%, локальному рынку требуются механизмы венчурного финансирования для поддержки команд на ранних стадиях.

В-третьих, локальные успехи нуждаются в суверенной инфраструктуре. Зависимость от зарубежных API допустима на этапе прототипирования, но национальная безопасность диктует необходимость инвестиций в собственные вычислительные кластеры и разработку нормативной базы [3].

Заключение. Пройденный путь от анализа теоретических концепций до компиляции рабочего APK-файла доказывает ключевой тезис: внедрение нейросетевых API кратно снижает порог входа в создание сложных, семантически точных продуктов [9]. Разработанное мобильное приложение «Сөздік» де-факто иллюстрирует сдвиг парадигмы – искусственный интеллект перестал быть закрытой лабораторной технологией для транснациональных корпораций и превратился в масштабируемый ресурс для решения реальных прикладных задач. Истинный потенциал этой синергии будет раскрыт в тот момент, когда технологические инициативы разработчиков, академическая наука и государственная поддержка сольются в единую производственную цепь, работающую на повышение интеллектуального капитала страны.

Список литературы

[1] Токаев, К.-Ж. Послание Президента Республики Казахстан народу Казахстана от 8 сентября 2025 года / К.-Ж. Токаев. – Астана: Акорда; 2025.

[2] Narbekkyzy, A. Developing students' critical thinking skills through the use of Artificial Intelligence (AI) in Kazakh language lessons / A. Narbekkyzy // Eurasian Science Review. – 2025. – № 2. – PP. 1791-1802.

[3] Иманалиева, И. AI regulation in healthcare: Recommendations for Kazakhstan: магистерская диссертация / И. Иманалиева. – Астана: Назарбаев Университет; 2025.

[4] Kasenova, L. M. Using Artificial Intelligence Tools in Kazakh Literature Lessons: Advantages and Risks / L. M. Kasenova // Eurasian Science Review. – 2025.

[5] DeepL API Technical Documentation [Электронный ресурс] / DeepL SE. – 2025. – URL: https://www.deepl.com/docs-api/ (дата обращения: 14.02.2025).

[6] Android Developers Official Documentation [Электронный ресурс] / Google. – 2025. – URL: https://developer.android.com/ (дата обращения: 14.02.2025).

[7] Kotlin Programming Language Official Documentation [Электронный ресурс] / JetBrains. – 2025. – URL: https://kotlinlang.org/docs/home.html (дата обращения: 14.02.2025).

[8] Tleshova, Z. Reliability of AI in Foreign Language Speaking Assessment: Comparing Automated and Human Scoring among Undergraduate IT Students in Kazakhstan / Z. Tleshova, Z. Tusselbayeva, A. Ichshanova // Высшая школа Казахстана. – 2025. – № 50(2). – PP. 18-31.

[9] Rakhimova, D. Legal AI in Low-Resource Languages: Building and Evaluating QA Systems for the Kazakh Legislation / D. Rakhimova, A. Turarbek, V. Karyukin // Computers. – 2025. – № 14(9). – P. 354.

[10] Bazarkulova, A. Kazakh Handwriting Recognition / A. Bazarkulova // Journal of Emerging Technologies and Computing. – 2024. – № 62(1). – PP. 88-102.

Авторский вклад

Молдасейтов Диас Айдынович – постановка проблемы исследования, разработка архитектуры мобильного приложения, написание программного кода на языке Kotlin, интеграция нейросетевого API DeepL, проведение функционального тестирования.

Таныкпаева Балауса Ерденовна – разработка методологии исследования, проверка теоретической базы применения ИИ в образовании, научное руководство проектом и оценка эффективности образовательных механик приложения.

DEVELOPMENT OF AN INTELLIGENT MOBILE APPLICATION FOR LEARNING FOREIGN LANGUAGES USING NEURAL NETWORK API

D.A. Moldaseitov
4th year student of the Faculty of Business and Digital Technologies, S. Seifullin Kazakh Agro Technical Research University

B.E. Tanykpaeva
Senior Lecturer, Master of Technical Sciences, Institute of Business and Digital Technologies, S. Seifullin Kazakh Agro Technical Research University

Abstract. The article discusses the development of a native Android mobile application designed for effective foreign language learning.

Keywords: mobile application, artificial intelligence, Android, MVVM architecture, machine translation, spaced repetition, educational technologies, DeepL API.

УДК 004.65

BIG DATA И ПРОДВИНУТАЯ АНАЛИТИКА ПЕРЕПИСЫВАЮТ ПРАВИЛА ДЕЛОВОГО МИРА



С.С. Шайымов
Преподаватель кафедры «Автоматизации» института Телекоммуникаций и информатики Туркменистана
smyradow615@gmail.com



А.Б. Чарыев
Преподаватель кафедры «Автоматизации» института Телекоммуникаций и информатики Туркменистана
arslan.caryyew93@gmail.com



Б.Н. Атанепесов
Преподаватель института Телекоммуникаций и информатики Туркменистана
bayramhhd@mail.ru

С.С.Шайымов

Окончил Государственный энергетический институт Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.

А.Б.Чарыев

Окончил Государственный энергетический институт Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.

Б.Н.Атанепесов

Окончил институт Телекоммуникаций и информатики Туркменистана. Работает в институте Телекоммуникаций и информатики Туркменистана.

Аннотация. На протяжении десятилетий бизнес-решения принимались в тишине кабинетов, опираясь на опыт, интуицию и, порой, простое везение руководителей. Однако сегодня мы наблюдаем фундаментальный сдвиг. В мире, где каждое движение потребителя, каждый клик и каждая транзакция оставляют цифровой след, полагаться только на человеческое чутье становится не просто рискованно, а экономически опасно. Мы вступили в эпоху, когда «мягкие» предположения заменяются «жесткими» данными. Большая аналитика – это не просто новый набор инструментов, это новая философия управления, превращающая хаос неопределенности в математически выверенную стратегию.

Ключевые слова: нейротехнология, нейрофидбек, Big Data, образования.

Введение. Современные роботизированные системы функционируют в средах с высокой динамикой и неопределённостью. В таких условиях использование BIG DATA обеспечивает своевременную корректировку управляющих параметров на основе анализа крупных потоков сенсорной информации. Это согласуется с выводами, указывающими на важность статистической обработки больших наборов данных для минимизации ошибок позиционирования [1].

Фундамент современных бизнес-решений строится на концепции Data-Driven Decision Making. В отличие от классического подхода, где данные использовались лишь для подтверждения уже принятых решений, современный бизнес позволяет данным диктовать вектор развития. Это требует не только технологической оснащенности, но и психологической готовности лидеров признать, что алгоритм может видеть рынок глубже, чем самый опытный топ-менеджер.

Оптимизация начинается с понимания того, что данные – это не статичная картина прошлого, а живой поток, позволяющий моделировать будущее. Когда компания интегрирует продвинутую аналитику в свои процессы, она переходит от реактивного состояния «исправления ошибок» к проактивному состоянию «предотвращения проблем». Это меняет саму структуру организации: иерархические барьеры рушатся, так как право на решение теперь принадлежит тому, чьи доводы подкреплены наиболее точной аналитической моделью.

Чтобы превратить разрозненные массивы Big Data в ценные бизнес-инсайты, необходима сложная инфраструктура, которую часто сравнивают с кровеносной системой организма. На входе мы имеем колоссальные объемы «сырой» информации: логи транзакций, данные GPS-трекеров, активность в социальных сетях и даже погодные сводки. Все это стекается в так называемые «озера данных» (Data Lakes), где информация хранится в исходном виде до востребования.

Здесь в игру вступают технологии продвинутой аналитики. Машинное обучение и нейронные сети выступают в роли современных алхимиков, превращающих свинец неструктурированных цифр в золото управленческих решений. Предиктивные модели анализируют исторические закономерности, выявляя скрытые корреляции, которые недоступны человеческому глазу. Например, алгоритм может обнаружить, что спрос на определенный вид товаров в конкретном регионе коррелирует не с сезоном, а с изменениями в стоимости логистики у конкурентов за три недели до этого. Именно такие микро-инсайты в совокупности дают колоссальное преимущество в масштабе всей корпорации.

Маркетинг традиционно считался одной из самых творческих и «неточных» сфер бизнеса. Продвинутая аналитика изменила это навсегда. Сегодня мы говорим не о массовых рынках, а о гиперсегментации. Компании больше не продают товар «женщинам от 25 до 35 лет». Вместо этого они обращаются к конкретному человеку в тот самый момент, когда его потребность в продукте достигает пика.

Анализ больших данных позволяет выстраивать «путь клиента» (Customer Journey) с точностью до секунды. Используя предиктивную аналитику, бизнес может предсказать вероятность оттока клиента еще до того, как сам клиент об этом задумался. Если паттерны поведения пользователя – частота входов в приложение, время ожидания ответа поддержки, изменение суммы чека – начинают напоминать поведение тех, кто уже ушел, система

автоматически запускает протокол удержания. Это может быть персональная скидка, звонок менеджера или специально подобранный контент. Такая оптимизация маркетинговых затрат позволяет не тратить бюджет на тех, кто лоялен, и точечно спасать тех, кто колеблется.

В физическом мире оптимизация бизнес-решений через Big Data проявляется наиболее наглядно в цепочках поставок. Для глобальных ритейлеров или производственных гигантов ошибка в прогнозировании спроса на 1% может означать миллионные убытки. Продвинутая аналитика позволяет синхронизировать производство, складские запасы и логистику в единый саморегулирующийся механизм.

Современные системы управления складом, использующие Big Data, способны предсказывать, какие товары будут востребованы в конкретном районе города в ближайшие выходные, и заранее перемещать их на ближайшие распределительные пункты. Это минимизирует «последнюю милю» доставки и сокращает время ожидания для клиента. В логистике алгоритмы оптимизации маршрутов учитывают тысячи факторов: от пробок и ремонтов дорог до расхода топлива конкретным водителем. В результате бизнес получает не просто экономию на бензине, а максимально эффективное использование своего самого дорогого ресурса – времени.

Одной из самых инновационных областей применения Advanced Analytics стало управление персоналом, или People Analytics. Оптимизация бизнес-решений здесь касается найма, обучения и удержания талантов. Анализируя данные об успешных сотрудниках, компании создают модели идеального кандидата для конкретной роли. Это снижает текучесть кадров и затраты на повторный поиск специалистов.

Более того, аналитика позволяет выявлять скрытых лидеров и «узкие места» в коммуникациях внутри компании. Анализ метаданных корпоративной почты и мессенджеров (без нарушения конфиденциальности содержания) помогает понять, как на самом деле движется информация в организации. Если два ключевых отдела перестают взаимодействовать, аналитика подает сигнал руководителю, позволяя принять управленческое решение до того, как это скажется на финансовых результатах проекта.

Несмотря на восторги от эффективности, использование Big Data несет в себе серьезные вызовы. Чем больше компания знает о своих процессах и клиентах, тем большую ответственность она несет за сохранность этих данных. Оптимизация решений не должна превращаться в цифровую диктатуру.

Проблема «черного ящика» остается одной из самых острых в продвинутой аналитике. Когда алгоритм принимает решение об отказе в кредите или об увольнении сотрудника, бизнес должен быть способен объяснить логику этого процесса. В противном случае возникает риск потери доверия со стороны общества и регуляторов. Этичная аналитика – это баланс между эффективностью и человечностью, и поиск этого баланса сам по себе является важнейшим бизнес-решением современности.

Заключение: На пороге автономного бизнеса. Будущее, к которому нас ведут Big Data и Advanced Analytics, – это создание полностью автономных бизнес-экосистем. Мы приближаемся к моменту, когда рутинные решения по закупкам, ценообразованию и распределению ресурсов будут приниматься искусственным интеллектом без участия человека. Роль руководителя сместится от операционного контроля к стратегическому видению и этическому надзору.

Оптимизация через данные – это путь без финишной черты. Каждый новый байт информации делает модели точнее, а решения – эффективнее. В этом бесконечном цикле совершенствования и кроется главный секрет успеха в экономике XXI века: побеждает не тот, у кого больше ресурсов, а тот, кто быстрее и точнее превращает данные в действия.

Список литературы

[1] Чарыев А.Б., Шайымов С.С., Аннабердиев И.А., «Важность использования BIG DATA для оптимизации управления роботами-манипуляторами». Международный научный журнал «ИННОВАЦИОННАЯ НАУКА» №2-1/2025.

[2] Шайымов С.С., Гочов А.С., Веллиева С.Т., «Значение BIG DATA в цифровую эпоху». Международный научный журнал «ИННОВАЦИОННАЯ НАУКА» №2-2/2025.

Авторский вклад

Авторы внесли равнозначный вклад.

THE AGE OF PREDICTION: HOW BIG DATA AND ADVANCED ANALYTICS ARE REWRITING THE RULES OF THE BUSINESS WORLD

S.S.Shayymov
Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan

A.B.Charyev
Lecturer of the Department of "Automation" of the Institute of Telecommunications and Informatics of Turkmenistan

B.N.Atanepesov
Lecturer at the Institute of Telecommunications and Informatics of Turkmenistan

Abstract. For decades, business decisions were made in the quiet of offices, relying on experience, intuition, and, sometimes, sheer luck. However, today we are witnessing a fundamental shift. In a world where every consumer movement, every click, and every transaction leaves a digital trace, relying solely on human intuition is becoming not just risky, but economically dangerous. We have entered an era where "soft" assumptions are being replaced by "hard" data. Big analytics is not just a new set of tools; it is a new management philosophy that transforms the chaos of uncertainty into mathematically sound strategy.

Keywords: neurotechnology, neurofeedback, Big Data, education.

УДК 004.8:616.728.48-073

СПОСОБЫ АУГМЕНТАЦИИ ДАННЫХ ПРИ РЕАЛИЗАЦИИ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ДЛЯ ДИАГНОСТИКИ НАРУШЕНИЙ РАВНОВЕСИЯ И РАБОТЫ ГОЛЕНОСТОПА



Н.С. Глушаченко
Магистрант кафедры программного обеспечения информационных технологий БГУИР
nglushachenko1@gmail.com



Д.В. Деменковец
Старший преподаватель кафедры программного обеспечения информационных технологий БГУИР, магистр
demenkovets@bsuir.by

Н.С. Глушаченко

Окончил Белорусский государственный университет информатики и радиоэлектроники по специальности «Программное обеспечение информационных технологий». Область научных интересов связана с анализом и разработкой моделей и алгоритмов построения интеллектуальных систем поддержки принятия решений

Д.В. Деменковец

Окончил Белорусский государственный университет информатики и радиоэлектроники по специальности «Программное обеспечение информационных технологий». Область научных интересов связана с программированием встроенных систем, методами тестирования запоминающих устройств, промышленной автоматизацией, программированием промышленных микроконтроллеров.

Аннотация. Рассмотрена задача построения системы поддержки принятия решений для диагностики нарушений равновесия и работы голеностопного сустава. Проведён сравнительный анализ классической стабилометрии и двигательно-когнитивных тестов с биологической обратной связью на балансировочном диске с акселерометром. Выполнен обзор открытых датасетов и констатировано отсутствие наборов данных с балансировочного диска, что обосновывает проецирование данных силовых платформ на координатное пространство нестабильной платформы. Описаны четыре метода аугментации временных рядов, позволяющие сформировать сбалансированную обучающую выборку при дефиците клинических данных.

Ключевые слова: система поддержки принятия решений, стабилометрия, биологическая обратная связь, балансировочный диск, центр давления, голеностопный сустав, аугментация данных

Введение. Особую роль в повышении эффективности восстановления пациентов с нарушениями опорно-двигательного аппарата играют двигательно-когнитивные тесты с использованием биологической обратной связи (БОС). В таких тестах пациент выполняет целенаправленные действия с помощью движений своего тела: управляет виртуальным объектом на экране, перемещая центр давления (ЦД) путём изменения углов наклона голеностопа или перемещая центр давления своего тела с целью зафиксироваться в определённой точке-мишени [1, 2].

Для количественной оценки процесса управления позой применяются два взаимодополняющих подхода: стабилометрия и двигательные тесты с использованием БОС. При проведении стабилометрии пациент удерживает равновесие в статической позе и регистрируюся колебания ЦД [10]. В двигательно-когнитивных тестах с БОС пациент целенаправленно перемещает свой ЦД в указанные зоны на экране [5]. Именно второй тип тестов позволяет выявить нарушения голеностопа, которые не обнаруживаются в статических пробах. Кроме силовых платформ для оценки состояния голеностопного сустава применяются балансировочные диски, оснащённые акселерометрами. В отличие от силовых платформ, балансировочный диск создаёт управляемую нестабильность: пациент стоит на качающейся поверхности, а встроенные датчики регистрируют углы наклона платформы. Подобный подход позволяет выявлять пациентов с хронической нестабильностью голеностопа и обеспечивает их реабилитацию. При этом, ключевым диагностическим параметром выступает угол наклона. Принципиальное преимущество нестабильных платформ заключается в целенаправленной нагрузке на голеностопный сустав [12]. Несмотря на высокую информативность обоих подходов, с ростом объёмов регистрируемых данных возникает проблема их ручной интерпретации. Для снижения нагрузки на врачей-ортопедов и автоматизации процесса постановки диагноза рассматривается применение специализированных программных средств [7].

Аппаратное обеспечение исследований. Традиционно для проведения стабилометрических исследований и анализа работоспособности двигательного аппарата применяются стационарные аппаратно-программные комплексы и силовые платформы. Подобные системы с высокой точностью регистрируют динамику смещения центра давления, однако их существенными недостатками являются высокая стоимость, большие габариты и жёсткая привязка к специализированным учреждениям здравоохранения [9, 10].

В противовес традиционным стационарным подходам может применяться портативный координационно-диагностический комплекс, который может быть установлен непосредственно на балансировочный диск. В таком случае для регистрации движений пациента используется встроенный акселерометр, вычисляющий углы наклона [7, 11]. Изменения положения центра давления испытуемого регистрируются в виде углов отклонения от положения равновесия платформы. Разработка портативных постурографических устройств на основе акселерометров описана в работе Dolinay et al.

[11], где продемонстрирована возможность создания компактного и недорогого устройства, регистрирующего наклоны платформы с помощью акселерометра и передающего данные в реальном времени для анализа постурального контроля. Данный подход подтверждает применимость акселерометрических датчиков как альтернативы дорогостоящим силовым платформам. В качестве альтернативы стационарным системам проводятся исследования применимости портативных решений. Показано, что даже игровые контроллеры (в частности, Nintendo Wii Balance Board) обеспечивают измерения ЦД, сопоставимые с лабораторными силовыми платформами [4].

Принципиальное отличие координационно-диагностического комплекса на балансировочном диске от классической силовой платформы состоит в том, что регистрируемой величиной являются не координаты ЦД в миллиметрах, а углы наклона в градусах. Аналогичный подход реализован в работе Laessoe et al. [12], где инструментальный wobble board с акселерометрическими датчиками продемонстрировал надёжную дифференциацию здоровых субъектов и пациентов с хронической нестабильностью голеностопа по вариации медиолатерального угла наклона. При этом стабилометрические метрики (СКО, средняя скорость, площадь эллипса), первоначально определённые для координат ЦД в миллиметрах, математически тождественны при применении к угловым координатам (в градусах). Для обеспечения высокого качества данных, поступающих в систему поддержки принятия решений, были оптимизированы временные характеристики записи. Если на ранних этапах развития стабилометрических исследований частота дискретизации фиксировалась на уровне 5 Гц из-за технологических ограничений, то в современных исследованиях оптимальной частотой дискретизации является 100 Гц [9].

Анализ открытых стабилометрических данных. Поскольку сбор обширных клинических баз ограничен, важным этапом является анализ существующих открытых наборов данных. Наиболее релевантным набором для задач, связанных с анализом колебаний ЦД, является A Public Data Set of Human Balance Evaluations [15]. Набор включает записи 163 субъектов в возрасте от 18 до 86 лет, протестированных в четырёх условиях: открытые/закрытые глаза на жёсткой и мягкой поверхностях. Каждое из четырёх условий повторяется три раза, что даёт около 1 956 отдельных записей.

Набор Balance-UCSD-VPL [16] содержит 11 028 клипов по 120 фреймов с частотой 60 Гц. Записи содержат временные ряды CoP в медиолатеральном и антеропостериорном направлениях для 21 здорового субъекта, а также синхронизированные 3D-координаты скелетных точек тела. Субъекты выполняли случайные движения на силовой платформе без перемещения стоп, что включает как статические, так и квазидинамические задачи.

Набор OLST [17] содержит синхронизированные данные ЦД для 32 субъектов (17 из которых старше 64 лет) при выполнении теста на одной ноге.

Анализ данных двигательных тестов и мультимодальные наборы данных. Открытые датасеты, содержащие записи тестов с активным перемещением ЦД к мишени, значительно менее распространены, чем данные статической стабилометрии. Это обусловлено тем, что подобные протоколы (LOS-тесты, игровые тренажёры с БОС) часто реализуются на проприетарном оборудовании с закрытыми форматами данных. Набор Balance-Multiview [18] содержит свыше 6 миллионов фреймов от 20 субъектов. Включает синхронизированные 2D-координаты скелета (4 камеры) и данные ЦД. Субъекты выполняют не только статические, но и динамические постуральные задачи с перемещениями, что приближает протокол к тестам на предел устойчивости. Мультимодальный набор Suprapostural Coordination Dataset [19], включает данные 48 здоровых молодых людей, выполнявших задания как на стабильной поверхности, так и на балансировочной платформе (медиолатеральная нестабильность). Однако данный набор ориентирован на оценку двигательных функций при свободном перемещении пациентов и напрямую не затрагивает оценку работы голеностопа. Следует отметить недостаток данных,

полученных при использовании нестабильной платформы. Большая часть данных получена при использовании силовой платформы, которая позоляет оценить только работу вестибулярного аппарата и не затрагивает оценку подвижности голеностопа. Поэтому были рассмотрены методы проецирования данных ЦД, полученных на силовых платформах, на координатное пространство балансировочного диска (пересчёт миллиметров в углы наклона) с последующей аугментацией временных рядов для формирования обучающей выборки.

Аугментация стабилометрических данных для построения СППР. Были реализованы четыре базовых метода аугментации временных рядов. Каждый из методов математически моделирует реальные физиологические или аппаратные флуктуации, возникающие при сборе данных на балансировочной платформе:

1. Добавление Гауссовского шума (Jittering). К исходным координатам X и Y добавляется случайный шум с нормальным распределением. С физиологической точки зрения этот метод имитирует естественный мышечный микротремор [11].

2. Масштабирование амплитуды (Scaling). Амплитуда отклонения ЦД напрямую зависит от роста, веса и высоты центра масс пациента. Масштабирование позволяет синтезировать данные, имитирующие пациентов различной комплекции [8, 14].

3. Вращение системы координат (Rotation). Данная трансформация критически важна для нивелирования ошибок позиционирования. На практике пациенты (особенно дети) не всегда ставят стопы на балансировочный диск идеально ровно относительно осей датчика [8]. Обучение на повёрнутых траекториях позволяет алгоритму выделять инвариантные признаки патологии, не зависящие от постановки ног [14].

4. Временная деформация (Time Warping). В контексте двигательно-когнитивных тестов с БОС скорость реакции пациентов варьируется. Деформация имитирует различную скорость сближения пациента с мишенью и задержки когнитивной реакции, сохраняя при этом общую топологию пройденного пути [8, 14].

Заключение. В настоящей работе рассмотрены способы аугментации данных для реализации системы поддержки принятия решений для диагностики нарушений равновесия и работы голеностопного сустава. Обоснована актуальность двигательно-когнитивных тестов с БОС и их преимущество перед статической стабилометрией для выявления скрытых нарушений голеностопа, а также показана роль инструментальных балансировочных дисков как портативной альтернативы стационарным комплексам. Систематизированы доступные датасеты и констатировано отсутствие открытых наборов данных с записями координат ЦД непосредственно с балансировочного диска. Описаны методы расширения данных, каждый из которых имеет физиологическое или аппаратное обоснование. Использование аугментации позволяет сбалансировать обучающую выборку и повысить обобщающую способность классификаторов по миноритарным классам патологий. Полученные результаты свидетельствуют о том, что интеграция недорогих портативных балансировочных платформ способна обеспечить диагностическую точность, сопоставимую с традиционными стационарными комплексами. Архитектура предоставляет врачу-ортопеду объективный, оцифрованный инструмент для оценки динамики реабилитации.

Список литературы

[1] Prieto T. E. et al. Measures of postural steadiness: differences between healthy young and elderly adults //IEEE Transactions on biomedical engineering. – 1996. – Т. 43. – №. 9. – С. 956-966.

[2] Collins J. J., De Luca C. J. Open-loop and closed-loop control of posture: a random-walk analysis of center-of-pressure trajectories //Experimental brain research. – 1993. – Т. 95. – №. 2. – С. 308-318.

[3] Tremblay M. S. et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth //International journal of behavioral nutrition and physical activity. – 2011. – Т. 8. – №. 1. – С. 98.

[4] Kennedy M. W. et al. Enhanced feedback in balance rehabilitation using the Nintendo Wii Balance Board //2011 IEEE 13th international conference on e-health networking, applications and services. – IEEE, 2011. – С. 162-168.

[5] Lange B. et al. Development of an interactive game-based rehabilitation tool for dynamic balance training //Topics in stroke rehabilitation. – 2010. – Т. 17. – №. 5. – С. 345-352.

[6] Девялтовская М.Г., Платонов А.В., Козловский Д.А. Биологическая обратная связь с визуальным каналом и стабилометрией: учеб.-метод. пособие. – Минск: БелМАПО, 2021. – 23 с.

[7] Глушаченко Н.С., Деменковец Д.В., Куйко Н.С. Способ сбора данных при использовании координационно-реабилитационного комплекса для исследования координации детей с нарушением опорно-двигательного аппарата // Компьютерные системы и сети: сб. ст. 60-й науч. конф. аспирантов, магистрантов и студентов БГУИР. – Минск, 2024. – С. 710–715.

[8] Fawaz H. I. et al. Data augmentation using synthetic data for time series classification with deep residual networks //arXiv preprint arXiv:1808.02455. – 2018.

[9] Scoppa F. et al. Clinical stabilometry standardization: basic definitions–acquisition interval–sampling frequency //Gait & posture. – 2013. – Т. 37. – №. 2. – С. 290-292.

[10] Гаже П.М., Вебер Б. Постурология / Пер. с франц.; под ред. М.М. Алексаняна. – СПб.: СПбМАПО, 2008. – 316 с.

[11] Dolinay J. et al. Posturography device based on accelerometer //Int J Syst Appl Eng Develop. – 2014. – Т. 8. – С. 155-62.

[12] Laessoe U. et al. Evaluation of functional ankle instability assessed by an instrumented wobble board //Physical therapy in sport. – 2019. – Т. 35. – С. 133-138.

[13] Quijoux F. et al. Center of pressure characteristics from quiet standing measures to predict the risk of falling in older adults: a protocol for a systematic review and meta-analysis //Systematic reviews. – 2019. – Т. 8. – №. 1. – С. 232.

[14] Halmich C. et al. Data augmentation of time-series data in human movement biomechanics: A scoping review //PloS one. – 2025. – Т. 20. – №. 7. – С. e0327038.

[15] Santos D. A., Duarte M. A public data set of human balance evaluations //PeerJ. – 2016. – Т. 4. – С. e2648.

[16] Balance-UCSD-VPL [Электронный ресурс]. – Режим доступа: https://github.com/UCSD-VPL/Balance-UCSD-VPL – Дата доступа: 24.03.2026.

[17] Copeland D. et al. Multimodal Synchronized Motion Capture, Force Plate, and Radar Dataset of the One-Legged Stand Test for Fall-Risk Assessment //PhysioNet. – 2026.

[18] Balance-Multiview [Электронный ресурс]. – Режим доступа: https://github.com/UCSD-VPL/Balance-Multiview – Дата доступа: 24.03.2026

[19] Mangalam M. et al. A multimodal biomechanical and eye-tracking dataset of suprapostural coordination in healthy young adults //Scientific Data. – 2025. – Т. 12. – №. 1. – С. 1311.

Авторский вклад

Глушаченко Никита Сергеевич – аналитический обзор и отбор открытых наборов стабилометрических данных. Реализация алгоритмов искусственного расширения временных рядов для решения проблемы дисбаланса медицинских данных путем добавления шума, масштабирования, вращения координат и временной деформации. Оценка применимости методов аугментации. Сравнение использования силовых и нестабильных платформах в задачах стабилометрии и оценки работы голеностопа

Деменковец Денис Викторович – постановка исследовательской задачи по автоматизации диагностики нарушений равновесия и созданию системы поддержки принятия решений. Формирование методологии цифровой обработки стабилометрических сигналов.

METHODS FOR DATA ENHANCEMENT IN THE IMPLEMENTATION OF A DECISION SUPPORT SYSTEM FOR DIAGNOSING BALANCE DISORDERS AND ANKLE FUNCTION

N. S. Glushachenko
Master's student of the Department of Software for Information Technologies, BSUIR

D.V. Demenkovets
Senior Lecturer, Department of Software for Information Technologies, BSUIR, Master of Technical Sciences

Abstract. This paper considers the problem of developing a decision support system for diagnosing balance disorders and assessing ankle joint function. A comparative analysis was conducted of conventional stabilometry and motor-cognitive tests using biofeedback on a balance disc equipped with an accelerometer. A review of publicly available datasets was conducted, and it was found that no datasets were available from the balancing disk, which justifies projecting data from stable platforms onto the coordinate space of an unstable platform. Four methods for augmenting time series are described, which allow for the creation of a balanced training sample when clinical data is scarce.

Keywords: decision support system, stabilometry, biofeedback, balance disc, center of pressure, ankle joint, data augmentation.

УДК: 519.6

ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ МЕТОДА ПОТЕНЦИАЛЬНЫХ ПОЛЕЙ В ГРУППОВОМ УПРАВЛЕНИИ ЗА СЧЁТ ПОЛИГОНАЛЬНОГО ЗАПОЛНЕНИЯ И РАСПРЕДЕЛЁННОГО ПОИСКА ПУТИ





Т.А. Васяева
Декан факультета информационных систем и технологий, доцент кафедры автоматизированных систем управления ДОННТУ, кандидат технических наук, доцент
vasyaeva@gmail.com

П.Г. Мельник
Аспирант кафедры автоматизированных систем управления ДОННТУ
zizaobycnaa@gmail.com

Т.А. Васяева
Окончила Донецкий национальный технический университет. Область научных интересов связана с разработкой методов и алгоритмов искусственного интеллекта, организацией учебного и научно-исследовательского процессов в техническом университете.

П.Г. Мельник
Окончил Донецкий национальный технический университет. Область научных интересов связанна с разработкой методов навигации автономных систем.

Аннотация. Разработан механизм полигонального заполнения локального минимума для мульти-агентной системы вместо привычных построений плоскостей или сфероидов Гаусса, что гарантирует целостность группы при выходе из локального минимума.

Ключевые слова: потенциальные поля, групповое управление, локальный минимум, архитектура «ведущий-ведомый», полигональное заполнение, распределённый поиск пути, трёхмерная навигация, автономные агенты.

Введение. Групповое управление беспилотными устройствами по принципу «ведущий-ведомый» (leader-follower) является одним из наиболее распространенных подходов в мульти-агентной робототехнике благодаря своей простоте, масштабируемости и эффективности [1]. Такие системы находят применение в широком спектре задач: от совместной транспортировки грузов до мониторинга опасных зон и поисково-спасательных операций.

Ключевым требованием к функционированию групп ведущий-ведомый является способность ведомых устройств сохранять заданную пространственную конфигурацию относительно движущегося лидера при движении в среде с препятствиями. Для обеспечения безопасной навигации ведомые роботы должны одновременно решать две задачи: следовать за лидером и избегать столкновений с препятствиями и другими членами группы.

Одним из наиболее эффективных подходов к решению этой проблемы являются методы искусственных потенциальных полей (Artificial Potential Fields, APF). Однако методы APF имеют фундаментальный недостаток – подверженность локальным минимумам[1,2].

Проблема. В текущих реализациях метода потенциальных полей для избегания локальных минимумов зачастую используется метод «заполнения ям». Этот подход подразумевает создание искусственного отталкивающего поля в найденном локальном минимуме для того, чтобы вытолкнуть агента из него. Заполнение обычно происходит либо плоскостью, перпендикулярной вектору притяжения к цели, либо сфероидом Гаусса в области локального минимума. Оба варианта имеют свои недостатки. Для заполнения плоскостью такими можно считать проблему перекрытия других участков препятствия в случае, когда локальный минимум при небольшом размере сильно «глубже» соседних областей, для сфероидов классической проблемой является выходы за пределы минимума, что потенциально может перекрыть оптимальные маршруты[3].

В мульти-агентной системе недостатки предыдущих методов решения становятся на порядок весомее. Ни один из них не способен гарантировать сохранность группы агентов, а также избегание ситуаций, в которых группа покидает локальный минимум, а один из агентов остается либо в новом локальном минимуме, либо отталкивается в другую сторону от группы.

Постановка задачи исследования. Таким образом, задачу можно было сформулировать так: для группы устройств в трёхмерном пространстве нужно было разработать метод обнаружения и выхода из локальных минимумов, лишенный недостатков классических решений. Данный метод не должен был использовать глобальные планировщики пути, и в целом агенты ничего не должны были знать об окружающей среде вне своей зоны видимости.

Как часть работы, необходимо было разработать такой метод заполнения пространства, который:

- Обеспечил бы гарантированный выход всей группы агентов в одном направлении
- Обеспечил бы сохранность группы как единого объекта
- С малым количеством итераций полностью перекрывал бы возможность нового попадания в тот же локальный минимум
- Не перекрывал ничего, кроме локального минимума

Также было принято решение сразу остановиться на архитектуре группы устройств «ведущий-ведомые»[4]. Это обусловлено как большим потенциалом для модификаций поведения системы, так и большей практической направленностью: на практике почти все мульти-агентные системы используют именно её из-за лучшей координации и простоты управления.

Предлагаемое решение. Проведя анализ существующих методов, были выявлены следующие недостатки существующих решений:

- Недостаточная скорость заполнения локального минимума
- Фрагментарность заполнения
- Риск «зажать» одного из агентов между полем заполнения и препятствием

Опираясь на тот факт, что мульти-агентная система подразумевает наличие минимум трёх агентов, было принято решение о заполнении локального минимума набором полигональных фигур, образующих замкнутую геометрическую конструкцию.

В тестовой мульти-агентной системе существуют две роли – лидер и ведомые, и два состояния у каждого – покой и поиск выхода из локального минимума. В режиме покоя лидер следует к цели, опираясь на классический метод потенциальных полей, а в режиме поиска выхода – остается на месте в качестве якоря системы. Ведомые в режиме покоя следуют за лидером, имея его как точку притяжения для своих потенциальных полей. В режиме поиска выхода именно ведомые, опираясь на положение лидера, ищут точки, которые позже станут основой для полигонального заполнения.

Таким образом, схематичное изображение принципов работы мульти-агентной системы выглядит следующим образом(рисунок 1):

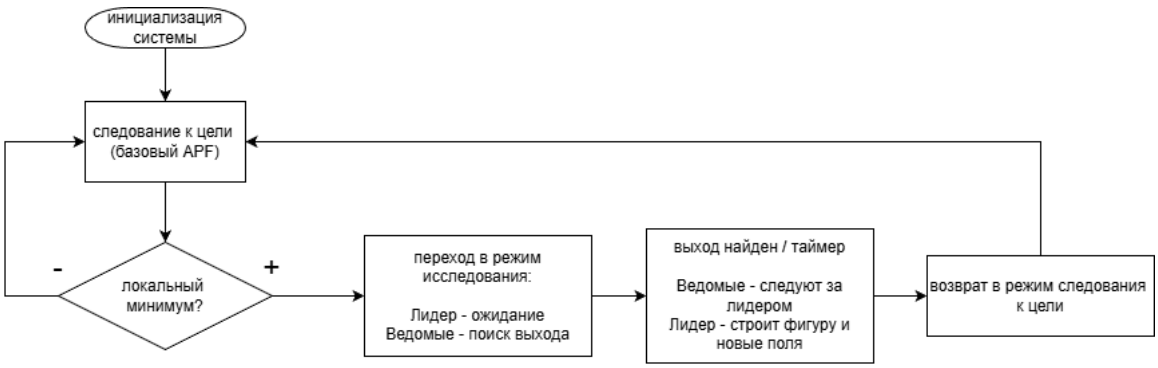


Рисунок 1. Схематичное изображение принципа работы системы

Локальный минимум определяется как падение скорости лидера меньше определённого значения на заданный отрезок времени.

Режим поиска включается после обнаружения лидером локального минимума и длиться заданное время, после чело система возвращается в режим следования к цели. В режиме поиска ведомые следуют вдоль стены препятствия в течении заданного количества времени, или до обнаружения потенциального прохода дальше, после чего посылают свои координаты лидеру.

Наиболее важной частью данной работы является поведение группы в момент попадания в локальный минимум.

Общая идея решения заключается в следующем: послать всех агентов-ведомых вдоль стен локального минимума (метод правой руки) в различных направлениях. После истечения определённого срока или нахождения нового потенциального пути, все ведомые посылают свои текущие координаты лидеру. Лидер собирает координаты ведомых, а также добавляет к ним ещё две точки. Первая – точка локального минимума[5], оп сути – своё положение. Эта точка нужна для закрытия нижней части формируемой фигуры. Вторая нужна для закрытия верхней части фигуры, и вычисляется так:

Пусть заданы:

- A*1* – точка локального минимума (позиция лидера в момент застревания);
- B1,B2 – позиции двух ведомых, на основе которых вычисляется точка притяжения.

Тогда единая точка притяжения C определяется следующим образом:

Находим среднюю точку между двумя ведомыми:

$$M=B1+B22$$

Определяем вектор от точки локального минимума к средней точке:

$$v=M-A$$

Находим перпендикулярный вектор в плоскости XZ:

$$v=-vz\ 0\ vx$$

Нормализуем перпендикулярный вектор:

$$v\hat{}=v\|v\|$$

Вычисляем точку C :

$$C=A+v2+\|v\|2v\hat{}$$

Усредняем высоту точки С по высотам точки локального минимума и ведомых:

$$Cy=Ay+My2$$

Таким образом, формируется замкнутая выпуклая фигура, состоящая из двух наборов полигонов, собранных вокруг одной из двух точек. Схематически, данный вариант для двухмерной среды можно представить следующим образом:

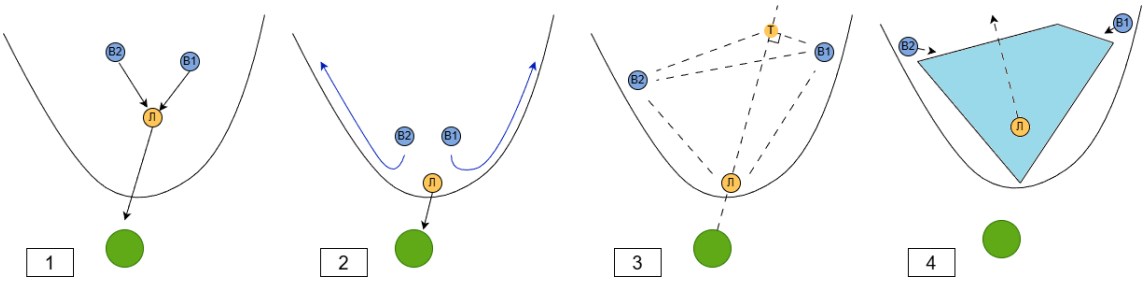


Рисунок 2. Двухмерная реализация метода полигонального заполнения

Эта конструкция обладает отталкивающим полем, и эффективно заполняет локальный минимум любой сложности, т.к опирается не на примитивы, а на замкнутую систему полигонов. Это позволяет избегать проблем, свойственных другим алгоритмам заполнения[6].

Выше на схемах показан максимально схематичный вариант. В действительности же описываемый подход предназначен в первую очередь для трёхмерных сред, где заполнение полигональными фигурами должно быть наиболее эффективным. Для этого была разработана тестовая среда с использованием языка C# (рисунок 3). Далее будет продемонстрирован алгоритм на практике во время одного из тестовых пусков.

На рисунке 2 видны препятствия, конечная точка в виде зелёного куба, и группа агентов: оранжевый – лидер, его пройденный путь, зона видимости, параметры, цель движения и т.д, а также светло-синие – ведомые. Также видны их вектора притяжения к лидеру и отталкивания друг от друга для избежания коллизий.

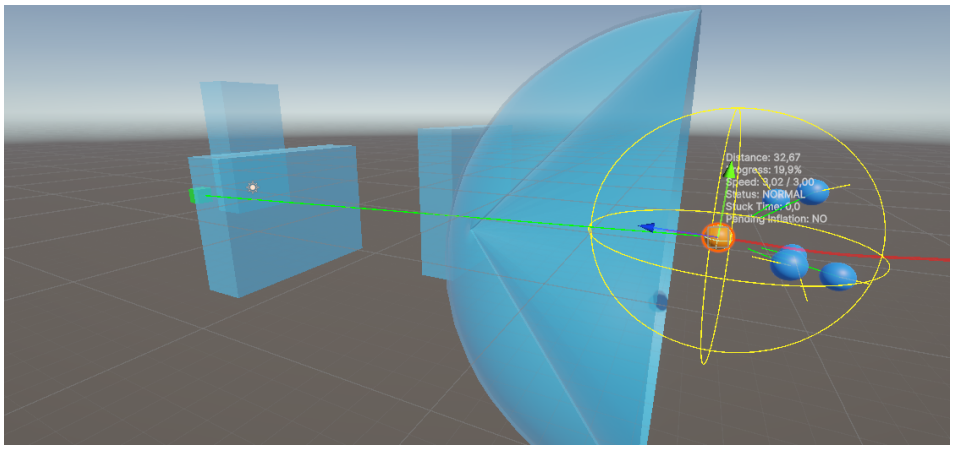


Рисунок 3. Работа трёхмерной среды

Когда лидер сталкивается с локальным минимумом, он переходит в режим ожидания, а все ведомые делят пространство на равные сектора и рассредотачиваются, ища выход, что показано на рисунке 4.

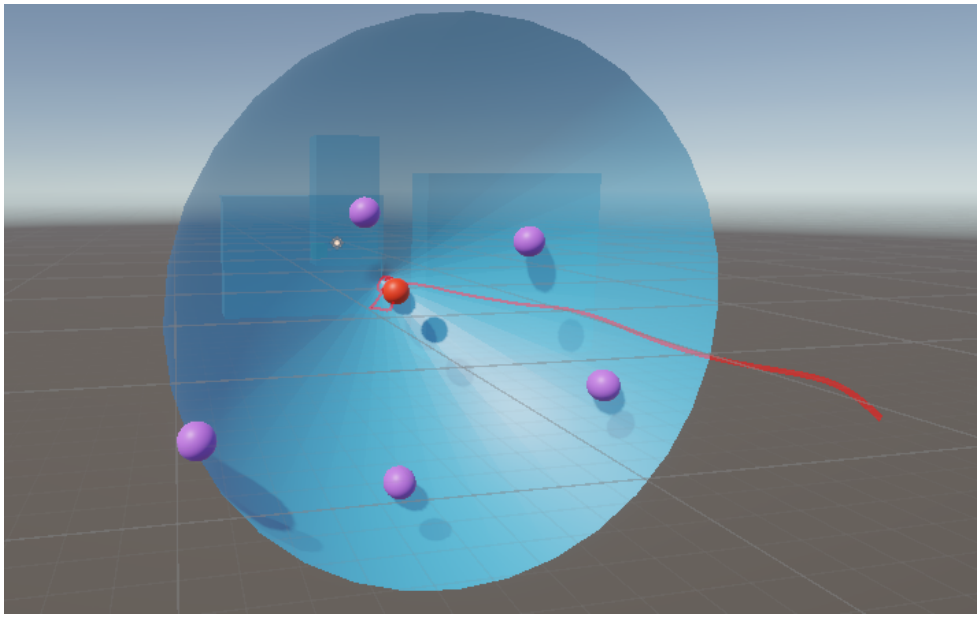


Рисунок 4. Группа в режиме поиска выхода из локального минимума

Как только выход будет найден, произойдёт заполнение пространства полигональной фигурой. Как только она будет создана, все ведомые перейдут обратно к режиму следования за лидером, что обеспечит сохранность группы как единой структуры. Сам же лидер, обретя опору в отталкивающем поле полигонального заполнения, продолжит свой путь к цели.

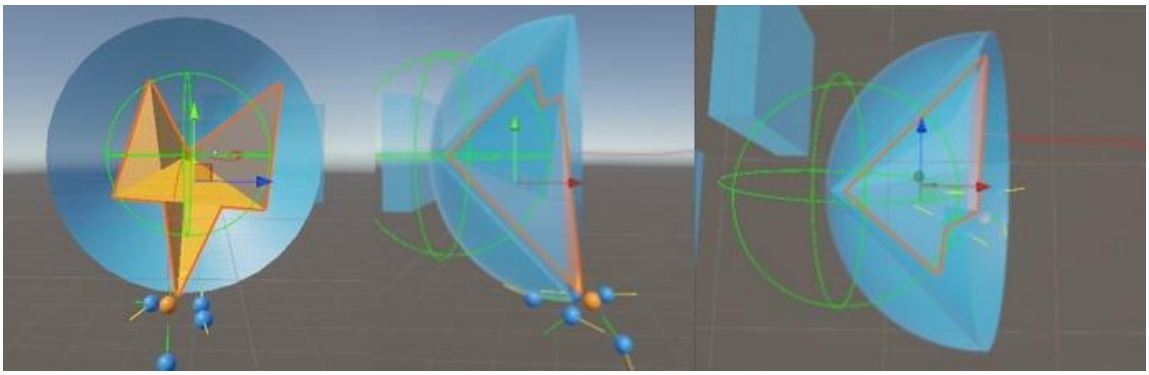


Рисунок 5. Полигональное заполнение в различных проекциях

Как видно на рисунке 5, даже одной итерации алгоритма хватило на то, чтобы практически полностью заполнить собой препятствие. Внешне может показаться, что заполнение недостаточно, и могут возникнуть новые локальные минимумы в местах, где не хватает заполненности, однако учитывая тот факт, что любая реализация метода потенциальных полей предполагает определённую «дальность видимости», то даже при таком заполнении вся группа сконцентрировано покинула локальный минимум. Для сравнения, вот примеры заполнения того же препятствия сфероидом (рисунок 6)

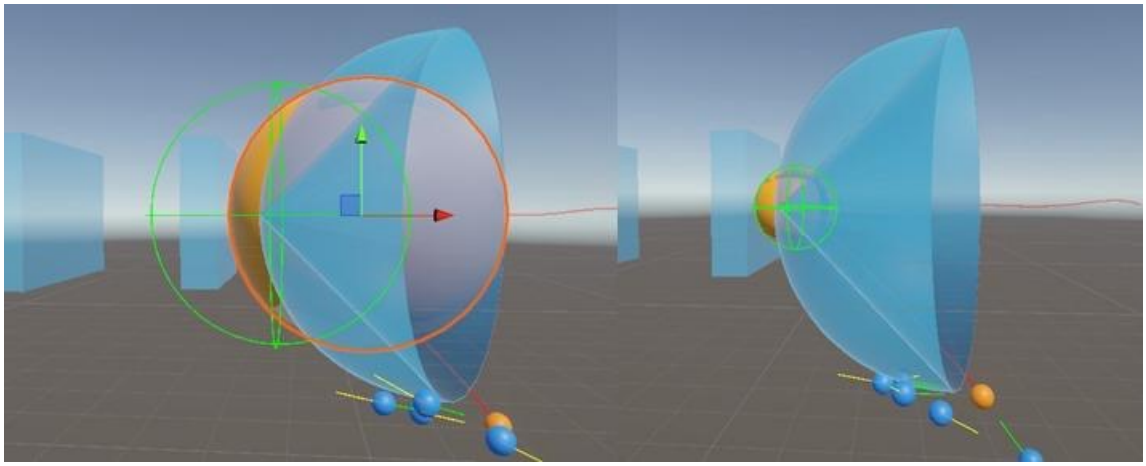


Рисунок 6. Пример заполнения локального минимума сфероидом

Как видно из рисунка, сфероиды, при всей их простоте, для подобного заполнения подходят гораздо меньше. Даже с учётом сдвига центра сферы от точки локального минимума[7], остались зоны, в которых может произойти застревание, а также, в силу особенностей формы, саму группу устройств может разбросать по окружности.

В действительности же заполнение сферами подразумевает куда меньшие диаметры для избежания случайных перекрытий маршрута, что кратно увеличивает время выхода из локального минимума[8,9].

Также демонстрируется ещё один минус.

Сфера перекрыла своим полем и часть пространства за препятствием, что может быть критичным в некоторых моментах.

Эксперимент. Реальную эффективность данного метода можно оценить только в сравнении с другими методами в трёхмерном пространстве.

Для этого было создано трёхмерное окружение, в котором были реализованы, помимо предлагаемого алгоритма, также алгоритм заполнения сфероидами со смещением.

Также, в связи с тем, что эффективность метода зависит от количества агентов в системе, было принято решение проводить замеры при разном количестве агентов.

Таким образом, были составлены следующие сценарии тестирования:

Таблица 1. Сценарии тестирования

№	Кол-во ведомых	Тип заполнения
1	3	сфероидный
2	3	полигональный
3	5	сфероидный
4	5	полигональный
5	8	сфероидный
6	8	полигональный

Основным оцениваемым параметром будет время, за которое лидер достигает конечной точки. Однако также будет оцениваться:

- Кол-во переключений в режим поиска выхода – насколько удачно метод позволяет покидать область локального минимума
- Кол-во потерь части группы – если один из агентов группы отдалится от лидера на достаточное расстояние, он будет считаться «утерянным»

Проведя для каждого сценария по пятьсот запусков, удалось получить следующие результаты (таблица 2).

Таблица 2. Результаты тестирования методов полигонального заполнения и заполнения сфероидами

№	Кол-во ведомых	Тип заполнения	среднее время, сек	процент запусков с потерянными агентами	среднее время итерации	итераций до выхода из минимума
1	3	сфероидный	27.21	7.44%	3.21	3.48
2	3	полигональный	19.92	0	4.81	1.32
3	5	сфероидный	27.92	8.68%	3.08	3.29
4	5	полигональный	20.04	0.8%	4.72	1.18
5	8	сфероидный	27.61	10.31%	3.35	3.52
6	8	полигональный	20.61	1.2%	5.01	1.06

Анализ результатов и выводы. Как видно из результатов тестирования, разработанный метод заполнения показывает устойчивое преимущество по ряду параметров для данной ситуации. При отсутствии глобального планировщика, адаптивная форма заполнения демонстрирует высокую степень отказоустойчивости, что и было одной из целей разработки данного метода.

Явно видно, что время одной итерации попытки выхода у предложенного метода больше, в первую очередь из-за того, что агентам-ведомым нужно время, чтобы продвинуться вдоль стен препятствия. Однако эффективность данной попытки ощутимо выше, чем у метода заполнения сферами.

Также из результатов тестирования видно, что хотя процент потерь агентов при использовании данного метода очень небольшой, он всё же есть. Отчасти он обусловлен острыми краями одного из элементов окружения, однако это всё ещё одно из направлений для дальнейших улучшений.

Также было замечено, что не смотря на увеличение количества агентов в группе, процент потерь агентов растёт не пропорционально количеству агентов. Это связано с тем, что большее количество агентов увеличивает и качество заполнения локального минимума, что уменьшает риск отсечения агента даже об острые края препятствия.

Перспективным направлением для дальнейшего развития данного метода является работа с положением агентов в пространстве, а не простое следование за лидером и отталкивание от других агентов.

Расположение агентов в заранее известном порядке позволит улучшить зону их покрытия, что должно повысить качество заполнения.

Также важным моментом является тот факт, что хотя данный метод показал высокую эффективность, универсальным он не является, и предназначен строго для систем, в которых количество агентов больше трёх, не считая лидера. В противном случае алгоритму просто не хватит точек для построения достаточного количества полигонов.

Список литературы

[1] Khatib O. Real-Time Obstacle Avoidance for Manipulators and Mobile Robots. *The International Journal of Robotics Research*. 1986;5(1):90-98. DOI: 10.1177/027836498600500106.

[2] Choset H., Lynch K.M., Hutchinson S., Kantor G., Burgard W., Kavraki L.E., Thrun S. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. Cambridge: MIT Press; 2005. 632 p.

[3] Baziyad M., Al Shohna M., Rabie T. The Bulldozer Technique: Efficient Elimination of Local Minima Traps for APF-Based Robot Navigation. *arXiv preprint arXiv:2512.23672*. 2025.

[4] Barraquand J., Latombe J.C. Robot Motion Planning: A Distributed Representation Approach. *The International Journal of Robotics Research*. 1991;10(6):628-649. DOI: 10.1177/027836499101000604.

[5] Kong H., Yang C., Liu J. A Hybrid Path Planning Method for Mobile Robot Based on Artificial Potential Field Method. In: Yu H., Liu J., Liu L., Ju Z., Liu Y., Zhou D., eds. *Intelligent Robotics and Applications – ICIRA 2019*. Cham: Springer; 2019. p. 325-331. (Lecture Notes in Computer Science; vol. 11741). DOI: 10.1007/978-3-030-27538-9_28.

[6] LaValle S.M. *Planning Algorithms*. Cambridge: Cambridge University Press; 2006. 842 p. DOI: 10.1017/CBO9780511546877.

[7] Desai J.P., Ostrowski J.P., Kumar V. Modeling and Control of Formations of Nonholonomic Mobile Robots. *IEEE Transactions on Robotics and Automation*. 2001;17(6):905-908. DOI: 10.1109/70.976023.

[8] Ge S.S., Cui Y.J. New Potential Functions for Mobile Robot Path Planning. *IEEE Transactions on Robotics and Automation*. 2000;16(5):615-620. DOI: 10.1109/70.880813.

[9] Zhou Y., Li X., Zhang J. A Review on Local Minimum and Multiple Minimum Avoidance Techniques in Local Path Planning. *Journal of Intelligent & Robotic Systems*. 2021;102(1):25. DOI: 10.1007/s10846-021-01368-6.

Авторский вклад

Васяева Татьяна Александровна – руководство исследованием, подбор метрик, участие в проведении анализа полученных данных

Мельник Пётр Георгиевич – реализация тестового окружения и алгоритмической базы, разработка метода полигонального заполнения

IMPROVING THE EFFICIENCY OF THE POTENTIAL FIELD METHOD IN GROUP CONTROL THROUGH POLYGONAL FILLING AND DISTRIBUTED PATHFINDING

T.A. Vasiaeva
Dean of the Faculty of Information Systems and Technologies, Associate Professor of the Department of Automated Control Systems of DONNTU, PhD of Technical Sciences, Associate Professor

P.G. Melnik
Postgraduate student of the department automated DONNTU control systems
illuzium1999@gmail.com

Abstract: A mechanism for polygonal filling of local minima for a multi-agent system has been developed, replacing the traditional construction of planes or Gaussian spheroids. This mechanism guarantees the integrity of the group when exiting a local minimum.

Keywords: potential fields, group control, local minimum, master-slave architecture, polygonal filling, distributed pathfinding, 3D navigation, autonomous agents.

УДК 004.932.72

ИССЛЕДОВАНИЕ МЕТОДОВ КОМПЬЮТЕРНОЙ ОБРАБОТКИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ



Т.А. Васяева
Декан факультета информационных систем и технологий ДонНТУ, кандидат технических наук, доцент
tanechka.vasyaeva@yandex.ru



М.А. Рябко
Аспирант кафедры АСУ, ДонНТУ
maxryabko2706@yandex.ru

Т.А. Васяева
Кандидат технических наук, доцент, декан факультета «Информационные системы и технологии», ФГБОУ ВО «Донецкий национальный технический университет». Область научных интересов связана с исследованием и разработкой методов искусственного интеллекта, организацией учебного и научно-исследовательского процессов в техническом университете.

М.А. Рябко
Аспирант кафедры автоматизированных систем управления, ФГБОУ ВО «Донецкий национальный технический университет». Область научных интересов связана с совершенствованием методов компьютерной обработки цифровой информации в аспекте медицинских задач.

Аннотация. В статье рассматриваются современные методы компьютерной обработки медицинских изображений, применяемые для анализа цифровых микроскопических данных. Особое внимание уделяется обработке изображений высокого разрешения, полученных с использованием технологий цифровой микроскопии. Проведён обзор существующих алгоритмов сегментации и анализа медицинских изображений, включая классические методы цифровой обработки и современные нейросетевые подходы. Рассмотрены особенности обработки микроскопических изображений мазков крови, связанные с неоднородностью окрашивания и высокой плотностью объектов. Предложена методика предварительной обработки и сегментации изображений, основанная на декомпозиции исходного изображения на сегменты фиксированного размера с последующей бинаризацией.

Ключевые слова: Медицинские изображения, сегментация изображений, нейросетевые методы, микроскопические изображения, машинное обучение.

Введение. В последние годы наблюдается активное развитие информационных технологий в медицинской диагностике, сопровождающееся широким внедрением систем цифровой микроскопии и автоматизированного анализа медицинских изображений. Использование цифровых микроскопов и технологий Whole Slide Imaging (WSI) позволяет получать изображения сверхвысокого разрешения, содержащие значительный объём диагностически важной информации. Такие изображения применяются при исследовании биологических препаратов, включая гистологические и цитологические препараты, изображения, а также мазки крови. Вместе с тем увеличение объёма цифровых данных приводит к возникновению новых задач, связанных с их хранением, обработкой и анализом. Современным стандартом цифровой патологии является технология WSI, позволяющая получать электронные копии гистологических препаратов с разрешением, эквивалентным увеличению светового микроскопа (×20, ×40). Получаемые виртуальные слайды

представляют собой изображения гигапиксельного размера, что требует разработки специализированных методов их обработки, основанных на анализе пирамидальных представлений и фрагментации изображения на отдельные патчи. Высокая разрешающая способность WSI позволяет сохранять всю диагностически значимую информацию – от клеточной архитектоники до ядерной структуры, но одновременно порождает проблемы, связанные с вариабельностью окрашивания, наличием артефактов сканирования и необходимостью эффективного хранения и передачи данных. Актуальность темы обусловлена тем, что несмотря на значительное количество существующих исследований в данной области, остаются вопросы повышения устойчивости алгоритмов сегментации, оптимизации обработки изображений большого разрешения и интеграции различных методов анализа в единые вычислительные системы.

Цель и задачи работы.

Целью работы является исследование и разработка подходов к анализу медицинских изображений для повышения эффективности постановки диагноза.

Для достижения поставленной цели в работе решаются следующие задачи:

1. Анализ существующих методов предобработки гигапиксельных изображений.
2. Исследование методов бинаризации фрагментированного изображения для последующей работы с изображением.
3. Разработка метода сегментации диагностически значимых областей на гистологических препаратах;

Описание проблемы. Современные методы компьютерного зрения предлагают широкий спектр алгоритмов обработки изображений, включая фильтрацию, сегментацию, морфологический анализ и методы машинного обучения. Однако применение этих методов к медицинским изображениям связано с рядом специфических трудностей. К ним относятся неоднородность освещения, вариативность окрашивания препаратов, наличие шумов и артефактов, а также высокая плотность объектов на изображении. В рамках данного исследования были получены гистологические снимки щитовидной железы в формате NDPI. Формат NDPI представляет собой разновидность TIFF-подобного формата, используемого для хранения цифровых микроскопических изображений. Этот формат разработан для профессионального использования в областях цифровой патологии и гистопатологии, обеспечивая высокое разрешение и сохранность важных метаданных. NDPI-файлы отличаются высоким разрешением, необходимым для детального анализа микроскопических препаратов. Однако, его нестандартная структура заголовков ограничивает совместимость с большинством стандартных программ для работы с изображениями. Одной из ключевых задач анализа медицинских изображений является сегментация, то есть выделение областей интереса на изображении. Сегментация позволяет автоматически обнаруживать биологические структуры, определять их геометрические и морфологические характеристики, а также проводить количественный анализ диагностически значимых объектов. Как отмечается в ряде исследований [1, 6], именно сегментация является базовым этапом компьютерного анализа медицинских изображений и широко используется при выявлении патологических изменений и количественной оценке тканей и клеточных структур. Таким образом, задачей настоящего исследования является разработка и экспериментальное исследование методов компьютерной обработки медицинских изображений, включающих этапы предварительной обработки, сегментации и морфометрического анализа объектов.

Математическая постановка задачи. Пусть исходное изображение представлено в виде функции яркости:

$$I(x, y),\ (x, y) \in \Omega \qquad (1)$$

где I (x, y) – интенсивность пикселя изображения, Ω – область изображения.

В связи с большим размером исходного изображения (формат NDPI) и сложностью его обработки, выполняется его разбиение на набор непересекающихся сегментов:

$$\Omega = \bigcup_{i-1}^{N} \Omega_i \quad (2)$$

где N – общее количество изображения, Ω_i – i-й сегмент изображения. Это позволяет значительно снизить сложность обработки и перейти к анализу локальных частей изображения.

Каждому сегменту соответствует подизображение, которое так же описывается формулой 1.

Задача сегментации формулируется как преобразование изображения I (x, y) в бинарную маску S (x, y), разделяющую пиксели на классы «объект» и «фон» с использованием пороговой функции T (x, y).

$$S_i(\mathrm{x},\mathrm{y}) \begin{cases} 1, & I_i(x,y) > T(x,y); \\ 0, & I_i(x,y) \le T(x,y); \end{cases} \quad (3)$$

В зависимости от выбранного метода сегментации пороговая функция может задаваться различными способами.

Выбор инструментальных средств и алгоритма работы. Для эффективной работы с NDPI существуют специальные программы, такие как: программное обеспечение Hamamatsu для просмотра и анализа изображений, Aperio Technologies ImageScope, кросс-платформенная утилита NetScope Viewer, открытый проект OpenSlide и популярная платформа QuPath, которая поддерживает разнообразные форматы, включая NDPI. В таблице 1 приведена сравнительная характеристика программных продуктов, работающих с медицинскими изображениями.

Таблица 1. Сравнительная характеристика программных продуктов, работающих с медицинскими изображениями

Программа	QuPath	Aperio ImageScope	NetScope Viewer
Стоимость	Бесплатно	Платная лицензия	Платная лицензия
Доступность	Открытый исходный код	Покупка	Покупка
Поддерживаемые форматы	Много форматов (TIFF, PNG, NDPI, SVS, MRXS)	Специализированные форматы (SVS, MRXS)	Стандартные форматы (TIFF, PNG)
Экспорт в стандартные форматы	Поддерживается	Частично поддерживается	Поддерживается
Встроенный редактор кода	Поддерживается (Python, Groovy)	Ограниченный	Не поддерживается
Классификация клеток и аннотирование изображения	Поддерживается	Частично поддерживается	Не поддерживается

Анализ таблицы 1 позволяет сделать выбор QuPath для работы с полученными изображениями. Выбранная программа является специализированным редактором

изображений, которая используется врачами, исследователями и специалистами лабораторной диагностики, а также позволяет проводить различные манипуляции с медицинскими изображениями.

На рисунке ниже представлен гистологический снимок щитовидной железы, полученный с помощью микроскопического исследования. Различная интенсивность окрашивания тканей позволяет выделить клетки, которые отмечены фиолетовым цветом.

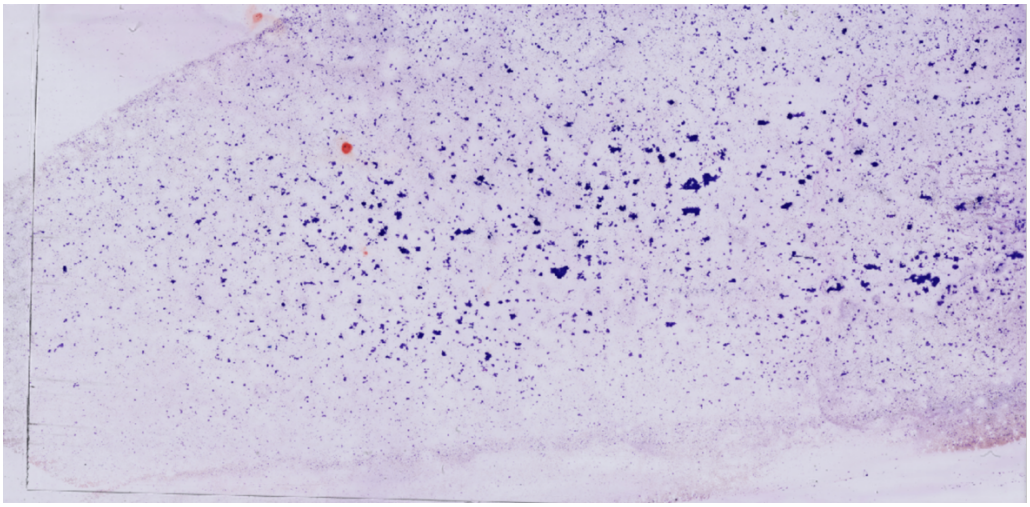


Рисунок 1. Изображение гистологического снимка щитовидной железы

На рисунке 2 представлена схема обработки изображений как последовательность действий.

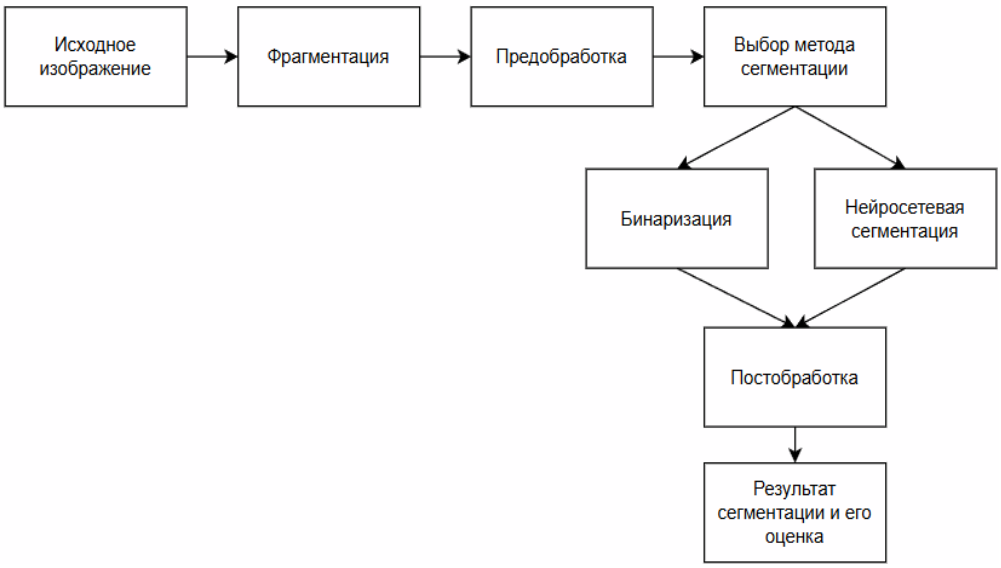


Рисунок 2. Предложенная схема обработки медицинских изображений

Предварительная обработка и фрагментация изображения. Реализация предложенного метода включает последовательность этапов обработки изображения, начиная с фрагментации и заканчивая оценкой качества сегментации. На этапе сегментации предусмотрен выбор одного из нескольких методов, включая классические алгоритмы бинаризации и нейросетевой подход. Важным этапом является подбор параметров алгоритмов, таких как пороговые значения при бинаризации, параметры фильтрации и

архитектура нейросети, что позволяет адаптировать метод к особенностям обрабатываемых изображений. До начала работы с изображением была проведена его предварительная обработка, включающая в себя уборку шума и артефактов изображения. Этап предобработки был выполнен с помощью созданного программного кода на Python.

Следующим этапом в обработке медицинского изображения является фрагментация. Необходимость фрагментации изображения обуславливается специфичностью формата изображения и его разрешением. Современные нейронные сети и прочие инструменты для работы с гистологическими снимками не предназначены для работы с таким форматом и размером данных. Для фрагментации изображения с помощью встроенного редактора кода, был написан скрипт на языке Python, который позволяет разделить исходное изображение на 2500 отдельных более мелких изображений в формате .png, что значительно упрощает дальнейшую работу, например выделение зон интересов или создание обучающей и тестовой выборок для обучения нейронной сети, которые будут разработаны для аннотирования изображения. Количество изображений, полученных в результате фрагментации, задается вручную, однако, данное количество фрагментов обеспечивает хорошее соотношение между уровнем детализации и необходимостью масштабирования результатов на полномасштабные анализы, что важно для воспроизведения результатов и будущих исследований. Полученное фрагментированное изображение, а также результат (фрагмент) работы написанного скрипта показаны на рисунках 3-4.

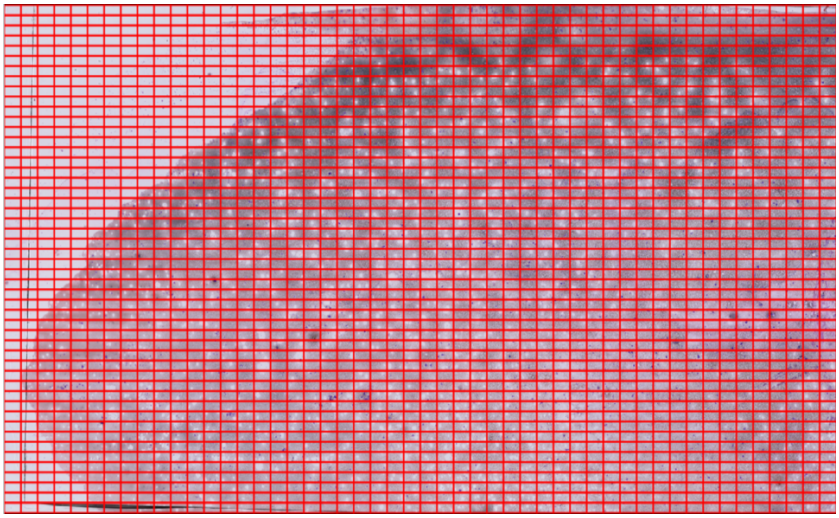


Рисунок 3. Фрагментированный гистологический снимок щитовидной железы

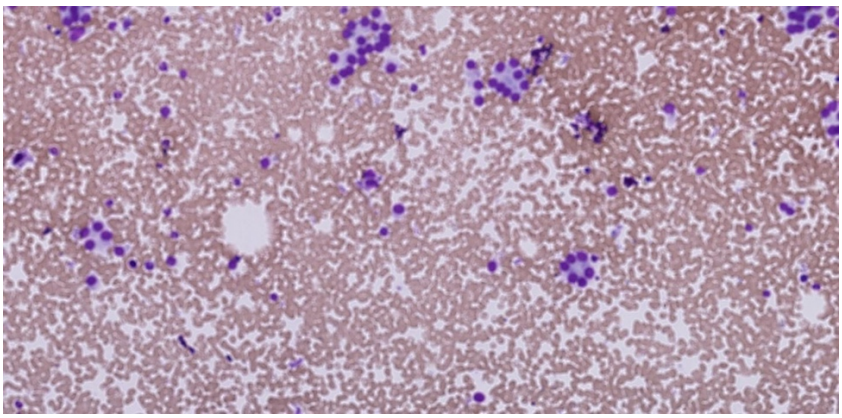


Рисунок 4. Полученный фрагмент исходного изображения гистологического снимка щитовидной железы

Применение классических методов сегментации медицинских изображений. В качестве базового подхода к сегментации были рассмотрены методы бинаризации, основанные на пороговой обработке изображений. Бинаризация позволяет преобразовать изображение в двухуровневое представление, что упрощает выделение объектов. Это позволяет перейти от сложного изображения к четко выделенным структурам. После проведения бинаризации будет доступен более упрощенный подсчет объектов, например количество клеток или плотность объектов, также можно более точно рассчитать площадь объектов и распределение клеток на выбранном изображении.

Для проведения исследования было выбрано 3 наиболее популярных метода бинаризации изображения, а именно:

1. Глобальная пороговая бинаризация;
2. Метод Оцу;
3. Адаптивная бинаризация.

После выполнения бинаризации на изображении формируется множество объектов, представляющих собой связные компоненты бинарной маски. Далее, для каждого из объектов проводится морфометрический анализ, при котором вычисляется площадь объектов, их количество, средняя площадь объектов на сегменте и другие характеристики изображения. Полученные характеристики используются для количественной оценки результатов сегментации и сравнения различных методов бинаризации. Для исследования методов бинаризации было выбрано 200 сегментов исходного изображения, после чего был написан скрипт на языке Python, который обрабатывает массив сегментированных фотографий поочередно каждым из перечисленных выше методов бинаризации. Выбранное количество сегментов для проведения экспериментов является достаточным для проведения эксперимента по следующим причинам: использование 200 сегментов позволило охватить достаточное количество различных вариантов изображений, включая нормально окрашенные фрагменты, недоокрашенные, переокрашенные и поврежденные участки, а так же позволило уменьшить аппаратные затраты и гарантировало получение приемлемой статистики для подтверждения выводов по выбору методов обработки медицинского изображения. После завершения работы программного модуля, обработанные изображения сохраняются в отдельную директорию, а также создается таблица, содержащая количественные характеристики сегментации для различных методов бинаризации. Рассмотрим полученные результаты (таблица 2) на примере сравнения методов бинаризации, для изображения, показанного на рисунке 5:

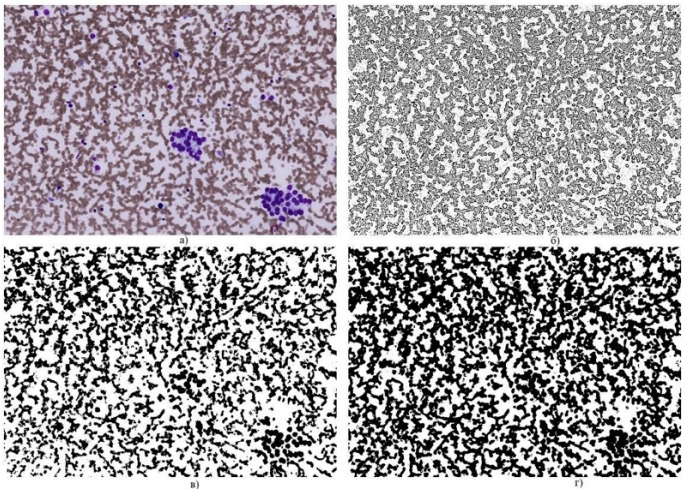


Рисунок 5. Сравнение методов бинаризации на примере сегмента исходного изображения, где а) – исходное изображение, б) – адаптивная бинаризация, в) – глобальная пороговая бинаризация, г) – метод Оцу.

Полученная таблица содержит следующие столбцы: имя изображения, метод бинаризации и количество объектов. Увеличение количества выделенных объектов свидетельствует о более высокой чувствительности метода к локальным изменениям интенсивности изображения.

Таблица 2. Сравнительная таблица параметров бинаризации

Имя изображения	Метод бинаризации	Количество объектов
tile_25.png	Глобальная пороговая	481
tile_25.png	Оцу	717
tile_25.png	Адаптивный	2538

Проведённый эксперимент позволил выявить особенности применения различных методов бинаризации при обработке микроскопических изображений. Результаты показывают, что выбор метода пороговой обработки оказывает существенное влияние на количество выделенных объектов и их морфометрические характеристики.

Глобальная пороговая бинаризация продемонстрировала наименьшее количество выделенных объектов. Метод Оцу показал более устойчивые результаты к вариациям яркости и контраста изображения, обеспечивая автоматический выбор порогового значения на основе статистических характеристик изображения. Адаптивная бинаризация позволяет учитывать локальные особенности изображения, что приводит к увеличению количества выделенных объектов. Однако данный метод также более чувствителен к шумам и может выделять дополнительные мелкие структуры, не относящиеся к биологическим объектам.

Проведённые эксперименты показали, что классические методы бинаризации не обеспечивают устойчивого выделения объектов на медицинских изображениях вследствие вариативности цветовых характеристик и неоднородности фона.

Применение нейросетевых методов сегментации медицинских изображений. Для повышения качества сегментации была разработана модель на основе сверточной нейронной сети [2, 7]. Обучение модели осуществлялось на наборе вручную аннотированных изображений, полученных с использованием специализированных инструментов разметки в программе QuPath. Пример аннотированного изображения показан на рисунке 6.

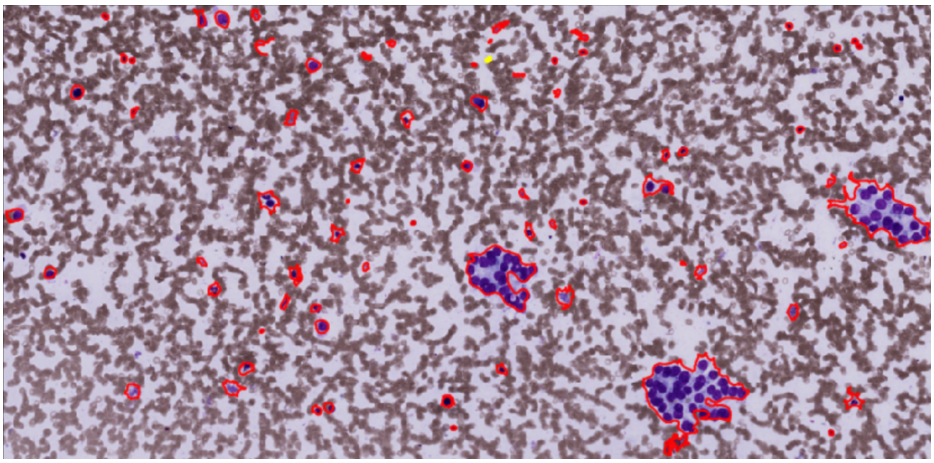


Рисунок 6. Пример аннотированного изображения

Нейросетевая модель реализует отображение входного изображения в бинарную маску, в которой пиксели, соответствующие объектам интереса, выделяются единичными значениями, а фон – нулевыми. Архитектура модели включает последовательность

сверточных слоёв, обеспечивающих извлечение пространственных и цветовых признаков, а также слоёв восстановления разрешения, позволяющих формировать выходное изображение того же размера, что и входное. Выбор сверточной архитектуры обусловлен её способностью эффективно работать с изображениями, учитывать локальные зависимости и автоматически извлекать информативные признаки без явного задания порогов [4]. В реализованной модели количество входных слоёв – 3, выходных слоёв – 1, уровней свертки – 2. Архитектура модели включает два основных блока: кодирующий и декодирующий. Кодирующая часть состоит из последовательности сверточных слоёв с функцией активации ReLU и операцией подвыборки MaxPooling с шагом 2, что позволяет уменьшить пространственное разрешение и выделить высокоуровневые признаки изображения. Функция активации выходного слоя – сигмоидальная. Декодирующая часть реализует восстановление пространственного разрешения с использованием транспонированных сверток (ConvTranspose), обеспечивая формирование сегментационной маски.

В качестве функции потерь использована бинарная кросс-энтропия, позволяющая оценивать расхождение между предсказанной и эталонной масками. Обучение модели осуществлялось на наборе аннотированных изображений, с использованием оптимизатора Adam. Для оценки качества сегментации использовались метрики Dice (показывает, насколько хорошо предсказанная маска совпадает с истинной маской) и IoU (оценивает степень пересечения между предсказанной и реальной областями сегментации), позволяющие количественно оценить степень совпадения предсказанных и эталонных областей. Результат работы нейронной сети в сравнении с исходным фрагментом показаны на рисунке 7.

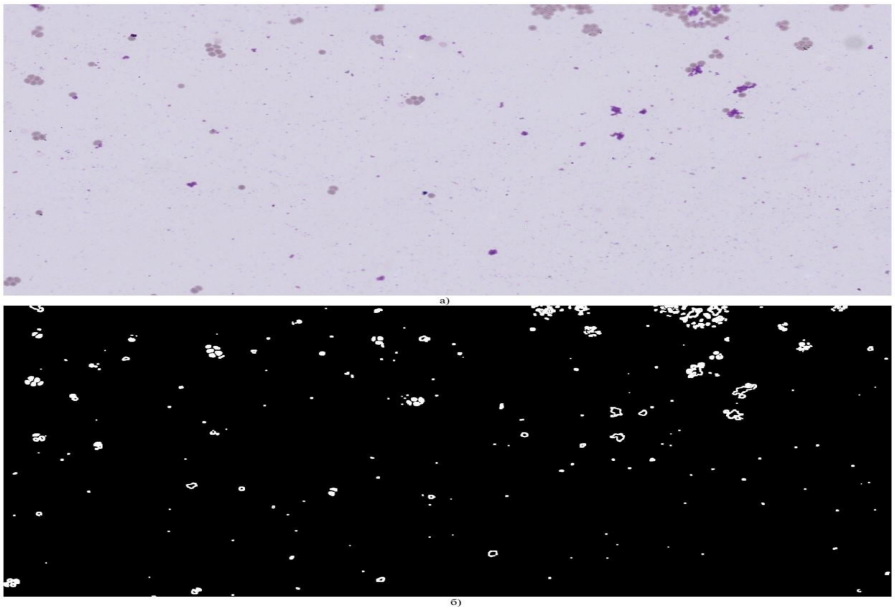


Рисунок 7. Результат работы нейронной сети, где а) – исходное изображение, б) – полученная в результате работы нейронной сети бинарная маска

В результате применения нейросетевого подхода удалось добиться более устойчивого и точного выделения объектов по сравнению с классическими методами.

Заключение. В ходе проведенного исследования были рассмотрены современные методы компьютерной обработки медицинских изображений.

Проведен обзор методов сегментации, включая традиционные методы цифровой обработки и нейросетевые модели. Разработан и реализован алгоритм предварительной

обработки изображений, заключающийся в декомпозиции исходного изображения на фрагменты фиксированного размера с последующим проведением сегментации.

В рамках исследования были реализованы и протестированы классические и нейросетевые подходы к сегментации изображений.

Сравнительный анализ показал, что нейросетевой подход обеспечивает более точное и стабильное выделение объектов, в то время как методы бинаризации могут использоваться в качестве базового или предварительного этапа обработки. Полученные результаты подтверждают целесообразность применения сверточных нейронных сетей для анализа медицинских изображений.

Список литературы

[1] Litjens G., Kooi T., Bejnordi B. et al. A survey on deep learning in medical image analysis // Medical Image Analysis. – 2017. – Vol. 42. – P. 60–88.

[2] Ермоленко С. В., Каширина И. Л., Старичкова Ю. В. Постобработка результатов сегментации медицинских изображений // Computational nanotechnology. 2025. №2.

[3] Горбунова А. В., Шмакова Я. В., Калугина О. Ф., Прохоров М. В., Бобров А. И., Кошечкин К. А. Использование методов компьютерного зрения и больших языковых моделей для проведения доклинических исследований // Медицина. 2024. №3.

[4] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). – 2015. – P. 234–241.

[5] Komura D., Ishikawa S. Machine learning methods for histopathological image analysis // Computational and Structural Biotechnology Journal. – 2018. – Vol. 16. – P. 34–42.

[6] Гонсалес Р., Вудс Р. Цифровая обработка изображений / Р. Гонсалес, Р. Вудс; пер. с англ. – М.: Техносфера, 2012. – 1104 с.

[7] Campanella G., Hanna M., Geneslaw L. et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images // Nature Medicine. – 2019. – Vol. 25. – P. 1301–1309.

Авторский вклад

Васяева Татьяна Александровна – руководство исследованием методов цифровой обработки медицинских изображений.

Рябко Максим Алексеевич – постановка задачи исследования, изучение задачи исследования, разработка решения проблемы, исследование методов и анализ полученных результатов.

RESEARCH OF COMPUTER PROCESSING METHODS FOR MEDICAL IMAGES

T.A. Vasiaeva
Dean of the Faculty of Information Systems and Technologies of DNTU, PhD of Technical sciences, Associate Professor

M.A. Riabko
Graduate student of the Department of Automated Control Systems, DNTU

Abstract. The article discusses modern methods of computer processing of medical images, which are used for the analysis of digital microscopic data. Special attention is paid to the processing of high-resolution images obtained using digital microscopy technologies. The review of existing algorithms for segmentation and analysis of medical images is carried out, including classical methods of digital processing and modern neural network approaches. The features of processing microscopic images of blood smears, associated with the heterogeneity of staining and high density of objects, are considered. A method for image preprocessing and segmentation based on decomposing the original image into fixed-size segments, followed by object detection and morphometric analysis, has been proposed.

Keywords: Medical images, image segmentation, neural network methods, microscopic images, machine learning.

УДК 004.054

ИССЛЕДОВАНИЕ ВОЗМОЖНОСТЕЙ СЕРВИСА YANDEX LOAD TESTING ДЛЯ ОЦЕНКИ ПРОИЗВОДИТЕЛЬНОСТИ ВЕБ-СЕРВИСОВ



К.А. Кошлатый
Студент третьего курса факультета компьютерного проектирвания БГУИР
alfnish159@gmail.com@gmail.com



И.В. Лихтар
Студент третьего курса факультета компьютерного проектирвания БГУИР
llikhtarilya141112005@gmail.com



А.Ю. Ефремова
Ассистент кафедры проектирования информационно-компьютерных систем БГУИР
al617e13@gmail.com

К.А. Кошлатый
Обучается в Белорусском университете информатики и радиоэлектроники. Область научных интересов связана с анализом данных и методами тестирования.

И.В. Лихтар
Обучается в Белорусском университете информатики и радиоэлектроники. Область научных интересов связана с исследованием методологий разработки.

А.Ю. Ефремова
Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с обеспечением информационной безопасности и надежного функционирования цифровых систем.

Аннотация. В статье рассматривается применение Yandex cloud testing как эффективного инструмента для оценки производительности веб-сервисов.

Ключевые слова: тестирование, тест-кейсы, оптимизации проектных решений, облачное тестирование.

Введение. В условиях современного мира и гонки технологий все большее количество пользовательских и бизнес-процессов переноситься в веб среду. Веб-сервисы используются повсеместно: магазины, поликлиники, аптеки, военная индустрия. При высокой нагрузке из-за множества подключений и больших объемов данных от сервиса требуется стабильность работы и приемлемое время отклика.

Объектом тестирования является веб-сервис для аналитики и поддержания бизнес-процессов тату салона. Сервис построен на архитектуре микросервисов с архитектурным стилем RESTful API.

Целью работы является изучение теоретических основ нагрузочного тестирования и анализ возможностей современных облачных средств, применяемых для проверки производительности веб-сервисов.

Теоретические основы тестирования. Нагрузочное тестирование рассматривается как часть более обширного тестирования производительности, при котором система находится под ожидаемой и/или целевой нагрузкой. При этом оцениваются ее устойчивость и стабильность: выдержит ли сервис высокую нагрузку и большой наплыв данных или

пользователей. В современных условиях все большую популярность приобретают облачные сервисы тестирования. Их применения позволяет сократить время тестирование, что позволяет снизить временные и материальные затраты. Они дают возможность запускать тесты без развертывания сложной локальной архитектуры, быстро масштабировать объемы нагрузки и получать результаты в простой для понимания форме.

Тестирование производительности помогает собрать измеримые данные о рабочей нагрузке. Ранний запуск тестов помогает создавать рабочие нагрузки в соответствии с правильными спецификациями [1].

Нагрузочное тестирование является одним из ключевых видов тестирования производительности и применяется для оценки поведения программной системы при ожидаемой рабочей нагрузке. Его основная задача состоит в том, чтобы определить, способна ли система стабильно функционировать при нормальном и близком к пиковому количестве пользователей, запросов или операций.

Облачное нагрузочное тестирование представляет собой подход, при котором генерация нагрузки, управление тестовыми сценариями и сбор данных происходят с использованием облачной инфраструктуры.

Одной из главных особенностей облачного нагрузочного тестирования является масштабируемость. При использовании облачной платформы объем генерируемой нагрузки может увеличиваться значительно быстрее, чем при локальном подходе, поскольку пользователю не требуется самостоятельно подготавливать дополнительные технические средства, настраивать распределение нагрузки и поддерживать работоспособность всей тестовой среды.

Еще одной особенностью является централизованный сбор и визуализация результатов. Облачные решения, как правило, предоставляют отчеты, графики и панели мониторинга, в которых отражаются ключевые показатели теста: время отклика, число успешных и ошибочных запросов, интенсивность нагрузки и поведение системы во времени.

Отдельно следует отметить, что облачный подход особенно удобен в учебных, исследовательских и прикладных задачах, где важно быстро получить результат без значительных затрат на поддержку собственной тестовой инфраструктуры.

Можно выделить несколько требований к средам нагрузочного тестирования:

– масштабируемость;

– система аналитики и мониторинга;

– возможность интеграции с CI/CD [2]

Именно совокупность этих критериев и показывает, насколько инструмент пригоден для объективной оценки сервиса.

Yandex Load Testing представляет собой облачный сервис, предназначенный для проведения нагрузочного тестирования с возможность создания настроенного генератора нагрузки в облаке, запуска тестов через интерфейс, просматривания отчетов и хранения результатов предыдущих запусков для последующего сравнения. Практическая ценность сервиса заключается в том, что он объединяет в одной среде подготовку тестовой инфраструктуры, выполнение нагрузочного сценария и анализ полученных результатов. Пользователь получает возможность не только сгенерировать нужный поток запросов к тестируемому веб-сервису, но и отслеживать поведение системы по ключевым метрикам, включая время отклика, квантили, коды ответов и динамику нагрузки.

Архитектура Yandex cloud testing состоит из нескольких основных компонентов: тестируемый сервис, агент, конфигурация теста и payload.

Тестируемый сервис является целевым объектом, на который направляется нагрузка. Это может быть веб-приложение, API, HTTP(S)-сервис или иная система, доступная агенту [3].

Агент же является виртуальной машиной, на которой размещены инструмент нагрузочного тестирования, генераторы нагрузки и коннектор для взаимодействия с сервисом Yandex Load Testing.

Сервис был выбран ввиду ряда преимуществ, описанных далее.

Основное преимущество –облачный характер сервиса, позволяющий выполнять работу гораздо более эффективно и нетребовательно к комплектующим технического средства, на котором мы работаем.

Второе – встроенная аналитика, благодаря которой сервис хорошо подходит для анализа производительности backend-компонентов, но не заменяет инструменты UI/E2E-тестирования.

Третье – поддержка разных генераторов и сценариев. Пользователь может применять Pandora, Phantom и, по сервисной странице, JMeter, выбирать формат тестовых данных и использовать файлы из Object Storage.

Проведение исследования. Перед проведением нагрузочного тестирования необходимо подготовить тестовое окружение в Yandex Cloud. Для работы с Yandex Load Testing требуется тестируемый сервис, агент генерации нагрузки, конфигурация теста и тестовые данные. В рамках исследования в качестве тестируемого приложения рассматривается информационная система тату-салона с аналитическим модулем, а генерация нагрузки должна выполняться с помощью отдельного агента Load Testing.

Подготовка окружения включает создание сервисного аккаунта для работы агента, настройку сети и создание виртуальной машины агента в сервисе Load Testing. В документации Yandex Cloud для агента указывается сервисный аккаунт, а также подсеть с настроенным NAT-шлюзом; для взаимодействия агента с сервисом требуется исходящий доступ по порту 443 (рисунок 1).

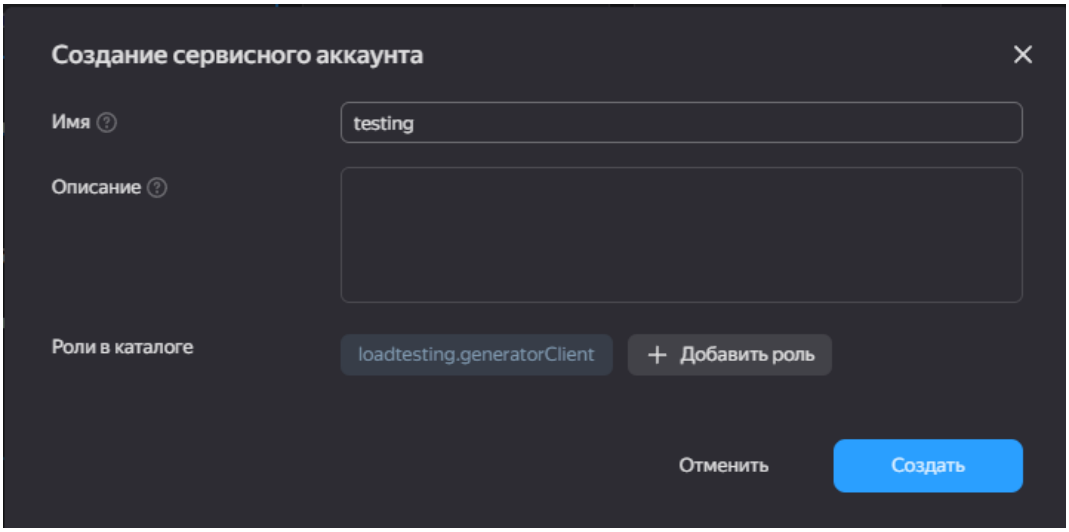


Рисунок 1. Интерфейс создания сервисного аккаунта

При подготовке окружения также важно учитывать параметры самого агента, поскольку от его ресурсов зависит корректность генерации нагрузки. Мониторинг состояния агента может быть использован для контроля загрузки процессора, памяти, сети и других параметров во время тестирования.

Yandex Cloud предоставляет несколько типов агентов, отличающихся вычислительными характеристиками: small, medium и Large. Каждый отличает по количеству памяти и вычислительным способностям. Small обладает наименьшим количеством оперативной памяти и объемом хранилища.

Он подходит для небольших обучающих проектов. Medium уже более продвинутый агент, с ним можно вести тестирование небольших коммерческих проектов. Large уже обладает внушительными объемами памяти, с которым можно смело тестировать большие коммерческие проекты со сложной структурой и большим количеством внутренних процессов.

Это необходимо для того, чтобы источник нагрузки сам не стал ограничивающим фактором эксперимента.

В контексте исследования выбран агент со средними характеристиками (рисунок 2).

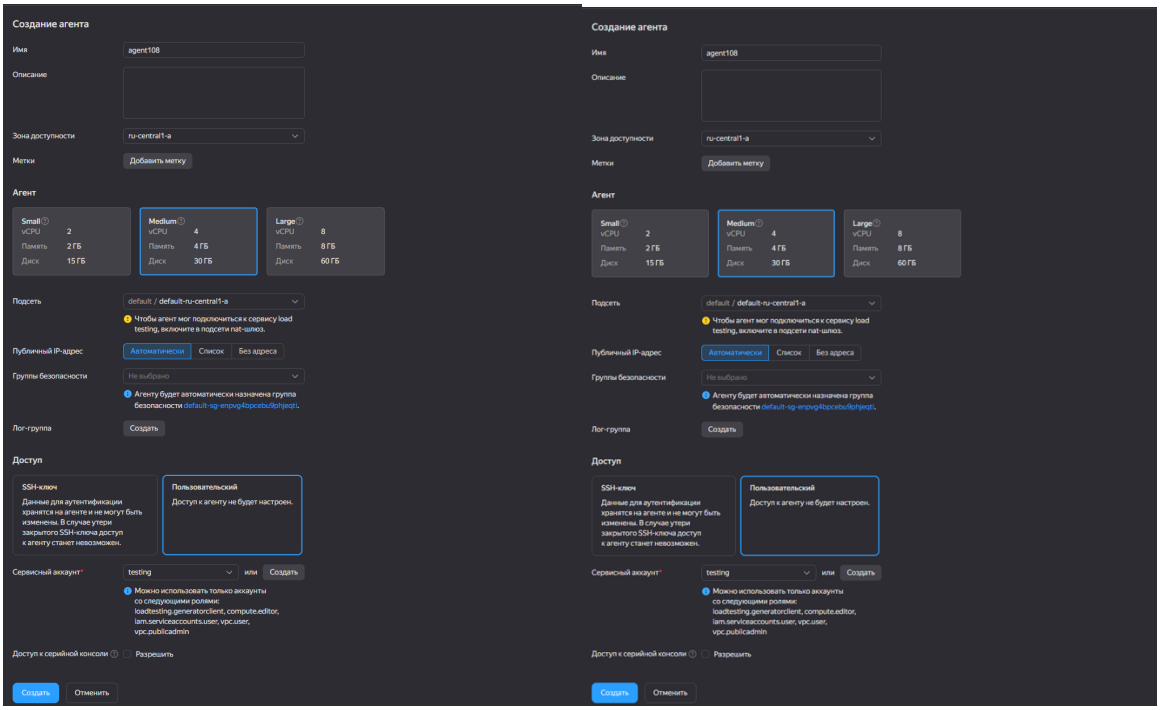


Рисунок 2. Интерфейс создания агента

Таким образом, на этапе подготовки тестового окружения формируется инфраструктурная основа для дальнейшей настройки нагрузочного теста: определяется тестируемый сервис, создается агент генерации нагрузки, настраиваются сетевые параметры и обеспечивается техническая готовность среды к проведению эксперимента.

На этапе настройки теста в Yandex Load Testing выбирается ранее созданный агент и задаются основные параметры нагрузочного тестирования.

Сервис позволяет создавать тест через интерфейс управления, где указываются цель тестирования, генератор нагрузки, параметры нагрузки и дополнительные настройки. Для первого сценария целесообразно использовать настройку через форму, так как она позволяет задать основные параметры без подготовки сложной конфигурации.

Дополнительно при настройке теста задаются тестовые данные, количество тестирующих потоков, время принудительного завершения теста и параметры мониторинга (рисунок 3).

Эти настройки необходимы для того, чтобы нагрузка формировалась контролируемо, а состояние агента можно было отслеживать в процессе выполнения теста.

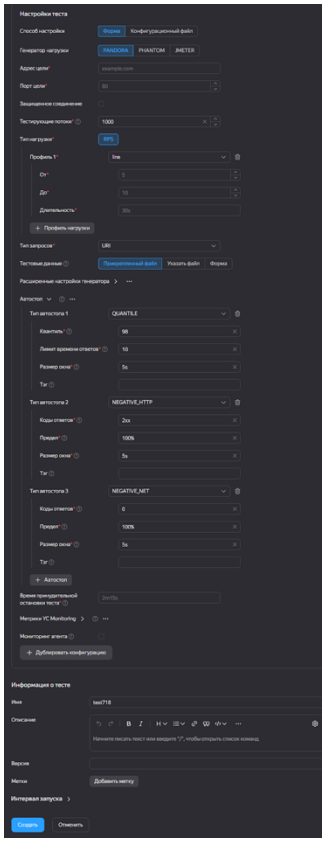


Рисунок 3. Создание и конфигурация теста

Для более полной оценки поведения информационной системы тату-салона нагрузочное тестирование целесообразно проводить не в рамках одного запуска, а в виде серии прогонов с различными параметрами нагрузки. Такой подход позволяет сопоставить поведение системы при умеренной, повышенной и постепенно возрастающей интенсивности запросов, а также подготовить данные для последующего анализа времени отклика, ошибок и устойчивости работы (таблица 3).

Таблица 3.1 – Серия прогонов нагрузочного тестирования

№ прогона	Название теста	Характер нагрузки	Назначение
1	tattoo-backend-smoke	Базовая проверка	Проверка доступности сервиса и корректности обработки запросов
2	tattoo-backend-const-low	Низкая постоянная нагрузка	Оценка работы системы в штатном режиме
3	tattoo-backend-line-growth	Линейно возрастающая нагрузка	Анализ изменения поведения сервиса при росте интенсивности запросов
4	tattoo-backend-step-load	Ступенчатая нагрузка	Выявление реакции системы на резкое увеличение нагрузки

Проведение серии прогонов также важно с точки зрения сопоставления результатов. Использование нескольких запусков с разными профилями нагрузки позволяет определить, при каких условиях начинают проявляться признаки деградации производительности, и подготовить основу для дальнейшего сравнения полученных показателей в следующей главе.

В рамках анализа результатов нагрузочного тестирования информационной системы тату-салона основное внимание уделялось поведению серверной части приложения при различных вариантах нагрузки. Для оценки использовались изменение времени отклика, характер кодов ответов и общая устойчивость работы сервиса. Такой подход позволяет рассматривать производительность системы комплексно.

При базовом сценарии сервис работает стабильно: сохраняется приемлемое время отклика и корректная обработка запросов. Это свидетельствует о том, что при невысокой интенсивности обращений серверная часть приложения справляется с нагрузкой без выраженных признаков деградации.

При постоянной низкой нагрузке система также остается устойчивой, что позволяет оценить ее работу в условиях длительной обработки однотипных запросов.

Более показательными являются сценарии с увеличением нагрузки. При линейном росте интенсивности запросов возрастает нагрузка на серверную логику и базу данных, из-за чего время отклика может постепенно увеличиваться. При ступенчатой нагрузке такие изменения проявляются еще заметнее, так как сервис должен адаптироваться к более резкому росту числа обращений.

Дополнительное значение имеет характер кодов ответов. Преобладание успешных ответов указывает на корректную работу сервиса, тогда как появление отдельных ошибок может свидетельствовать о росте нагрузки на приложение, базу данных или инфраструктуру. В совокупности с изменением времени отклика это дает общее представление о поведении серверной части проекта под нагрузкой.

Для обобщения результатов была составлена сводная таблица, отражающая поведение сервиса при разных вариантах тестирования (таблица 4).

Таблица 4. Сводные результаты нагрузочного тестирования информационной системы

Название теста	Профиль нагрузки	Время отклика	Коды ответов	Общая оценка результата
tattoo-backend-smoke	базовая проверка	стабильное, низкое	преобладают успешные ответы	сервис работает корректно, признаков перегрузки не наблюдается
tattoo-backend-const-low	постоянная низкая нагрузка	стабильное, с незначительными колебаниями	в основном успешные ответы	система сохраняет устойчивость при длительной умеренной нагрузке
tattoo-backend-line-growth	линейно возрастающая нагрузка	умеренно увеличивается по мере роста нагрузки	сохраняется преобладание успешных ответов, возможны единичные отклонения	сервис в целом устойчив, однако заметна зависимость времени отклика от интенсивности запросов
tattoo-backend-step-load	ступенчатая нагрузка	заметно возрастает на отдельных этапах	возможны отдельные ошибки при повышении нагрузки	при резком увеличении нагрузки наблюдается ухудшение производительности, что указывает на чувствительность системы к интенсивным обращениям

На основании данных, представленных в таблице, можно сделать вывод, что сервис в наибольшей степени сохраняет устойчивость при базовой и умеренной постоянной нагрузке. При более сложных сценариях, связанных с постепенным или ступенчатым увеличением интенсивности запросов, система продолжает функционировать, однако демонстрирует более выраженную зависимость времени отклика от уровня нагрузки. Это свидетельствует о том, что серверная часть приложения пригодна для работы в стандартных режимах, но при росте интенсивности обращений ее производительность начинает снижаться.

Таким образом, Yandex Load Testing можно считать подходящим средством для оценки производительности и устойчивости серверной части информационной системы тату-салона.

Заключение. Анализ сервиса Yandex Load Testing показывает, что он соответствует основным требованиям, предъявляемым к современным средствам нагрузочного тестирования. Yandex Load Testing наиболее эффективен при оценке реакции веб-сервиса на рост числа запросов, определения пределов его стабильной работы и выявления потенциальных узких мест.

Список литературы

[1] Recommendations for performance testing // Microsoft Learn. – 2025. – URL: https://learn.microsoft.com/en-us/power-platform/well-architected/performance-efficiency/performance-test (дата обращения: 19.03.2026).

[2] Features // Distributed Load Testing on AWS. – AWS Documentation. – URL: https://docs.aws.amazon.com/solutions/latest/distributed-load-testing-on-aws/features.html (дата обращения: 19.03.2026).

[3] Yandex Load Testing overview // Yandex Cloud Documentation. – 2025. – URL: https://yandex.cloud/en/docs/load-testing/concepts/ (дата обращения: 20.03.2026).

Авторский вклад

Кошлатый Кирилл Александрович – проведение исследования методик нагрузочного тестирования, анализ теоретических основ нагрузочного тестирования, изучение особенностей облачного нагрузочного тестирования и требований к современным средствам нагрузочного тестирования, участие в подготовке выводов по применению Yandex Load Testing.

Лихтар Илья Владимирович – анализ возможностей сервиса Yandex Load Testing, рассмотрение его архитектуры, основных компонентов, преимуществ и ограничений, формулирование итоговых выводов по применению сервиса для нагрузочного тестирования веб-сервисов.

Ефремова Александра Юрьевна – руководство исследованием и постановка задачи по изучению возможностей сервиса для оценки производительности веб-сервисов.

EVALUATING WEB SERVICE PERFORMANCE USING YANDEX LOAD TESTING

K.A. Koshlatyi
Student of the department Information and Computer-Aided Systems Design, BSUIR

I.V. Likhtar
Student of the department Information and Computer-Aided Systems Design, BSUIR

A. Yu. Yafremava
Assistant of the department Information and Computer-Aided Systems Design, BSUIR

Abstract. The article examines the use of Yandex Cloud Testing as an effective tool for evaluating the performance of web services.

Keywords: testing, test cases, optimization of design solutions, cloud testing.

UDC 004.6:004.9

APPLICATION OF BIG DATA AND ADVANCED ANALYTICS FOR OPTIMIZATION OF IT SOLUTIONS



O. P. Berdiyeva
The institute of Telecommunications and informatics of Turkmenistan, lecturer
bibi.ber.va@gmail.com

O. P. Berdiyeva
Graduated from the Turkmen National Institute of World Languages named after Dovletmammet Azadi. He works at the Institute of Telecommunications and Informatics in Turkmenistan.

Abstract. This article explores the transformative potential of Big Data and Advanced Analytics in enhancing the efficiency and performance of modern IT solutions. By analysing large-scale datasets and employing predictive modelling, organizations can significantly optimize their infrastructure, reduce latency, and improve overall system reliability. The study highlights key methodologies for integrating data-driven insights into the software development lifecycle to achieve superior technical outcomes.

Keywords: Big Data, Advanced Analytics, IT solution optimization, predictive modelling, data-driven insights, system performance.

Introduction

In the era of rapid digital transformation, IT systems generate vast amounts of data that hold critical insights for operational excellence. Leveraging Big Data and Advanced Analytics allows for a shift from reactive troubleshooting to proactive optimization. This paper examines how algorithmic solutions and machine learning models can be applied to refine IT processes and deliver more robust software architectures.

Main Content

1. Data-Driven Architecture Optimization

Modern IT solutions require dynamic architectures that can adapt to changing workloads. Big Data analytics enables the transition from static configurations to data-driven, elastic infrastructures. By analyzing historical traffic patterns, IT systems can implement Auto-scaling algorithms that adjust server capacity in real-time. This reduces operational costs by ensuring that resources are only consumed when necessary, thereby optimizing the overall cost-to-performance ratio.

2. Machine Learning Algorithms in IT Operations (AIOps)

Advanced Analytics integrates Machine Learning (ML) into IT operations, a field known as AIOps.

- **Regression Analysis:** This is used to predict future system load based on historical data trends. For instance, an IT solution can forecast storage requirements for the next quarter with high precision.
- **Clustering Techniques:** By using k-means clustering, system administrators can group similar types of network traffic, allowing for the optimization of data packets and reducing latency in high-speed networks.

3. **Enhancing User Experience (UX) through Behavioral Analytics**

The optimization of an IT solution is incomplete without considering the end-user. Big Data allows for the collection and analysis of user interaction logs. Through Heatmap Analysis and Path Tracking, developers can identify which features of a software solution are causing friction. Advanced analytics then provides data-backed recommendations for UI/UX adjustments, leading to higher user retention and system efficiency.

4. **Real-time Monitoring and Self-healing Systems**

The ultimate goal of applying Advanced Analytics in IT is the creation of self-healing systems.

- **Root Cause Analysis (RCA):** When a failure occurs, Big Data tools can scan millions of log entries in seconds to identify the exact point of failure.
- **Automated Remediation:** Once a pattern is recognized, the system can automatically execute scripts to restart services or re-route traffic, minimizing the Mean Time to Recovery (MTTR).

5. Performance Comparison Table 1

Table 1. The following table illustrates the effectiveness of various models in IT optimization

Optimization Goal	**Analytical Model**	**Impact on Performance**
Traffic Forecasting	Time-Series Analysis	High (Predicts peak loads)
Security/Fraud	Anomaly Detection	Critical (Identifies breaches)
Storage Management	Linear Regression	Medium (Capacity planning)
System Reliability	Bayesian Networks	High (Risk assessment)

Conclusion

The integration of Big Data and Advanced Analytics is essential for the continuous improvement of IT solutions. These technologies not only streamline operations but also provide a competitive edge by ensuring that IT infrastructures are agile, resilient, and data-informed.

Reference List

[1] Smith, J. Machine Learning Applications for Generative Decision-Making. *Journal of Business Intelligence*. 2021; 15(2): 88-99.

[2] White, A. Big Data and IT Infrastructure: A Modern Approach. *Technical Review*. 2023; 10(4): 45-52.

Author's contribution

Authors make an equivalent contribution.

ПРИМЕНЕНИЕ БОЛЬШИХ ОБЪЕМОВ ДАННЫХ И ПЕРЕДОВОЙ АНАЛИТИКИ ДЛЯ ОПТИМИЗАЦИИ ИТ-РЕШЕНИЙ

О. П. Бердиева
Институт телекоммуникаций и информатики
Туркменистана, преподаватель
bibi.ber.va@gmail.com

Аннотация. В этой статье рассматривается преобразующий потенциал больших данных и продвинутой аналитики для повышения эффективности и быстродействия современных ИТ-решений. Анализируя крупномасштабные наборы данных и используя прогнозное моделирование, организации могут значительно оптимизировать свою инфраструктуру, сократить время ожидания и повысить общую надежность системы. В исследовании освещаются ключевые методологии интеграции аналитических данных на основе данных в жизненный цикл разработки программного обеспечения для достижения превосходных технических результатов.

Ключевые слова: большие данные, расширенная аналитика, оптимизация ИТ-решений, прогнозное моделирование, аналитические данные на основе данных, производительность системы.

УДК 004.94:551.583

КОНЦЕПЦИЯ ЦИФРОВЫХ ДВОЙНИКОВ И ИХ ПРИМЕНЕНИЕ ДЛЯ МОДЕЛИРОВАНИЯ КЛИМАТА



В.О. Тихоненко
Студент 3 курса специальности «Программная инженерия» факультета компьютерного проектирования
vlad6867911@gmail.com



Е.Д. Ракицкий
Студент 3 курса специальности «Программная инженерия» факультета компьютерного проектирования
egor.rakitskiy@gmail.com



Ю.А. Ёщик
Студент 3 курса специальности «Программная инженерия» факультета компьютерного проектирования
eschikjulia@gmail.com



А.В. Воробей
Асситент кафедры инженерной психологии и эргономики БГУИР, магистр технических наук
vorobey@bsuir.by

В.О. Тихоненко
Студент 3 курса специальности «Программная инженерия» факультета компьютерного проектирования.Область научных интересов: разработка архитектур программных платформ для моделирования сложных природных систем, клиент-серверные приложения, методы ассимиляции данных в реальном времени.

Е.Д. Ракицкий
Студент 3 курса специальности «Программная инженерия» факультета компьютерного проектирования.Область научных интересов: применение методов машинного обучения и нейронных сетей для построения суррогатных (клоновых) моделей в задачах прогнозирования, анализ больших данных.

Ю.А. Ёщик
Студент 3 курса специальности «Программная инженерия» факультета компьютерного проектирования.Область научных интересов: обеспечение физической согласованности прогнозов в самообучающихся системах, обработка потоков данных от IoT-сенсоров и спутниковых систем, интеграция физических моделей и ИИ.

А.В. Воробей
Окончила Белорусский государственный педагогический университет имени М. Танка» по специальности Биология и химия. Окончила аспирантуру БГУИР по специальности «Психология труда, инженерная психология, эргономика».

Аннотация. В статье рассматривается концепция программной платформы для создания самообучающихся цифровых двойников климатических явлений. Обосновывается необходимость преодоления ограничений традиционных статичных моделей, не способных адаптироваться к текущим условиям. Сформулированы три ключевых принципа построения: ассимиляция данных в реальном времени, редукция вычислительной сложности через клоновые модели и обеспечение физической согласованности прогнозов. Представлена архитектура программного решения на базе клиент-серверной модели и методы обновления цифровых двойников. Результаты работы направлены на повышение точности прогнозирования экстремальных погодных явлений и поддержку принятия решений.

Ключевые слова: цифровой двойник, климатическое моделирование, машинное обучение, ассимиляция данных, Big Data, клоновая модель, физическая согласованность, прогнозирование погоды.

Введение. Современное климатическое моделирование сталкивается с фундаментальным противоречием: физические законы, описывающие атмосферные процессы, хорошо известны, но их прямое численное решение для систем планетарного масштаба требует колоссальных вычислительных ресурсов и времени. Традиционные модели, основанные на решении систем уравнений гидротермодинамики на пространственной сетке, обеспечивают физическую строгость, но вынужденно упрощают микропроцессы и вторичные эффекты, что снижает точность прогнозов, особенно для экстремальных явлений. Дополнительно такие модели, как правило, статичны: данные и параметры фиксируются на этапе инициализации и не обновляются в процессе расчёта, что приводит к расхождению с реальным состоянием системы и ограничивает оперативность прогнозирования.

Параллельно стремительное развитие цифровизации, Интернета вещей, методов машинного обучения и анализа больших данных радикально изменило возможности наблюдения и моделирования сложных природных систем. Разворачивающиеся в реальном времени потоки разнородных данных от наземных метеостанций, спутниковых систем и других сенсоров создают предпосылки для перехода от закрытых статичных схем к динамическим моделям, способным адаптироваться к текущим условиям. В этом контексте концепция цифровых двойников, предполагающая тесно связанную физическую и виртуальную компоненты с двунаправленным обменом данными, становится естественной основой для переосмысления климатического моделирования и прогнозирования погоды [1].

Цель данной работы – представить концепцию программной платформы для создания самообучающихся цифровых двойников климатических и экстремальных погодных явлений, объединяющую физические модели, методы машинного обучения и инфраструктуру обработки больших данных. В статье формулируются ключевые принципы построения такой системы, обсуждаются сценарии её применения и описывается архитектура программного решения, ориентированного на работу с потоками данных в реальном времени и поддержку принятия решений.

Принципы построения самообучающейся модели. Создание эффективного цифрового двойника климатических и погодных явлений требует пересмотра традиционных подходов к моделированию. В отличие от статичных моделей, построенных исключительно на решении систем дифференциальных уравнений, предлагаемый подход базируется на трех фундаментальных принципах, обеспечивающих динамическую адаптацию, вычислительную эффективность и физическую достоверность прогнозов.

Первый принцип – ассимиляция данных в реальном времени – предполагает непрерывный приём и интеграцию информации от распределённой сети сенсоров (наземные метеостанции, спутниковые системы, радиолокаторы, аэрологические зонды и др.), характерной для современных климатических платформ [6-8]. В отличие от традиционных моделей, работающих в замкнутом режиме с фиксированными начальными условиями, цифровой двойник регулярно «подтягивает» своё состояние к наблюдаемой реальности, компенсируя неполноту физического описания и накопление численных ошибок. Ассимиляция включает сбор, предобработку, верификацию и интеграцию данных

в текущее состояние модели, что особенно важно при описании быстроразвивающихся экстремальных явлений – ураганов, смерчей, внезапных наводнений.

Схема обновления цифрового двойника на основе данных с IoT-сенсоров в реальном времени представлена на рисунке 1.



Рисунок 1. Схема обновления цифрового двойника

Второй принцип связан с редукцией вычислительной сложности посредством клоновых (surrogate) моделей, которые строятся на основе прогонов физической модели высокого разрешения в различных сценариях. Такой подход соответствует современным представлениям о масштабируемых предиктивных цифровых двойниках, сочетающих физические модели и методы машинного обучения. Физическая модель используется не для прямого прогноза в режиме реального времени, а для генерации представительного массива обучающих данных; на этой базе формируется быстрая surrogate-модель (например, на основе глубоких нейронных сетей), аппроксимирующая вход-выходное поведение полной модели при существенно меньших вычислительных затратах. Это обеспечивает получение оперативных прогнозов и возможность последующего дообучения модели по мере накопления новых данных.Однако при построении таких моделей возникает ряд трудностей. Традиционные методы моделирования часто основаны на проектных численных параметрах, что может приводить к большому количеству итераций проектирования. Для повышения точности модели исследователи предлагают различные стратегии: создание интеллектуальных систем на основе многомодальных датчиков, способных обновляться в реальном времени, а также применение динамического моделирования на основе данных, использующего алгоритмы глубокого обучения для обработки больших объемов информации.

Третий принцип – физическая согласованность – направлен на предотвращение появления правдоподобных, но физически некорректных прогнозов, характерных для чисто эмпирических схем. Клоновая модель должна не только минимизировать ошибку предсказания, но и соблюдать фундаментальные законы сохранения энергии, массы,

импульса и другие ограничения, что соответствует современным подходам к построению объяснимых цифровых двойников динамических систем. Современные исследования показывают, что подходы, комбинирующие физические модели и классификаторы машинного обучения, позволяют создавать интерпретируемые модели, способные не только прогнозировать, но и объяснять причины тех или иных явлений.

Различия в использовании традиционной модели анализа данных и использовании цифрового двойника показаны на рисунке 2.



Рисунок 2. Сравнение традиционных систем прогнозирования погоды с систем на основе цифровых двойников

Концептуальная модель цифрового двойника климата. Построение цифрового двойника для такой сложной и динамичной системы, как климат Земли, требует не только технических решений, но и четкого концептуального понимания того, что именно создается. В научной литературе и практических разработках существует терминологическая неопределенность, связанная с использованием понятий «цифровая тень» и «цифровой двойник». Разграничение этих категорий имеет принципиальное значение для корректной постановки задач и интерпретации получаемых результатов.

С быстрым развитием области цифровых двойников все больше ученых посвящают себя теоретическим и практическим исследованиям в этой сфере. Концептуальная близость между цифровыми тенями и цифровыми двойниками стимулировала обширные исследовательские усилия, однако важно подчеркнуть, что эти понятия имеют разную природу и играют различные роли во взаимодействии физического и виртуального миров.

Цифровая тень представляет собой цифровое изображение состояния конкретного физического объекта в определенный момент времени. По мнению Бергса, цифровые тени можно рассматривать как предшественника цифровых двойников [5]. Они в первую очередь ориентированы на сбор и анализ исторических данных, событий и моделей поведения для получения ценной информации и поддержки принятия решений. Цифровая тень не обладает обратной связью с физическим объектом – это одностороннее отображение, фиксирующее состояние объекта «как есть» в прошлом или настоящем, но не влияющее на его будущее поведение. Информационный поток в этом случае направлен исключительно от физического объекта к его виртуальному представлению.

Цифровой двойник, в отличие от цифровой тени, представляет собой полностью интегрированную и автоматизированную систему с двунаправленным потоком данных

между физическим объектом и его цифровым аналогом. Согласно определению Крицингера, это цифровое представление физического объекта, обеспечивающее автоматизированный обмен данными в обоих направлениях. Цифровой двойник способен не только воспринимать текущее состояние физического объекта, но и диагностировать его, прогнозировать будущие состояния и, в предельном случае, регулировать поведение физического объекта посредством взаимодействия и оптимизации решений. Барричелли рассматривает цифровой двойник как живую, интеллектуальную и развивающуюся модель, служащую виртуальным аналогом физического объекта или процесса. Благодаря самообучению на основе соответствующих цифровых моделей, цифровой двойник способен к автономной интеллектуальной эволюции.

Применительно к климатическим системам это разграничение приобретает особый смысл. Большинство существующих сегодня климатических моделей, включая те, что используются в международных проектах, по своей сути являются цифровыми тенями. Они оперируют историческими данными, строят ретроспективные анализы и делают прогнозы на будущее, но не имеют механизма непрерывной двунаправленной связи с реальной климатической системой в реальном времени. Данные, поступающие от метеорологических станций и спутников, используются для верификации и калибровки, но не для активной коррекции текущего состояния модели в процессе ее работы.

Сценарии моделирования. Цифровые двойники климатических систем, в зависимости от целей и временного горизонта, могут реализовывать различные сценарии моделирования. Эти сценарии охватывают как ретроспективный анализ, так и прогнозирование будущих изменений, а также исследование конкретных экстремальных явлений.

Многолетние модели предназначены для анализа долгосрочных климатических трендов и охватывают значительные временные промежутки. В рамках современных международных инициатив создаются модели, охватывающие недавнее прошлое (с 1990 года) и возможные будущие изменения климата до 2050 года. Моделирование прошлых событий используется для оценки качества модели перед применением прогноза на будущее, а также для количественной оценки того, как характеристики климата и погоды изменятся в течение следующих десятилетий. Такой подход позволяет верифицировать модели на известных данных и повышать достоверность долгосрочных прогнозов.

Сюжетные модели (storylines) представляют собой специализированный класс моделей, нацеленных на реконструкцию развития недавних экстремальных явлений. Волны жары, катастрофические наводнения, засухи, ураганы – эти события исследуются в различных климатических условиях для понимания механизмов их возникновения и развития. Сюжетные модели позволяют детально проанализировать конкретное событие, выявить ключевые факторы, способствовавшие его возникновению, и оценить, как аналогичное явление могло бы развиваться при иных начальных условиях или в изменяющемся климате. Такой подход особенно ценен для задач адаптации и планирования мер гражданской защиты.

Протоколы моделирования играют ключевую роль в обеспечении сопоставимости и надежности результатов, получаемых различными исследовательскими группами. Основой для большинства современных климатических исследований служит проект сравнения связанных моделей (CMIP – Coupled Model Intercomparison Project), который стандартизирует сценарии, входные данные и форматы выходных результатов [6]. Особое значение для данного исследования имеет проект сравнения моделей высокого разрешения (HighResMIP – High Resolution Model Intercomparison Project), нацеленный на разработку и оценку моделей с повышенным пространственным разрешением [7].

В рамках европейской инициативы Destination Earth (DestinE) эти протоколы дополнительно адаптируются для моделирования в масштабе километров, что существенно дополняет существующие климатические прогнозы [8]. Километровое разрешение

позволяет явным образом описывать процессы, которые ранее необходимо было параметризовать: конвекцию, образование облаков, взаимодействие с локальными особенностями рельефа. Важным элементом протоколов является проведение моделирования с повторяющимся воздействием 1990 года без изменения внешних воздействий с течением времени. Такие эксперименты позволяют количественно оценить дрейф модели (собственную внутреннюю изменчивость, не связанную с изменением внешних факторов) и смоделированную межгодовую изменчивость, а также предоставляют необходимый контекст для интерпретации исторических и сценарных моделирований.

Обработка данных и обновление модели. Эффективность цифрового двойника климата в значительной степени зависит от того, насколько логично и без разрывов выстроена работа с данными –от момента их получения до пересчёта параметров моделей. В крупных международных проектах по моделированию климата (CMIP, HighResMIP, инициатива Destination Earth) для этого задействуется разветвлённая наблюдательная инфраструктура: сеть наземных метеостанций, спутниковые измерения, радиолокационные комплексы, океанические платформы и другие источники [6-8].

В разрабатываемой платформе данные из этих каналов, приходящие с разной частотой и в разнородных форматах, сначала проходят единый цикл предварительной обработки, включающий фильтрацию шумов, поиск аномальных значений, восстановление пропусков и приведение к общим шкалам и форматам. Затем они объединяются в целостное описание текущего состояния климатической системы с учётом различий по пространству и времени и возможных расхождений между источниками. Для более тонкой очистки, реконструкции недостающей информации и выделения неочевидных закономерностей применяются методы машинного обучения и глубинные нейронные сети, что укладывается в современные практики работы с крупномасштабными климатическими данными [6-10].

Для поддержания актуальности цифрового двойника недостаточно однократной калибровки; требуется регулярное обновление структуры и параметров модели на основе новых данных и результатов численных экспериментов. В работе рассматриваются три группы методов: корректировка по фактическим данным, скользящее обучение и имитация обучающих наборов на основе физических моделей. В первом случае модель, предварительно настроенная по физическим соображениям, уточняется по реальным измерениям, что повышает точность, но ограничено наблюдаемостью внутренних параметров. При скользящем обучении (online/continual learning) цифровой двойник периодически дообучается на новых данных, что позволяет учитывать старение аппаратуры и изменение фоновых условий, но требует выбора рациональной частоты обновления. Подход имитации обучающих данных, реализованный в работе Kapteyn и соавторов, основан на комбинации физических моделей и вероятностных графовых структур: формируется интерпретатор (например, оптимальное дерево решений), а цифровой двойник обновляется по результатам синтетических сценариев, сгенерированных физической моделью. Перспективные методы, предложенные Tripura и коллегами, дополняют эту схему использованием библиотек физически мотивированных функций и разрежённого байесовского обучения для автоматического выявления новых возмущающих членов и их параметров в динамике системы. Совместное применение этих подходов позволяет удерживать цифровой двойник в актуальном состоянии даже при неполноте исходной информации и высокой неопределённости климатических процессов.

Архитектура программного решения. Для реализации описанных принципов и процессов необходима соответствующая программная архитектура, способная обеспечить обработку больших потоков данных, выполнение ресурсоемких вычислений и предоставление результатов пользователям в удобной форме. Визуальное представление архитектуры программного решения цифрового двойника климата представлена на рисунке 4.

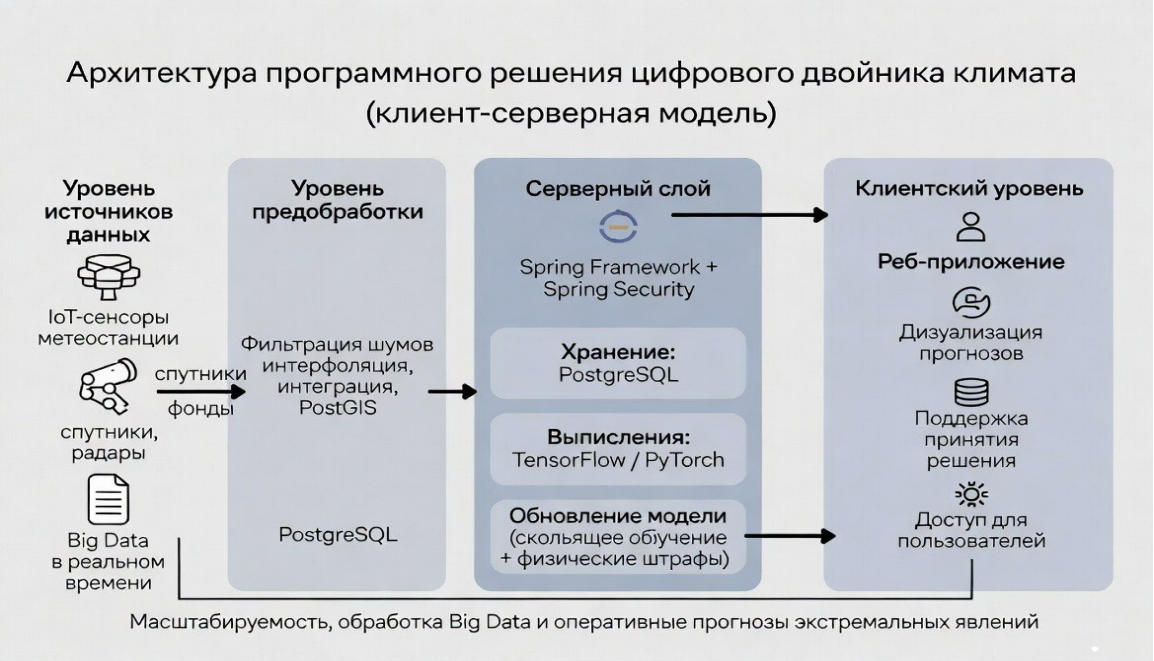


Рисунок 4. Архитектура программного решения цифрового двойника климата

В качестве базового подхода целесообразно рассмотреть клиент-серверную архитектуру. Особенности такой модели заключаются в том, что пользователь отправляет запрос на сервер, где он обрабатывается, после чего результат возвращается клиенту. Сервер при этом способен одновременно обслуживать множество пользователей, что критически важно для систем мониторинга и прогнозирования, где к данным могут обращаться как исследователи, так и лица, принимающие решения. Для реализации пользовательского доступа оптимальным представляется выбор веб-приложения. Главное преимущество такого решения – масштабируемость и доступность: возможность одновременной работы большого числа сотрудников различных ведомств и организаций. Веб-приложение не требует установки на локальные устройства, что снижает требования к вычислительным ресурсам пользователей и обеспечивает доступ с любой точки при наличии интернет-соединения.

В качестве программной платформы для серверной части предлагается рассмотреть экосистему Spring. Spring Security может быть использован для построения систем аутентификации и авторизации, разграничения прав доступа к данным и моделям различного уровня чувствительности. Для хранения и обработки данных целесообразно выбрать PostgreSQL, занимающий лидирующие позиции среди открытых реляционных СУБД и обладающий развитыми средствами работы с пространственными данными (расширение PostGIS), что особенно важно для геопривязанной климатической информации. Для реализации вычислительно интенсивных компонентов, связанных с машинным обучением и обработкой больших данных, архитектура предполагает интеграцию с фреймворками глубокого обучения (TensorFlow, PyTorch) и инструментами для работы с большими распределенными данными.

Практическая реализация описанных подходов может опираться на существующие открытые разработки. В качестве примера можно привести репозиторий polytope-examples в рамках инициативы Destination Earth, где представлены примеры реализации структур данных для цифровых двойников климатической направленности [11]. Эти примеры демонстрируют подходы к организации эффективного доступа к большим массивам геопространственных данных, их индексации и выборке для последующего использования в моделях. Предлагаемая архитектура обеспечивает необходимую масштабируемость,

гибкость и производительность для реализации всех этапов работы цифрового двойника – от приема данных с сенсоров до визуализации прогнозов и поддержки принятия решений.

Заключение. В работе представлена концепция программной платформы для создания самообучающихся цифровых двойников климатических и экстремальных погодных явлений, опирающаяся на развитие общего подхода к цифровым двойникам [1-5] и современные климатические инициативы CMIP, HighResMIP и Destination Earth [6-8]. Предложено перейти от статичных климатических моделей, основанных исключительно на решении систем дифференциальных уравнений, к динамическим системам, способным адаптироваться к текущим условиям за счёт ассимиляции данных в реальном времени и интеграции методов машинного обучения [6-9].

Сформулированы три ключевых принципа построения самообучающегося цифрового двойника: непрерывная ассимиляция разнородных наблюдений, редукция вычислительной сложности посредством surrogate-моделей, обученных на прогонных данных полнофизических моделей, и обеспечение физической согласованности прогнозов через включение физических ограничений в процедуру обучения [6-10]. Эти принципы согласуются с современными представлениями о предиктивных цифровых двойниках и позволяют сбалансировать точность описания процессов, вычислительную эффективность и оперативность обновления прогнозов. Описана программная архитектура платформы цифрового двойника климата, основанная на клиент–серверной модели, веб-интерфейсе, использовании реляционной СУБД с поддержкой пространственных данных и интеграции с фреймворками глубинного обучения, а также возможности опоры на открытые разработки, такие как polytope-examples в рамках инициативы Destination Earth. Такая архитектура обеспечивает масштабируемую обработку потоков климатических данных и доступ к результатам моделирования для широкого круга пользователей, включая исследователей и лиц, принимающих решения. Практическая значимость предложенного подхода заключается в потенциальном улучшении краткосрочного и среднесрочного прогнозирования экстремальных погодных явлений и создании инструментов поддержки решений в условиях усиливающейся климатической изменчивости [6-8].

Самообучающиеся цифровые двойники позволяют формировать специализированные численные сценарии развития опасных метеорологических и гидрологических событий на горизонтах от нескольких часов до нескольких дней, что создаёт основу для своевременного планирования мер адаптации и снижения рисков для населения и экономики [6-10].

Список литературы

[1] Grieves M. Virtually Intelligent Product Systems: Digital and Physical Twins. Presented at: Product Lifecycle Management (PLM) Course; 2002; University of Michigan, Detroit, USA.

[2] Kritzinger W., Karner M., Traar G., Henjes J., Sihn W. Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine. 2018;51(11):1016-1022. DOI: 10.1016/j.ifacol.2018.08.474.

[3] Zhuang C., Liu J., Xiong H. Connotation, architecture and trends of product digital twin. Computer Integrated Manufacturing Systems. 2017;23(4):753-768.

[4] Barricelli B.R., Casiraghi E., Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access. 2019;7:167653-167671. DOI: 10.1109/ACCESS.2019.2953499.

[5] Bergs T., Gierlings S., Brockmann T. The concept of digital twin and digital shadow in manufacturing. Procedia CIRP. 2021;101:81-84. DOI: 10.1016/j.procir.2021.02.010.

[6] CMIP – Coupled Model Intercomparison Project. Available at: https://wcrp-cmip.org/cmip-overview/ (accessed 05.03.2026).

[7] HighResMIP – High Resolution Model Intercomparison Project. Available at: https://www.highresmip.org/ (accessed 05.03.2026).

[8] European Centre for Medium-Range Weather Forecasts. Destination Earth: Climate Change Adaptation Digital Twin. Available at: https://destine.ecmwf.int/climate-change-adaptation-digital-twin-climate-dt/ (accessed 15.02.2026).

[9] Kapteyn M.G., Pretorius J.V.R., Willcox K.E. A probabilistic graphical model foundation for enabling predictive digital twins at scale. Nature Computational Science. 2021;1(5):337-347. DOI: 10.1038/s43588-021-00069-0.

[10] Tripura T., Chakraborty S. Sparse Bayesian learning for identifying digital twins of dynamical systems with symbolic recovery of governing physics. Mechanical Systems and Signal Processing. 2023;185:109776. DOI: 10.1016/j.ymssp.2023.110858.

[11] Destination Earth. Polytope Examples. Available at: https://github.com/destination-earth-digital-twins/polytope-examples (accessed 26.02.2026).

CONCEPT OF DIGITAL TWINS AND THEIR APPLICATION TO CLIMATE MODELING

V.O. Tikhonenko
Third-year student majoring in Software Engineering at the Faculty of Computer Aided Design

E.D. Rakitsky
Third-year student majoring in Software Engineering, Faculty of Computer Aided Design

Yu. A. Yeschik
Third-year student majoring in Software Engineering, Faculty of Computer Design

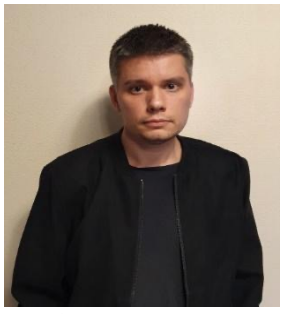
A.V. Vorobey
Assistant of the Department of Engineering Psychology and Ergonomics at BSUIR, Master of Technical Sciences

Abstract. The article presents a concept of a software platform for building self-learning digital twins of climate phenomena. It substantiates the need to overcome the limitations of traditional static models that cannot adapt to current conditions. Three key design principles are formulated: real-time data assimilation, reduction of computational complexity through surrogate models, and enforcement of physical consistency in forecasts. The architecture of the software solution based on a client–server model and methods for updating digital twins are described. The results are aimed at improving the accuracy of forecasts of extreme weather events and supporting decision-making.

Keywords: digital twin, climate modeling, machine learning, data assimilation, Big Data, surrogate model, physical consistency, weather forecasting.

УДК 004:021.004

ЭЛЕМЕНТЫ ОБРАБОТКИ ДАННЫХ В СИСТЕМАХ УПРАВЛЕНИЯ БЕСПИЛОТНЫМ АВТОМОБИЛЕМ



А.И. Селезнев
Ассистент кафедры электронных вычислительных машин
БГУИР, магистр
a.seleznev@bsuir.by

А.И. Селезнев

Окончил Белорусский государственный университет информатики и радиоэлектроники и магистратуру. Область научных интересов связана с искусственным интеллектом и нейросетями в технических системах.

Аннотация. В статье рассматриваются ключевые элементы обработки данных в системах управления беспилотным транспортом. Анализируются технологические решения, обеспечивающие безопасное и эффективное функционирование автономных транспортных средств. Рассматриваются средства аппаратной поддержки в управлении БА.

Ключевые слова: беспилотный автомобиль (БА), обработка данных в БА, сенсорные системы, аппаратная поддержка.

Введение. Развитие технологий беспилотного транспорта требует создания надежных систем обработки данных, способных в реальном времени анализировать информацию с различных сенсоров и принимать корректные решения.

Ключевой компонент работы беспилотных автомобилей (БА) – система обработки данных, которая преобразует информацию от датчиков в команды для движения [1].

Основные элементы обработки данных в БА [2]:

1. Сбор данных: датчики и сенсоры. Камеры (монокулярные, стерео, кругового обзора) фиксируют визуальную информацию: дорожную разметку, знаки, светофоры, других участников движения.

Разрешение – от HD до 4K, частота кадров – 30–60 Гц. Лидары (LIDAR) создают 3D-карту окружения с помощью лазерных импульсов.

Точность – до нескольких сантиметров, дальность – до 200 м.

Результат – «облако точек» с координатами (x,y,z) для каждого объекта.

Радары (RADAR) измеряют скорость и расстояние до объектов с помощью радиоволн. Работают в любых погодных условиях, дальность – до 300 м, точность по скорости – до 0,1 м/с. Ультразвуковые датчики используются для парковки и обнаружения препятствий на малых расстояниях (до 5–8 м).

Инерциальные измерительные устройства (IMU) отслеживают ускорение и угловую скорость автомобиля (акселерометры и гироскопы). GPS/ГЛОНАСС предоставляют глобальные координаты с точностью 2–5 м (обычные) или до 10 см (с RTK-коррекцией). Датчики одометрии (колесные энкодеры) фиксируют пройденное расстояние и скорость вращения колес.

2. Предварительная обработка данных. На этом этапе «сырые» данные очищаются и приводятся к единому формату. Калибровка датчиков применяется для выравнивания координат камер, лидаров и радаров в единой системе отсчета, компенсация искажений (например, дисторсии камер).

Фильтрация шумов производится с помощью применения фильтров (например, фильтра Калмана) для сглаживания данных с IMU и GPS, удаление ложных отражений в данных лидара. Синхронизация по времени нужна для объединения данных с разных датчиков с привязкой к единой временной метке (timestamp).

3. Восприятие окружения (Perception). Система анализирует окружение и идентифицирует объекты. Алгоритмы (YOLO, Faster R-CNN) на основе камер определяют классы объектов: автомобили, пешеходы, велосипеды. Лидарные данные обрабатываются с помощью PointNet или VoxelNet для 3D-обнаружения.

Сегментация (Semantic Segmentation) применяется для пиксельной классификация изображений (дорога, тротуар, здания). Отслеживание объектов (Tracking) – присвоение уникальных ID объектам и прогнозирование их траекторий (с помощью фильтров Калмана или SORT-алгоритмов). Локализация (Localization) нужна для сопоставления данных лидара/камер с высокоточной картой (HD-map) для определения позиции автомобиля с точностью до 10 см.

4. Планирование маршрута и принятие решений. На основе данных восприятия система строит маршрут и траекторию. Глобальное планирование (Global Path Planning) неообходимо для поиска оптимального маршрута от точки А до точки Б на HD-карте (алгоритмы A*, Dijkstra). Локальное планирование (Local Path Planning) служит для расчета

безопасной траектории движения с учетом динамических препятствий (алгоритмы RRT*, Model Predictive Control – MPC).

Принятие решений (Decision Making) включает в себя выбор необходимых действий: ускорение, торможение, перестроение, остановка.

Кроме того, необходимо учитывать ПДД и этические сценарии (например, при неизбежном столкновении).

5. Управление движением (Control) включает в себя реализацию запланированных действий с помощью регуляторов (PID, LQR), которые преобразуют желаемую траекторию в команды для актуаторов: угол поворота руля, газ, тормоз.

Обратная связь – данные с датчиков одометрии и IMU – используются для коррекции управляющих сигналов.

6. Вычислительная платформа включает в себя аппаратное обеспечение для обработки данных:

– центральные процессоры (CPU) выполняют общие задачи (логика, коммуникация);

– графические процессоры (GPU) ускоряют работу нейросетей для восприятия (например, NVIDIA DRIVE);

– специализированные чипы (ASIC, FPGA) оптимизированы под задачи ИИ (например, Google TPU, Tesla FSD Chip);

– системы охлаждения предотвращают перегрев при высоких нагрузках.

7. Коммуникация и безопасность реализуется с помощью V2X (Vehicle-to-Everything). Обмен данными с другими автомобилями (V2V) и инфраструктурой (V2I) происходит через Wi-Fi (DSRC) или 5G. Это особенно важно для получения информации о пробках, авариях, сигналах светофора. Кибербезопасность реализуется посредством шифрования данных и защиты от взлома бортовых систем.

8. Использование высокоточных карт (HD-Maps), которые содержат геометрия дорог, разметка, знаки, светофоры; 3D-модели окружения (деревья, здания).

Сценарий использования системы. На этапе сбора данных камера видит красный сигнал светофора, лидар обнаруживает пешехода на переходе, а радар фиксирует автомобиль впереди, замедляющийся со скоростью − 2 м/с².

На этапах предварительной обработки и восприятия окружения нейросеть классифицирует светофор как «красный», алгоритм сегментации выделяет пешехода в зоне перехода и фильтр Калмана прогнозирует траекторию переднего автомобиля.

Затем согласно этапу планирования система решает остановиться перед светофором, рассчитывается траектория торможения с ускорением −1,5 м/с².

На этапе управления PID-регулятор передает команду на тормоз и актуатор замедляет автомобиль.

Аппаратная поддержка: компьютеры и нейроморфные чипы [3].

Сложность алгоритмов ИИ и требования к времени реакции делают невозможной полную зависимость от облачных вычислений. Аппаратное обеспечение нового поколения является физическим фундаментом для работы всей системы.

Бортовые эдж-компьютеры. Эти устройства обеспечивают обработку данных непосредственно на борту автомобиля (на границе сети − «edge»).

Это позволяет значительно снизить задержки (латентность) при принятии решений, обеспечивая время реакции в миллисекунды, необходимое для безопасного управления.

Эдж-компьютеры выполняют первичную обработку сенсорных данных, запускают компактные версии нейросетей для критичных по времени задач и обеспечивают работу системы даже при потере связи с облаком.

Нейроморфные чипы. Это специализированные процессоры, архитектура которых имитирует структуру и принципы работы человеческого мозга, используя искусственные нейроны и синапсы.

Их ключевые преимущества для беспилотников заключаются в следующем:

− высокая энергоэффективность. Мозг человека потребляет около 20 Вт, и инженеры стремятся к аналогичной эффективности в нейроморфных системах. Это критически важно для электромобилей, где каждый ватт энергии на счету;

− высокая скорость обработки неструктурированных данных. Нейроморфные чипы хорошо справляются с задачами распознавания образов и обработки потоковых данных в реальном времени, что является основой восприятия в автономном вождении;

− возможность обучения на устройстве: в отличие от традиционных архитектур, нейроморфные системы потенциально способны к непрерывному обучению и адаптации непосредственно в процессе работы, без необходимости мощных облачных кластеров для переобучения моделей.

Заключение. Эффективная обработка данных является ключевым фактором успешного функционирования беспилотных транспортных средств.

Постоянное совершенствование алгоритмов и методов обработки позволяет повышать уровень автономности и безопасности систем управления.

Рассмотрены средства аппаратной поддержки.

Перспективами развития и дальнейших исследований являются: синергия технологий нейронных сетей, интернет вещей, блокчейн в управлении БА, их аппаратная поддержка; повышение точности датчиков, улучшение алгоритмов машинного обучения, внедрение 5G и V2X для обмена данными в реальном времени, создание единых стандартов безопасности и регулирования.

Список литературы

[1] Autonomous Vehicles Enabled by the Integration of IoT, Edge Intelligence, 5G, and Blockchain. *Sensors*. 2023; 23(4):1963.

[2] Петров В.В. Обработка данных в автономных системах. – СПб.: Техника, 2024.

[3] Нейроморфные системы: компьютеры, вдохновленные устройством человеческого мозга. *Habr*. 2020. URL: https://habr.com/ru/companies/mws/articles/457578/. – Дата доступа : 07.03.2026.

COMPONENTS OF DATA PROCESSING IN AUTONOMOUS VEHICLE CONTROL SYSTEMS

A.I. Seleznev

Teaching Assistant at the Department of Electronic Computing Machines, Belarusian State University of Informatics and Radioelectronics (BSUIR), Master's degree

Abstract. The article examines the key elements of data processing in autonomous vehicle control systems. It analyzes technological solutions, algorithms, and methods that ensure the safe and efficient operation of autonomous vehicles. Hardware support tools in UAV management are being considered.

Keywords: unmanned vehicle (UAV), data processing in the UAV, sensor systems, hardware support.

УДК 004.6:005.953.2

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ BIG DATA В СОВРЕМЕННОМ РЕКРУТИНГЕ



А.А. Гайсенок
Заведующий сектором сопровождения ERP БГУИР, магистрант кафедры экономической информатики БГУИР
a.gajsenok@bsuir.by



Г.А. Хацкевич
Профессор кафедры экономической информатики БГУИР, доктор экономических наук, кандидат технических наук, профессор
g.a.khatskevich@gmail.com

А.А. Гайсенок

Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов: интеллектуальные системы поддержки принятия решений, анализ данных в управлении персоналом, цифровые инструменты оценки персонала.

Г.А. Хацкевич

Окончил математический факультет Белорусского государственного университета. Область научных интересов связана с эконометрическим моделированием и прогнозированием социально-экономических процессов, а также развитием инновационного высшего образования.

Аннотация. В статье рассматриваются возможности применения технологий Big Data в процессе подбора персонала. Описаны основные этапы рекрутинга, которые могут быть автоматизированы с помощью анализа данных. Представлены примеры использования интеллектуальных систем для поиска, оценки и отбора кандидатов. Особое внимание уделено преимуществам внедрения таких технологий для компаний, а также возможным рискам и ограничениям.

Ключевые слова: Big Data, рекрутинг, подбор персонала, автоматизация HR, интеллектуальные системы, анализ данных.

Введение. Каждый день компании по всему миру размещают тысячи вакансий, а соискатели отправляют миллионы резюме. HR-специалисты тратят огромное количество времени на просмотр анкет, проведение собеседований и сравнение кандидатов. При этом найти действительно подходящего сотрудника становится все сложнее: количество информации растет, а время на принятие решений сокращается.

Здесь на помощь приходят технологии Big Data. Они позволяют автоматизировать рутинные задачи, находить скрытые закономерности и принимать более обоснованные решения. В этой статье мы рассмотрим, как именно анализ больших данных помогает в рекрутинге, какие инструменты для этого используются и какие выгоды получают компании.

Big Data в рекрутинге. Под Big Data в рекрутинге понимают сбор и анализ больших объемов информации о кандидатах, вакансиях и процессах подбора персонала. Источниками такой информации могут быть:

- резюме и сопроводительные письма;
- профили кандидатов в социальных сетях (LinkedIn, Habr, GitHub);
- результаты тестов и собеседований;

- данные о текущих сотрудниках компании;
- информация о рынке труда и уровне заработных плат [1].

Раньше эти данные существовали отдельно друг от друга, и HR-специалист мог анализировать их только вручную. Сегодня специальные системы собирают всю информацию в единое целое, обрабатывают ее и выдают готовые рекомендации.

Big Data на разных этапах рекрутинга. Процесс подбора персонала можно разделить на несколько этапов. На каждом из них технологии анализа данных приносят свою пользу.

Поиск кандидатов. Традиционно компании размещают вакансии на специальных сайтах и ждут откликов. Но часто лучшие специалисты не ищут работу активно – они уже где-то работают. Big Data позволяет находить таких «пассивных» кандидатов, анализируя их профессиональную активность в интернете: участие в конференциях, публикации статей, профили на GitHub. Система сама подбирает людей, чьи навыки и опыт соответствуют требованиям вакансии, и предлагает их рекрутеру [2].

Скрининг резюме. На одну вакансию может прийти сотни откликов. Просматривать каждое резюме вручную – долго и утомительно. Интеллектуальные системы автоматически читают тексты резюме, выделяют ключевые навыки, опыт работы, образование и сравнивают их с требованиями вакансии. В результате рекрутер получает список наиболее подходящих кандидатов, отсортированных по степени соответствия. В работе [3] была предложена программная система, включающая модуль расчета рейтинга соискателей, который позволяет автоматизировать процесс скрининга и повысить объективность оценки кандидатов.

Оценка кандидатов. После первичного отбора начинаются собеседования. Здесь Big Data тоже может помочь. Например, существуют платформы, которые анализируют видеозаписи интервью: оценивают речь, эмоции, манеру отвечать на вопросы. Другие системы предлагают кандидатам пройти онлайн-игры, в процессе которых оцениваются их коммуникативные способности, стрессоустойчивость и стиль мышления.

Прогнозирование успешности. Самое ценное, что дает анализ данных, – это возможность предсказать, насколько успешным будет кандидат в новой компании. Система сравнивает характеристики соискателя с данными о лучших сотрудниках компании и выдает прогноз: с какой вероятностью этот человек пройдет испытательный срок, как долго проработает и какой вклад принесет в результат.

Удержание сотрудников. Big Data работает не только на этапе найма, но и после него. Анализируя данные о текущих сотрудниках, можно выявить тех, кто может уволиться в ближайшее время, и вовремя принять меры, чтобы их удержать.

Какие технологии используются. За умными возможностями, описанными выше, стоят различные технологии анализа данных. Вот основные из них:

Обработка текста (NLP). Эта технология позволяет компьютеру «понимать» человеческий язык. Система может прочитать резюме, выделить из него важную информацию и даже определить, насколько опыт кандидата соответствует описанному в вакансии.

Машинное обучение. Алгоритмы учатся на исторических данных: какие кандидаты были приняты на работу, как они потом себя проявили. На основе этого обучения система начинает сама определять, какие характеристики делают кандидата потенциально успешным.

Анализ социальных сетей. Технология изучает связи между людьми в профессиональных сообществах. Это помогает не только находить новых кандидатов, но и получать рекомендации от текущих сотрудников, что часто оказывается самым надежным источником найма.

Преимущества внедрения. Компании, которые используют Big Data в рекрутинге, получают ряд важных преимуществ:

- Экономия времени. Автоматизация рутинных задач позволяет рекрутерам сосредоточиться на самом главном – общении с кандидатами и принятии финальных решений. Время закрытия одной вакансии может сократиться на 30–50 % [4].
- Экономия денег. Чем быстрее найдется подходящий сотрудник, тем меньше компания тратит на закрытие вакансии. Кроме того, точный отбор снижает количество ошибок найма, когда человек увольняется уже через несколько месяцев.
- Объективность. Человек может устать, отвлечься или неосознанно симпатизировать одним кандидатам больше, чем другим. Алгоритм лишен этих недостатков – он оценивает только факты и данные.
- Качество найма. Благодаря прогнозным моделям компании чаще находят тех, кто действительно будет успешен на новом месте. Это повышает эффективность работы команд и снижает текучесть кадров [5].

Риски и ограничения. Однако, как и у любой технологии, у Big Data в рекрутинге есть свои недостатки, о которых важно знать.

- Алгоритмическая предвзятость. Если в исторических данных, на которых обучалась система, были скрытые предубеждения (например, компания чаще нанимала мужчин на технические должности), алгоритм может воспроизвести эту же несправедливость. Поэтому важно регулярно проверять, не дискриминирует ли система кандидатов по полу, возрасту или другим признакам.
- Нехватка человеческого взгляда. Не все качества кандидата можно измерить цифрами. Интуиция, личная химия, культурное соответствие – это то, что иногда лучше чувствует человек, а не алгоритм.
- Защита данных. Сбор и обработка персональной информации о кандидатах требует соблюдения законодательства о защите данных. Компании должны быть прозрачны в том, как они используют информацию о соискателях.

Заключение. Технологии Big Data уже сегодня меняют рекрутинг, делая его быстрее, эффективнее и объективнее. Автоматический поиск кандидатов, интеллектуальный скрининг резюме, прогнозирование успешности – все это становится доступным для компаний любого размера.

Однако важно помнить, что технологии – это лишь инструмент. Они не заменяют человека-рекрутера, а помогают ему работать лучше. Окончательное решение о найме должно оставаться за человеком, который учитывает не только цифры, но и нюансы, которые не может уловить алгоритм.

В ближайшие годы можно ожидать дальнейшего развития этой области: появления более умных алгоритмов, интеграции с новыми источниками данных и создания комплексных платформ, которые будут сопровождать сотрудника на всем пути – от первого контакта до увольнения.

Список литературы

[1] LinkedIn. Global Talent Trends 2025 [Электронный ресурс]. – 2025. – Режим доступа: https://business.linkedin.com/talent-solutions/global-talent-trends. – Дата доступа: 10.03.2026.

[2] Boudreau, J. W. Retooling HR: Using Proven Business Tools to Make Better Decisions About Talent / J. W. Boudreau. – Boston: Harvard Business Review Press, 2010. – 256 p.

[3] Михалевич (Гайсенок), А. А. Программная поддержка рекрутинговой деятельности с модулем расчета рейтинга соискателей и системы KPI / А. А. Михалевич (Гайсенок) // Подготовка научных кадров: опыт, проблемы, перспективы : материалы VI Международной научно-практической конференции, Минск, 6 декабря 2024 года / Университет НАН Беларуси ; редкол.: М. Г. Жилинский (отв. ред.) [и др.]. – Минск : ИВЦ Минфина, 2024. – С. 63–66.

[4] Deloitte. 2025 Global Human Capital Trends [Электронный ресурс]. – 2025. – Режим доступа: https://www2.deloitte.com. – Дата доступа: 10.03.2026.

[5] Чуланова, О. Л. Цифровые технологии в управлении персоналом: тренды, возможности и риски / О. Л. Чуланова // Управление персоналом и интеллектуальными ресурсами в России. – 2023. – № 2. – С. 32–38.

Авторский вклад

Гайсенок Анна Александровна – постановка цели и задач исследования, анализ преимуществ и экономической эффективности внедрения Big Data в рекрутинг, описание технологических аспектов применения анализа данных в рекрутинге, анализ этапов подбора персонала, рассмотрение рисков и ограничений, формулировка выводов.

Хацкевич Геннадий Алексеевич – научное консультирование, корректировка методологии, доработка итогового текста.

APPLICATION OF BIG DATA TECHNOLOGIES IN MODERN RECRUITMENT

H.A. Haisenok
Head of ERP Support Section, BSUIR; Master's Student, Department of Economic Informatics, BSUIR

G.A. Khatskevich
Professor, Department of Economic Informatics, BSUIR, Doctor of Economic Sciences, Professor

Abstract. The article discusses the possibilities of using Big Data technologies in the process of personnel selection. The main stages of recruitment that can be automated using data analysis are described. Examples of using intelligent systems for searching, evaluating and selecting candidates are presented. Special attention is paid to the advantages of implementing such technologies for companies, as well as possible risks and limitations.

Keywords: Big Data, recruitment, personnel selection, HR automation, intelligent systems, data analysis.

УДК 004.9

ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ЧАТ-БОТА ДЛЯ ГЕНЕРАЦИИ РЕКОМЕНДАЦИЙ ПО ИНКЛЮЗИВНОМУ ОБРАЗОВАНИЮ С ПРИМЕНЕНИЕМ RAG-ТЕХНОЛОГИЙ





М.М. Тарасов
Магистрант Высшей школы электроники и компьютерных наук
ФГАОУ ВО «ЮУрГУ (НИУ)»
tosik4302@mail.ru

Л.Ю. Костылева
Доцент кафедры «Информационные системы и технологии» Высшей школы электроники и компьютерных наук ФГАОУ ВО «ЮУрГУ (НИУ)», кандидат технических наук
kostylevali@susu.ru

М.М. Тарасов

Окончил Южно-Уральский государственный университет (национальный исследовательский университет). Область научных интересов связана с разработкой информационных систем, методов и алгоритмов машинного обучения.

Л.Ю. Костылева

Окончила Южно-Уральский государственный университет (национальный исследовательский университет). Область научных интересов связана с исследованием проблем управления в организационных системах, методов и алгоритмов совершенствования управленческой деятельности.

Аннотация. В статье рассматривается концепция и архитектура разрабатываемого чат-бота, предназначенного для помощи преподавателям высших учебных заведений в организации образовательного процесса для студентов с ограниченными возможностями здоровья (ОВЗ). Предлагаемое решение базируется на современной RAG-технологии (Retrieval-Augmented Generation), которая позволяет интегрировать большие языковые модели с актуальной нормативно-методической базой. В качестве источника знаний выступают официальные методические пособия Министерства науки и высшего образования РФ. Описываются функциональные модули системы, обосновывается выбор технологического стека и приводятся ожидаемые результаты внедрения.

Ключевые слова: RAG-технология, чат-бот, инклюзивное образование, студенты с ОВЗ, искусственный интеллект в образовании, рекомендательная система.

Введение. Современная система высшего образования в Российской Федерации характеризуется устойчивым развитием инклюзивных практик и расширением доступности образовательных услуг для различных категорий граждан. В последние годы наблюдается рост числа студентов с ограниченными возможностями здоровья (ОВЗ), поступающих в вузы. Это связано как с совершенствованием нормативно-правовой базы, так и с формированием более доступной образовательной среды [1].

Ключевым органом, отвечающим за нормативное регулирование в этой области, выступает Министерство науки и высшего образования РФ. Именно оно разрабатывает подзаконные акты и инструктивные письма, регламентирующие процесс обучения студентов с инвалидностью и ограниченными возможностями здоровья. Базовые гарантии права на образование, а также описание механизмов создания адаптированных условий закреплены в Федеральном законе «Об образовании в Российской Федерации» [2].

Современное инклюзивное образование в вузах нельзя сводить только к установке пандусов или специального оборудования. Ключевым фактором становится готовность педагогов взаимодействовать с обучающимися, имеющими разные нозологии. Это требует от преподавателя понимания специфики усвоения материала, умения корректировать содержание курсов, подбирать адекватные формы контроля и выстраивать коммуникацию с учетом индивидуальных особенностей студента. Реализация таких задач невозможна без дополнительных психолого-педагогических знаний и способности быстро перестраивать методику работы.

На первый план выходят такие инструменты, как проектирование персональных траекторий обучения, внедрение дистанционных форматов, корректировка оценочных средств и использование вариативных способов подачи информации. Известны различные успешные способы применения современных цифровых технологий, в том числе чат-ботов и социальных роботов в инклюзивном образовании [3], направленные на персонализированную поддержку студентов с различными образовательными потребностями и ограниченными возможностями здоровья, которые охватывают широкий спектр нозологий, включая нарушения зрения, слуха и другие специфические трудности обучения [4].

Тем не менее, на деле внедрение современных технологий сталкивается с препятствиями в виде высокой загруженности преподавательского состава и запутанности нормативной документации. Действующие регламенты, касающиеся организации учебного процесса для студентов с ОВЗ, отличаются большим объемом и требуют детального изучения. Как следствие, оперативное нахождение ответов на возникающие в конкретной педагогической ситуации вопросы превращается в серьезную проблему [5].

Хотя официальные методические рекомендации существуют, использовать их в ежедневной рутине непросто. Перед преподавателем часто встают частные, но важные вопросы: как скорректировать формат текущего контроля, что должно быть отражено в рабочей программе, какие именно специальные условия обязательны к исполнению. Сбор информации из множества разрозненных источников отнимает много времени и чреват ошибками, связанными с неполнотой или неверной интерпретацией требований [6].

В эпоху цифровизации образования все более востребованными становятся интеллектуальные помощники и системы поддержки принятия решений. Использование чат-ботов для быстрого извлечения информации из загруженной документации представляет собой современный подход к автоматизации работы с документами, который сочетает технологии обработки естественного языка, машинного обучения и информационного поиска. Современные системы обработки естественного языка дают пользователям возможность формулировать запросы в обычной речи и получать точные ответы из текстовых документов без ручного анализа [7]. Алгоритмы на базе трансформеров очень эффективны в работе с текстом – модели BERT показывают высокую точность при классификации документов и при извлечении релевантной информации, уменьшают вероятность человеческих ошибок и заметно ускоряют обработку текстовых данных по сравнению с традиционным анализом [8]. Развитие больших языковых моделей позволяет автоматизировать обработку текстов и генерировать ответы в форме, удобной для человека. Но если применять такие модели в текущем виде без проверки информации по надёжным источникам, это может привести к неправильным или искажённым сведениям.

Эффективной альтернативой является технология дополненной генерации ответов – она объединяет возможности языковой модели с механизмом поиска по авторитетной базе знаний. Благодаря этому итоговый ответ строится на основе фрагментов актуальных официальных документов, что гарантирует его точность и достоверность. Для образовательной практики, где строгое соблюдение нормативов принципиально важно, такой подход представляется наиболее перспективным [9].

В настоящей работе предлагается концепция чат-бота, выполняющего функции интеллектуального ассистента для преподавателей вузов. Бот реализует RAG-архитектуру и обеспечивает диалоговый формат взаимодействия. Пользователь формулирует запрос в свободной форме, после чего система осуществляет поиск релевантных положений в базе нормативных и методических документов и генерирует структурированные рекомендации.

Предлагаемое решение направлено на повышение оперативности принятия педагогических решений, снижение временных затрат на поиск информации и повышение нормативной корректности организации обучения студентов с ОВЗ. Новизна подхода заключается в интеграции RAG-технологии с официальными методическими материалами и ориентации на практическую поддержку преподавателей в режиме реального времени. Разработка подобной системы соответствует современным тенденциям цифровизации высшего образования и отвечает актуальным потребностям инклюзивной образовательной среды.

Постановка задачи и формирование требований к системе. Разработка и создание прототипа чат-бота, способного на основе запроса преподавателя формировать релевантные, нормативно обоснованные рекомендации по адаптации образовательного процесса для студентов с различными нозологиями ОВЗ, обусловлено необходимостью повышения качества инклюзивного образования и снижения временных затрат преподавателей на поиск и интерпретацию регламентирующих документов. Предполагается, что система будет функционировать как интеллектуальный ассистент, обеспечивающий оперативный доступ к структурированной базе знаний, сформированной на основе официальных методических материалов и нормативных документов. Разрабатываемая система не заменяет эксперта в области инклюзии, но может выступать в роли инструмента поддержки принятия решений, повышая обоснованность, корректность и уверенность педагогов в своих действиях.

Для достижения поставленной цели необходимо решить ряд взаимосвязанных задач.

1. Анализ и систематизация источников базы знаний. Нужно провести отбор, структурирование и тематическую классификацию официальных нормативных документов и методических пособий, разработанных Министерством науки и высшего образования

Российской Федерации, а также иных регламентирующих материалов, касающихся организации обучения студентов с ОВЗ. Собранные документы необходимо обработать, выделяя фрагменты, в которых содержатся практико-ориентированные рекомендации. Имеющиеся в вузе локальные нормативные документы и актуальные наработки в области реализации инклюзивного образования также могут быть включены в базу знаний. В ходе систематизации источников следует также запланировать их будущие обновления – продумать способ оценки текущей версии, перечень критериев и атрибутов, позволяющих квалифицировать источник как актуальный либо устаревший, порядок поиска обновлённых версий уже задействованных материалов и правила добавления новых источников в рабочий массив данных.

2. Проектирование архитектуры RAG-системы. Требуется разработать модель архитектуры, включающую модуль предварительной обработки и векторизации текстовых документов с учетом формата файлов, векторную базу данных для хранения эмбеддингов и механизм поиска релевантных фрагментов. Архитектура должна обеспечивать масштабируемость, возможность обновления базы знаний и корректную обработку запросов различной сложности.

3. Интеграция генеративной модели. Интеграция должна предусматривать формирование промптов с учетом найденных фрагментов документов, а также настройку параметров генерации для обеспечения точности, формальной корректности, минимизации придуманных ответов, так называемых «галлюцинаций», и стилистической нейтральности ответов. В качестве генеративного ядра системы планируется использование свободно распространяемых генеративных трансформеров (GPT-моделей), подобранных с учетом имеющихся вычислительных возможностей.

4. Разработка пользовательского интерфейса. Взаимодействие пользователя с системой предполагается реализовать через веб-интерфейс или доступный мессенджер. Интерфейс должен обеспечивать удобный ввод запросов на естественном языке, получение структурированных ответов с отображением ссылок на источники, поддержку диалогового формата общения, возможность копирования текста вопросов и ответов, ведение структурированного журнала запросов и ответов с целью последующего анализа диалоговых паттернов.

5. Апробация и оценка эффективности. Планируется проведение пилотного внедрения прототипа в образовательной среде Южно-Уральского государственного университета. В рамках апробации предполагается анализ точности ответов, релевантности рекомендаций, удовлетворенности пользователей и оценка влияния системы на скорость принятия педагогических решений.

Ключевые функциональные требования к системе включают:

– приём и корректную интерпретацию запросов, сформулированных на естественном языке (например, «Какие требования к освещению в аудитории для слабовидящего студента?», «Как адаптировать лабораторную работу для студента с нарушением опорно-двигательного аппарата?»);

– автоматизированный поиск релевантных данных в базе знаний с применением алгоритмов векторного сходства для сопоставления семантики запроса и содержимого базы;

– генерацию связного и структурированного ответа, обладающего практической ценностью, с обязательным указанием источника информации или нормативного основания из документов, находящихся в базе знаний;

– поддержку истории диалога и учёт контекста предыдущих взаимодействий для повышения точности и релевантности рекомендаций;

– обеспечение защиты персональных данных пользователей и устойчивости функционирования системы в условиях повышенной нагрузки.

Дополнительно система должна отвечать требованиям надежности, расширяемости и возможности последующего масштабирования. В перспективе архитектура должна позволять обновлять базу знаний при появлении новых нормативных документов без существенной переработки программной части.

Предлагаемая архитектура решения. Предлагаемая архитектура построена на принципах Retrieval-Augmented Generation (RAG) [9] и включает несколько логически взаимосвязанных модулей, реализующих функциональные требования к системе и обеспечивающих полный цикл обработки запроса – от загрузки нормативных документов до формирования финального ответа преподавателю (рисунок 1).

1. Модуль загрузки и обработки данных. Источником знаний выступают официальные PDF-документы, размещенные на портале Министерства науки и высшего образования Российской Федерации и других официальных сетевых ресурсах – нормативно-правовые акты, методические рекомендации, инструкции, регламенты и иные материалы, касающиеся организации обучения студентов с ОВЗ.

На этапе загрузки документов реализуется автоматизированный конвейер обработки данных, обеспечивающий извлечение текста из PDF-файлов, очистку от служебной информации (колонтитулы, номера страниц, повторяющиеся заголовки), нормализацию текста (удаление лишних пробелов, приведение кодировки), разбиение на логические фрагменты (чанкинг) и извлечение метаданных.

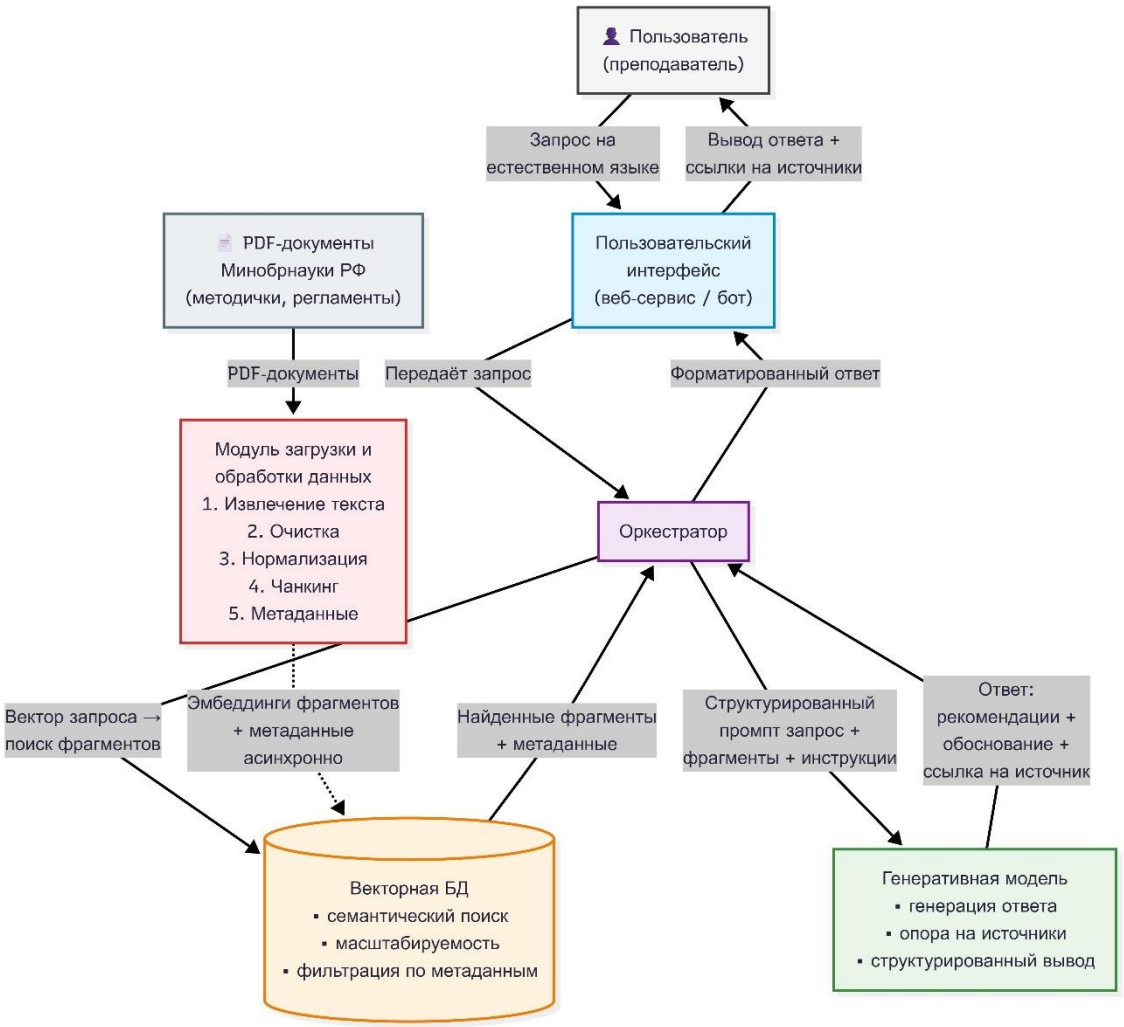


Рисунок 1. Концепция архитектуры чат-бота

Чанкинг осуществляется с учетом смысловой структуры документа, включающей сооветствующие разделы, подпункты, абзацы. Размер фрагмента подбирается с учетом контекстного окна применяемой GPT-модели таким образом, чтобы сохранялась семантическая целостность, но при этом обеспечивалась точность поиска. Каждый фрагмент снабжается метаданными (например, название документа, раздел, номер страницы, дата публикации), что в дальнейшем позволяет формировать корректные ссылки на источник.

2. Векторная база данных. После разбиения текстовые фрагменты преобразуются в векторные представления (эмбеддинги) с использованием специализированной модели. Эмбеддинг отражает семантическое содержание текста в виде числового вектора фиксированной размерности.

Полученные векторы сохраняются в векторной базе данных (например, ChromaDB, FAISS или Qdrant). Использование векторной БД обеспечивает быстрый семантический поиск по принципу близости, масштабируемость при увеличении объема документов, а также возможность фильтрации по метаданным (тип документа, год и т.д.).

Поиск релевантных фрагментов осуществляется по косинусной мере сходства между вектором пользовательского запроса и векторами, хранящимися в базе.

3. Оркестратор. Представляет собой центральный модуль системы, координирующий взаимодействие между компонентами. Получив запрос от преподавателя через интерфейс чат-бота, он выполняет следующие шаги:

1) преобразует текст запроса в вектор с использованием той же модели эмбеддингов;

2) выполняет поиск в векторной базе данных и извлекает наиболее релевантные фрагменты;

3) с помощью специальной модели (реранкера) структурирует извлеченные фрагменты по приоритету: самые релевантные фрагменты первыми подаются в генеративную модель, повышая точность и качество ответа;

4) формирует структурированный промпт для генеративной модели, включающий оригинальный запрос пользователя, извлеченные фрагменты нормативных документов и инструкции модели о необходимости опираться исключительно на предоставленные источники.

Таким образом обеспечивается контролируемая генерация ответа и минимизация риска появления недостоверной информации (галлюцинаций).

4. Генеративная модель. Модель получает подготовленный промпт и формирует связный, структурированный ответ, включающий конкретные рекомендации, нормативное обоснование и ссылку на источник (по метаданным фрагмента).

Ключевым принципом является ограничение генерации рамками предоставленных документов. Модель синтезирует ответ только на основе найденных фрагментов, что повышает надежность системы. В качестве генеративного ядра системы планируется использовать доступные GPT-модели, которые могут функционировать во внутреннем контуре организации и имеют удобный API.

5. Пользовательский интерфейс. Может быть реализован в нескольких вариантах – веб-сервиса или бота для доступного мессенджера. Интерфейс обеспечивает прием запросов на естественном языке, вывод структурированных ответов, отображение ссылок на источники и хранение истории диалога для учета контекста.

Описанная архитектура системы ориентирована на модульность, масштабируемость и возможность дальнейшего расширения функциональности.

Ожидаемые результаты и перспективы внедрения. На текущем этапе разработка системы находится в стадии формирования базового прототипа. Реализован чат-бот, обеспечивающий пользовательский интерфейс и обработку входящих сообщений, однако интеграция с полнофункциональным RAG-контуром и генеративной моделью еще не завершена. В настоящее время бот выполняет демонстрационную функцию и служит

платформой для тестирования логики взаимодействия с пользователем. Основные усилия следующего этапа разработки будут направлены на построение векторной базы знаний, настройку алгоритмов семантического поиска и интеграцию с генеративной моделью через API.

Ключевым шагом станет формирование структурированной базы знаний на основе официальных методических документов, их векторизация и размещение в специализированной базе данных. После завершения данного этапа система сможет обеспечивать релевантный поиск фрагментов и формировать ответы, опираясь на нормативные источники. Для оценки эффективности разработанного решения нужно выполнить анализ того, насколько точно извлекаются текстовые данные, грамотно ли выполнено деление документов на логические блоки и правильно ли сопоставляются запросы пользователей с релевантными фрагментами нормативных и методических источников.

Ожидается, что интеграция созданного чат-бота в деятельность Южно-Уральского государственного университета приведет к ряду ощутимых результатов – и в практической плоскости, и в организации учебного процесса. Педагоги будут тратить меньше времени на поиск нужной информации. Вместо того чтобы изучать многостраничные регламенты, преподаватель может просто задать вопрос и сразу получить понятный структурированный ответ с ссылкой на конкретный документ. Фокус сместится с бюрократических процедур на само обучение студентов, это ускорит подготовку к занятиям.

Кроме того, система призвана улучшить качество разработки адаптированных образовательных программ и планирования отдельных занятий с учетом потребностей студентов с ограниченными возможностями здоровья. Благодаря семантическому анализу и опоре на свежие нормативные данные, бот будет предлагать решения, которые учитывают характер заболевания, тип учебного занятия и выбранную форму проверки знаний. Это позволит перейти к более продуманной и методологически выверенной инклюзивной политике в университете.

В психологическом плане ожидается, что появление интеллектуального помощника снизит тревожность и неуверенность преподавателей, поскольку вопросы, касающиеся обучения студентов с ОВЗ, зачастую порождают чувство неопределенности и страх нарушить существующие предписания. Наличие инструмента, предоставляющего обоснованные рекомендации со ссылкой на официальные документы, призвано повысить уверенность педагогов в корректности принимаемых решений.

В перспективе бот может стать единым цифровым окном доступа к актуальной нормативной информации в сфере инклюзивного образования. Централизация источников в рамках одной системы упростит навигацию и обеспечит единообразие интерпретации требований. В дальнейшем это может способствовать формированию единой методической культуры работы со студентами с ОВЗ внутри университета.

Развитие проекта предполагает расширение базы знаний за счет включения локальных нормативных актов университета, регламентов кафедр, положений о текущем и промежуточном контроле, а также доступных лучших практик других образовательных организаций. Возможна интеграция аналитических модулей, позволяющих выявлять наиболее частые запросы преподавателей и формировать рекомендации по совершенствованию внутренних регламентов.

Таким образом, реализация и последующее масштабирование системы имеют потенциал стать значимым элементом цифровой трансформации инклюзивного образования, обеспечивая устойчивую поддержку преподавателей и повышение качества образовательного процесса в целом.

Заключение. В статье представлена концепция интеллектуального чат-бота, предназначенного для поддержки преподавателей высших учебных заведений в вопросах организации инклюзивного образовательного процесса. Предложенное решение

ориентировано на формирование удобного цифрового инструмента, обеспечивающего оперативный доступ к нормативной и методической информации, связанной с обучением студентов с ограниченными возможностями здоровья. В основе архитектуры системы лежит применение технологии RAG, сочетающей семантический поиск по базе знаний с возможностями генеративной модели. Такой подход позволяет формировать ответы, опирающиеся на официальные документы и методические рекомендации, что обеспечивает высокую релевантность, нормативную корректность и практическую применимость выдаваемых рекомендаций. Использование диалогового формата взаимодействия делает систему доступной и удобной в повседневной работе преподавателя. Предполагается, что дальнейшее развитие системы позволит не только повысить оперативность принятия педагогических решений, но и укрепить методическую культуру инклюзивного образования в университете.

Социальная значимость проекта заключается в его направленности на создание более доступной, комфортной и справедливой образовательной среды. Внедрение подобного инструмента может способствовать снижению барьеров в обучении студентов с ОВЗ и повышению качества их образовательного опыта, что соответствует современным приоритетам развития высшего образования.

Список литературы

[1] Инклюзивная образовательная среда вуза (обзор зарубежных исследований) / С. Т. Кохан, С. В. Власова, П. Ангелова, С. Павлович // Ученые Записки Забайкальского Государственного Университета. – 2024. – Т. 19. – № 2. – С. 41–50.

[2] Федеральный закон «Об образовании в Российской Федерации» от 29.12.2012 N 273-ФЗ (последняя редакция) \ КонсультантПлюс. – URL: https://www.consultant.ru/document/cons_doc_LAW_140174/ (дата обращения: 29.03.2026). – Текст : электронный.

[3] Social Robots in Special Education: A Systematic Review / G. A. Papakostas, G. K. Sidiropoulos, C. I. Papadopoulou [et al.] // Electronics. – 2021. – Vol. 10. – Social Robots in Special Education. – № 12. – P. 1398. – DOI: 10.3390/electronics10121398.

[4] Pérez, J. Q. Rediscovering the use of chatbots in education: A systematic literature review / J. Q. Pérez, T. Daradoumis, J. M. M. Puig // Computer Applications in Engineering Education. – 2020. – Vol. 28. – Rediscovering the use of chatbots in education. – № 6. – P. 1549–1565.–DOI: 10.1002/cae.22326.

[5] Рубцов, В. В. Инклюзивное образование: актуальный дискурс и перспективы развития / В. В. Рубцов, С. В. Алехина // Психологическая наука и образование. – 2024. – Т. 29. – Инклюзивное образование. – № 5. – С. 4.

[6] Салаватова, А. М. Нормативно-правовое обеспечение применения информационных технологий в социальной сфере, связанной с реабилитацией людей с инвалидностью / А. М. Салаватова, О. И. Истрофилова, Д. В. Ершов. – ООО «Издательский дом «Среда», 2024. – С. 19–22.

[7] Conversations with Documents: An Exploration of Document-Centered Assistance / M. ter Hoeve, R. Sim, E. Nouri [и др.]. – Текст : электронный // Proceedings of the 2020 Conference on Human Information Interaction and Retrieval : CHIIR ’20. – New York, NY, USA : Association for Computing Machinery, 2020. – С. 43–52. – URL: https://dl.acm.org/doi/10.1145/3343413.3377971 (дата обращения: 19.03.2026).

[8] A Comprehensive Analysis of Deep Learning Techniques for Documentation Classification / V. Kumar, A. Dabas, I. Jain, K. Pawar. – Текст : электронный // 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS). – Coimbatore, India : IEEE, 2021. – С. 228–235. – URL: https://ieeexplore.ieee.org/document/9395861/ (дата обращения: 19.03.2026).

[9] Retrieval-augmented generation for knowledge-intensive NLP tasks / P. Lewis, E. Perez, A. Piktus [et al.] – Текст : электронный // Proceedings of the 34th International Conference on Neural Information Processing Systems : NIPS ’20. – Red Hook, NY, USA : Curran Associates Inc., 2020. – URL: https://dl.acm.org/doi/abs/10.5555/3495724.3496517 (дата обращения 19.03.2026).

Авторский вклад

Тарасов Максим Михайлович – постановка задачи и проведение исследования, разработка требований, концепции и архитектуры системы, алгоритмов обработки данных.

Костылева Лилия Юрьевна – постановка задачи исследования, анализ источников, общее редактирование и структурирование текста, подготовка к публикации.

DESIGN AND DEVELOPMENT OF A RAG-ENABLED CHATBOT FOR GENERATING RECOMMENDATIONS IN INCLUSIVE EDUCATION

M.M. Tarasov
Master's student of School of Electronic Engineering and Computer Science, SUSU

L.Yu. Kostyleva
Associate Professor of Department of Information Systems and Technologies, School of Electronic Engineering and Computer Science, SUSU,
PhD of Technical Sciences

Abstract. The article examines the concept and architecture of a developing chatbot designed to assist faculty instructors at higher education institutions in organizing the educational process for students with disabilities (limited health capabilities). The proposed solution is based on modern Retrieval-Augmented Generation (RAG) technology, which enables the integration of large language models with up-to-date regulatory and methodological resources. Official methodological guidelines from the Russian Ministry of Science and Higher Education serve as the knowledge source. The system's functional modules are described, the rationale for the technology stack is justified, and the expected outcomes of implementation are outlined.

Keywords: Retrieval-Augmented Generation (RAG) technology, chatbot, inclusive education, students with disabilities, artificial intelligence in education, recommender system.

УДК 004.94:629.7.023:678.01:539.3

DATA-DRIVEN ПОДХОД К ОЦЕНКЕ НАПРЯЖЕННО-ДЕФОРМИРОВАННОГО СОСТОЯНИЯ СИЛОВОГО КАРКАСА БПЛА ИЗ ПОЛИМЕРНЫХ КОМПОЗИТОВ





Е.И. Бавбель
Аспирант кафедры проектирования информационно-компьютерных систем БГУИР
egorigorevichw9@gmail.com

В.Ф. Алексеев
Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент
alexvikt.minsk@gmail.com

Е.И. Бавбель
Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений и обработки информации.

В.Ф. Алексеев
Окончил Минский радиотехнический институт. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.

Аннотация. В статье рассматриваются особенности структурно-прочностного анализа силового каркаса беспилотного летательного аппарата, выполненного из композиционных материалов. Приведены основные расчетные зависимости, учитывающие анизотропные свойства материалов, а также рассмотрены характерные виды нагружения элементов конструкции. Проведен анализ напряженно-деформированного

состояния силовых элементов и оценено влияние конструктивных параметров на прочностные характеристики.

Ключевые слова: композиционные материалы, силовой каркас, прочность, напряжение, деформация, беспилотный летательный аппарат.

Введение. Современные летательные аппараты характеризуются возрастающими требованиями к снижению массы конструкции при одновременном обеспечении высокой прочности и жесткости [1]. Данные требования обусловлены необходимостью повышения энергетической эффективности, увеличения продолжительности полета и улучшения эксплуатационных характеристик. В этих условиях широкое распространение получили композиционные материалы, обладающие высокими удельными прочностными характеристиками и возможностью направленного формирования свойств за счет ориентации армирующих волокон.

Применение композиционных материалов в несущих элементах авиационных конструкций приводит к необходимости пересмотра традиционных подходов к расчету прочности, поскольку их поведение существенно отличается от изотропных металлических сплавов.

Силовой каркас летательного аппарата представляет собой совокупность элементов, воспринимающих основные аэродинамические и инерционные нагрузки. К числу таких элементов относятся лонжероны, нервюры, обшивка и силовые элементы фюзеляжа. При использовании композиционных материалов анализ прочности данных элементов требует учета анизотропии, слоистой структуры и возможных механизмов разрушения, включая межслойное расслоение и потерю устойчивости.

Процесс разрушения композиционных конструкций при статических и циклических нагружениях зависит от множества факторов: свойств волокон и матрицы, схемы укладки слоев, особенностей отверждения, температуры и содержания влаги. Накопление повреждений в виде микротрещин, отслоений и разрывов волокон приводит к снижению жесткости и прочности как отдельных слоев, так и структуры в целом, что требует разработки специализированных методов оценки напряженно-деформированного состояния.

Традиционные методы конечно-элементного моделирования обеспечивают достаточную точность расчета изотропных конструкций, однако для композиционных материалов необходима модификация подходов с учетом слоистой структуры и анизотропии свойств. Развитие методов машинного обучения и интеллектуального анализа данных (Big Data) открывает новые возможности для создания предиктивных моделей поведения композитных конструкций на основе накопленных экспериментальных и расчетных данных [2].

В связи с этим актуальной является задача разработки методов структурно-прочностного анализа, позволяющих адекватно оценивать напряженно-деформированное состояние элементов силового каркаса БПЛА из полимерных композитов и обеспечивать требуемый уровень надежности конструкции. Применение data-driven подходов позволяет существенно сократить объем вычислительных экспериментов при сохранении достаточной точности прогнозирования прочностных характеристик [3].

В результате исследований была выполнена разработка методологии оценки напряженно-деформированного состояния силового каркаса БПЛА из полимерных композитов на основе интеллектуального анализа данных.

Оценка работы коробчатого силового элемента из композиционного материала. В качестве расчетной схемы рассмотрен участок силового каркаса крыла, выполненный в виде замкнутого коробчатого элемента. Такая схема представляет практический интерес, поскольку сочетает сравнительно малую массу, высокую изгибную жесткость и способность воспринимать крутящие нагрузки. Несущая способность конструкции в этом случае определяется совместной работой поясов, стенок и обшивки.

При действии распределенной аэродинамической нагрузки в сечении крыла возникают изгибающий момент, поперечная сила и крутящий момент. В рамках предварительного анализа полагаем, что изгиб вносит основной вклад в формирование напряженного состояния. Тогда максимальный изгибающий момент в корневом сечении может быть записан в виде [4, с. 45]:

$$M_{max} = \frac{qL^2}{2}, \tag{1}$$

где: q–интенсивность распределенной нагрузки; L–длина консольной части.

В качестве исходных данных для вычислительного эксперимента приняты следующие значения: $q = 180$Н/м, $L = 1.8$м.

Тогда:

$$M_{max} = \frac{180 \cdot 1.8^2}{2} = 291.6\ .$$

Полученное значение позволяет заключить, что даже при сравнительно малой интенсивности распределенной нагрузки в корневой части крыла формируется существенный изгибающий момент. Таким образом, подтверждается, что определяющее значение при проектировании силового каркаса имеет именно корневая зона.

Для оценки напряженного состояния в корневом сечении крыла принято допущение, что коробчатый лонжерон обладает эквивалентным поперечным сечением с осевым моментом инерции $I = 1.4 \cdot 10^{-7}$м^4 , а расстояние от нейтральной оси до наиболее удаленного волокна составляет $y = 0.022$м. Тогда нормальные напряжения при изгибе определяются выражением [5, с. 30]:

$$\sigma = \frac{My}{I}. \tag{2}$$

Выполнив подстановку численных значений исходных параметров в соотношение (2), приходим к следующему результату:

$$\sigma = \frac{291.6 \cdot 0.022}{1.4 \cdot 10^{-7}} \approx 45.8 \cdot 10^6 \text{ Па} = 45.8 \text{ МПа}.$$

Полученный уровень напряжений не является критическим для углепластиковых материалов, однако сам по себе этот результат еще не означает достаточную надежность конструкции. Для композиционного элемента принципиально важно учитывать направление армирования, так как прочность вдоль волокон и поперек волокон различается на порядок.

Если принять продольный модуль упругости материала $E_1 = 130$ ГПа, а поперечный $E_2 = 9$ ГПа, то различие в деформационной реакции композиционного материала становится очевидным. При одинаковом уровне нормальных напряжений σ, то деформация вдоль направления армирования составит [6, с. 157]:

$$\varepsilon_1 = \frac{\sigma}{E_1}, \tag{3}$$

а поперечная деформация:

$$\varepsilon_2 = \frac{\sigma}{E_2}. \tag{4}$$

Тогда для найденного уровня напряжений:

$$\varepsilon_1 = \frac{45.8 \cdot 10^6}{130 \cdot 10^9} \approx 3.52 \cdot 10^{-4},$$
$$\varepsilon_2 = \frac{45.8 \cdot 10^6}{9 \cdot 10^9} \approx 5.09 \cdot 10^{-3}.$$

Сопоставление этих значений показывает, что поперечная деформация на порядок выше продольной. Данный результат свидетельствует о том, что при неоптимальной ориентации армирующих слоев даже умеренный уровень внешней нагрузки может обусловливать избыточную деформативность конструкции [7, с. 215]. Следовательно, при проектировании коробчатого силового элемента продольные слои $(0°)$ должны воспринимать основную часть изгибных напряжений, тогда как слои с ориентацией $\pm 45°$ призваны обеспечивать работу на сдвиг и повышать устойчивость стенок [8, с. 228].

Дополнительно оценим касательные напряжения в стенке коробчатого элемента. При воздействии поперечной силы:

$$Q = qL, \tag{5}$$

она принимает значение:

$$Q = 180 \cdot 1.8 = 324 \text{ Н}.$$

Если же эффективная площадь восприятия сдвига составляет $A_s = 6.5 \cdot 10^{-4}\text{м}^2$, то тогда среднее касательное напряжение можно определить как [9, с. 85]:

$$\tau = \frac{Q}{A_s}. \tag{6}$$

Подставляя исходные данные в (6) можно получить значение:

$$\tau = \frac{324}{6.5 \cdot 10^{-4}} \approx 0.50 \text{ МПа}.$$

Величина касательных напряжений здесь сравнительно невелика. Но для композитов именно сдвиг нередко становится главной причиной расслоения [8, с. 156]. Особенно это критично в зонах примыкания стенок к поясам коробчатого сечения – там возникает сложное напряженное состояние.

Прочность силового элемента оценим по критерию максимальных напряжений [8, с. 112]. Зададим допустимое напряжение вдоль армирования $[\sigma] = 800$ МПа, тогда коэффициент запаса по нормальным напряжениям можно найти по (7):

$$n = \frac{[\sigma]}{\sigma}. \tag{7}$$

$$n = \frac{800}{45.8} \approx 17.5.$$

Запас прочности вдоль армирования получился большой. Но это не значит, что вся конструкция избыточна. Реальную опасность представляют потеря устойчивости стенок, расслоение и местные напряжения в узлах крепления.

При выбранных исходных данных коробчатый элемент проходит по нормальным напряжениям. Интереснее другое: его работа определяется не столько уровнем напряжений, сколько анизотропией материала. Продольные слои хорошо работают на изгиб. Поперечные и наклонные – влияют на жесткость, устойчивость и сдвиг.

Деформации вдоль и поперек волокон различаются сильно. Значит, заменять композит эквивалентным изотропным материалом нельзя – занизим поперечные деформации и пропустим риски локального разрушения.

Даже при малых касательных напряжениях надо проверять межслойную прочность. Именно она определяет долговечность при переменных нагрузках. Поэтому проектировать каркас имеет смысл с учетом и геометрии, и схемы укладки, и реальных нагрузок.

Влияние схемы армирования на жесткость и массу. Композиты хороши тем, что свойства конструкции можно настраивать через укладку слоев. Для силового каркаса БПЛА это означает: жесткость и прочность меняются не только геометрией, но и ориентацией волокон.

В простейшем случае многослойный пакет состоит из слоев $0°$, $90°$и $\pm 45°$. Слои, ориентированные вдоль продольной оси элемента, в основном воспринимают нормальные напряжения изгиба. Слои с ориентацией $\pm 45°$ работают преимущественно на сдвиг и обеспечивают устойчивость конструкции к крутящим и поперечным нагрузкам. Слои $90°$ повышают поперечную жесткость и препятствуют локальному растрескиванию матрицы.

Для оценки изгибной жесткости эквивалентного коробчатого элемента, выполненного из слоистого композиционного материала, может быть использована следующая зависимость [8, с. 143]:

$$D = \sum_{k=1}^{n} \bar{Q}_{11}^{(k)} \frac{z_k^3 - z_{k-1}^3}{3}, \tag{8}$$

где: $\overline{Q_{11}^{(k)}}$ – приведённый модуль упругости k-го слоя в направлении оси x глобальной системы координат, вычисляемый через преобразование матрицы жёсткости слоя $[Q]$ с учётом угла ориентации волокон θ_k:

$$\overline{Q_{11}^{(k)}} = Q_{11}\cos^4\theta_k + 2(Q_{12} + 2Q_{66})\sin^2\theta_k\cos^2\theta_k + Q_{22}\sin^4\theta_k$$

где: $Q_{11}, Q_{12}, Q_{22}, Q_{66}$ – компоненты матрицы жёсткости однонаправленного слоя в главных осях материала, θ_k – угол укладки k –го слоя;
приведенная жесткость k-го слоя в направлении изгиба;

z_k, z_{k-1}– координаты границ слоя по толщине пакета.

Данное выражение показывает, что наибольший вклад в изгибную жесткость вносят внешние слои, удаленные от нейтральной поверхности. Это имеет принципиальное значение при проектировании силового каркаса: наиболее жесткие слои целесообразно располагать в наружных зонах сечения, тогда как внутренние слои могут использоваться для обеспечения сдвиговой прочности и устойчивости.

Наиболее рациональными являются схемы армирования, в которых продольные слои обеспечивают восприятие основных изгибных напряжений, а наклонные и поперечные слои – устойчивость, сдвиговую прочность и сопротивление локальному разрушению.

Сопоставление расчетных параметров и обобщение результатов анализа. Выполненный расчет и анализ силового элемента каркаса позволяют выявить ключевые закономерности, определяющие эффективность применения композиционных материалов в конструкциях летательных аппаратов. Полученные зависимости показывают, что прочностные и жесткостные характеристики определяются не только уровнем внешней нагрузки, но и совокупностью геометрических параметров, а также структурой материала.

Для обобщения результатов рассмотрим основные параметры, характеризующие работу силового элемента: уровень нормальных напряжений, относительные деформации, изгибную жесткость, прогиб и удельную эффективность конструкции. Сопоставление этих величин позволяет оценить влияние схемы армирования и геометрии на общее поведение конструкции. Сопоставление основных расчетных параметров силового элемента приведено в таблице 1.

Таблица 1. Сопоставление расчетных параметров силового элемента

Параметр	**Обозначение**	**Значение**	**Характер влияния**
Изгибающий момент	M	≈ 290 Н·м	Определяет уровень напряжений
Нормальные напряжения		≈ 45–50 МПа	Основной критерий прочности
Касательные напряжения		≈ 0.5 МПа	Влияют на межслойную прочность
Продольная деформация	1	~10^{-4}	Низкая, обеспечивает жесткость
Поперечная деформация	2	~10^{-3}	Значительно выше из-за анизотропии
Относительная изгибная жесткость	D	↑ до 1.33	Растет при увеличении доли слоев 0°
Относительный прогиб	f	↓ до 0.75	Снижается при увеличении жесткости
Относительная масса	m	↑ до 1.25	Возрастает с числом слоев
Удельная эффективность		≈ 1.06	Ограниченный рост

Анализ представленных данных показывает, что увеличение продольной жесткости за счет изменения схемы армирования приводит к снижению деформаций конструкции, однако сопровождается ростом массы. Выигрыш в удельной эффективности растет заметно медленнее, чем абсолютная жесткость.

Анизотропия композита сильно меняет характер деформации. Деформации вдоль и поперек волокон различаются принципиально – значит, слои надо ориентировать по направлению нагрузок. Иначе конструкция станет мягче и появятся локальные перегрузки.

Если сравнить нормальные и касательные напряжения, видно: даже при небольших сдвигах именно они могут разрушить конструкцию – за счет расслоения. Приходится учитывать не только прочность материала, но и его слоистую структуру.

В итоге работоспособность силового каркаса зависит не от одного параметра, а от их связки. Хорошая конструкция получается, когда геометрия, схема армирования и нагрузки подобраны совместно.

Заключение. Проведенный анализ силового элемента каркаса, выполненного из композиционного материала, показал, что его прочностные и жесткостные характеристики в значительной степени определяются не только уровнем действующих нагрузок, но и геометрическими параметрами сечения, а также схемой армирования.

Установлено, что при действии изгибающих нагрузок основное влияние на напряженное состояние оказывает изгибающий момент, достигающий максимального значения в корневой части конструкции. При этом даже при относительно невысоких значениях напряжений характер деформирования существенно зависит от анизотропии материала, что проявляется в значительном различии продольных и поперечных деформаций.

Показано, что увеличение доли слоев, ориентированных вдоль направления основных напряжений, приводит к росту изгибной жесткости и снижению прогибов, однако сопровождается увеличением массы конструкции и изменением ее чувствительности к

сдвиговым нагрузкам. В результате эффективность такого решения должна оцениваться с учетом совокупного влияния всех параметров.

Дополнительно отмечено, что при анализе композиционных конструкций определяющее значение могут иметь не максимальные нормальные напряжения, а сдвиговые и межслойные напряжения, а также устойчивость тонкостенных элементов. Это требует применения комплексного подхода к проектированию, включающего учет структуры материала и условий нагружения.

Полученные результаты подтверждают, что рациональное использование композиционных материалов в силовых элементах летательных аппаратов позволяет обеспечить требуемый уровень прочности при снижении массы конструкции. При этом ключевым фактором является согласование геометрии, схемы армирования и характера нагружения на этапе проектирования.

Список литературы

[1] Тарнопольский, Ю. М. Конструкционные полимерные композиционные материалы в авиационной и ракетно-космической технике / Ю. М. Тарнопольский, В. В. Хитров. – М. : ЦНИИТМАШ, 2014. – 320 с.

[2] Vassilopoulos, A. P. Fatigue life prediction of composites and composite structures / A. P. Vassilopoulos. – Cambridge : Woodhead Publishing, 2010. – 628 p.

[3] Liu, B. Surrogate-based aerodynamic shape optimization of hypersonic flows considering transonic performance / B. Liu [et al.] // Aerospace Science and Technology. – 2022. – Vol. 129. – P. 107849.

[4] Феодосьев, В. И. Сопротивление материалов : учебник для вузов / В. И. Феодосьев. – 10-е изд., перераб. и доп. – Москва : Изд-во МГТУ им. Н.Э. Баумана, 1999. – 588 с.

[5] Селиванов Ю. Т. Сопротивление материалов : учеб. пособие : в 2 ч. / Ю. Т. Селиванов. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2007. – Ч. 1. – 120 с.

[6] Димитриенко, Ю.И. Механика композиционных материалов : учебное пособие / Ю.И. Димитриенко. – Москва : Изд-во МГТУ им. Н.Э. Баумана, 2014. – 416 с.

[7] Jones R. M. Mechanics of Composite Materials / R. M. Jones. – 2nd ed. – Philadelphia : Taylor & Francis, 1999. – 538 p.

[8] Васильев, В. В. Механика конструкций из композиционных материалов / В. В. Васильев. – Москва : Машиностроение, 1988. – 272 с.

[9] Селиванов Ю.Т. Сопротивление материалов : учебное пособие : в 2 ч. / Ю.Т. Селиванов. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2007. – Ч. 1. – 120 с.

Авторский вклад

Бавбель Егор Игоревич – постановка задачи исследования, разработка расчетной модели силового элемента каркаса, выполнение анализа напряженно-деформированного состояния, интерпретация полученных результатов и подготовка текста статьи.

Алексеев Виктор Федорович – формулировка научной задачи, общее руководство исследованием, корректировка методологии расчета и редактирование текста статьи.

DATA-DRIVEN ASSESSMENT OF THE STRESS-STRAIN BEHAVIOR OF A COMPOSITE POLYMER AIRFRAME FOR UNMANNED AERIAL VEHICLES

E.I. Bavbel
Postgraduate student, Department of Information Computer Systems Design BSUIR

V.F. Alekseev
Associate Professor, Department of Information Computer Systems Design, BSUIR, PhD of Technical sciences, Associate Professor

Abstract. The paper considers the structural-strength analysis of a load-bearing frame element made of composite materials. The influence of geometric parameters and reinforcement scheme on the stress-strain state of the structure is investigated. The main relationships describing the behavior of anisotropic materials are presented. The obtained results demonstrate the effectiveness of composite materials in reducing weight while maintaining structural strength.

Keywords: composite materials, load-bearing structure, stress, deformation, stiffness, unmanned aerial vehicle

УДК 519.7

ОПТИМИЗАЦИЯ ДВУХУРОВНЕВЫХ ПРЕДСТАВЛЕНИЙ ЛОГИЧЕСКИХ СХЕМ С УЧЕТОМ ЭНЕРГОПОТРЕБЛЕНИЯ



Л.Д. Черемисинова
Профессор кафедры экономической информатики БГУИР,
Доктор технических наук, профессор,
l.cheremisinova@bsuir.by

Л.Д. Черемисинова
Область научных интересов: логико-комбинаторные задачи, логическое проектирование и тестирование дискретных систем управления, реализация параллельных алгоритмов управления.

Аннотация. Рассматривается проблема оптимизации двухуровневых логических И–ИЛИ представлений логических схем по критерию энергопотребления. Приводятся оценки энергопотребления, использование которых в процессе логического проектирования схем позволяет снизить энергопотребление при их реализации на кристалле КМОП СБИС. Предлагаются модификации известных методов минимизации булевых функций в классе дизъюнктивных нормальных форм путем добавления в них эвристик, направляющих процесс минимизации к получению представлений, реализуемых схемами с меньшим энергопотреблением.

Ключевые слова: логический синтез, КМОП СБИС, рассеивание мощности, минимизация булевых функций.

Введение. Сокращение энергопотребления цифровых блоков является одной из важнейших проблем, возникающих при проектировании заказных СБИС, выполняемых по КМОП технологии. Это связано с тем, что рассеивание мощности становится препятствием на пути повышения плотности монтажа микросхем [1, 2], а также с увеличением числа номенклатуры цифровых систем, которые должны работать длительное время без подзарядки батареи питания. В частности проблема снижения рассеивания мощности микросхем становится определяющей в космических применениях [3] вследствие автономности энергопитания космических систем. В последние годы фактор энергопотребления при проектировании интегральных схем стал играть такую же важную роль, как площадь и быстродействие. Как правило, при проектировании стараются достичь высокой эффективности при ограниченном потреблении электроэнергии. Проблема энергосберегающего проектирования до сих пор является по большому счету искусством. Это обусловлено отсутствием средств оценки эффективности используемых в процессе проектирования эвристик, направленных на снижение рассеивания мощности реализуемой в кремнии схемы. Проблемами проектирования устройств с низким рассеиванием мощности занимаются ведущие в области САПР фирмы, такие как Cadence Design Systems, Synopsys, Mentor Graphics, Apache Design, и др.

Снижение энергопотребления проектируемой схемы может обеспечиваться на разных уровнях проектирования: от алгоритмического и системного, до топологического и схемного уровней. При этом, чем более ранним является этап, тем большее влияние он оказывает на качество проекта. В частности на логическом уровне (за счет построения

удачной логической структуры) можно достичь сокращения рассеивания мощности на 10–20% без ущерба для быстродействия и сложности схемы [4]. САПР микроэлектронных схем должны иметь средства, позволяющие оценивать и минимизировать энергопотребление схем в процессе их проектирования, чтобы избежать дорогостоящей процедуры их перепроектирования на стадии схемотехнического проектирования.

В настоящей статье рассматриваются задача снижения рассеивания мощности (энергопотребления) при минимизации двухуровневых представлений логических схем по критерию энергопотребления. Рассматриваются оценки энергопотребления на этом этапе, использование которых позволяет обеспечить минимум энергопотребления реализованной на кристалле СБИС схемы. Под энергопотреблением далее понимается оценка средней мощности, рассеиваемой микросхемой. Рассматривается случай схем, реализуемых по КМОП-технологии. Предполагается, что основные схемотехнические решения подлежащей проектированию схемы, такие как частота синхронизации и напряжение питания, фиксированы.

Составляющие мощности, потребляемой микросхемой. В КМОП-технологии всю рассеиваемую микросхемой мощность можно разделить на статическую и динамическую. Первая составляющая обусловлена наличием токов утечки или статических проводящих путей между шинами питания [1, 4]. Она присутствует даже тогда, когда схема бездействует (находится в режиме ожидания). У большинства хорошо спроектированных КМОП-схем статическая рассеиваемая мощность очень мала. Около 80% всей рассеиваемой энергии приходится на ее динамическую составляющую [2], порождаемую нестационарным поведением узлов микросхемы.

Наиболее существенной причиной потребления мощности КМОП-схемой в динамике является динамическая рассеиваемая мощность P_{dyn}, обусловленная токами заряда/разряда паразитных емкостей транзисторов и линий связи во время переходных процессов, когда сигналы на выходах узлов микросхемы переключаются. Эти емкости учитываются в виде емкостной нагрузки C_L выхода узла [1]. При смене входного сигнала с 0 на 1 ток, протекающий через транзистор с *p*-каналом, заряжает емкость C_L, при смене входного сигнала с 1 на 0 емкость C_L разряжается током, текущим через транзистор с *n*-каналом. В каждом из этих случаев на сопротивлении открытого транзистора рассеивается мощность, выражаемая следующим соотношением [5, 6]:

$$P_{dyn} = \frac{1}{2} V_{dd}^2 f_{clk} E_s C_L,$$

где V_{dd} – напряжение питания, f_{clk} – частота синхронизации, определяющая среднюю периодичность изменения состояния входа схемы, E_s – среднее число переключений сигнала на выходе узла в течение одного такта работы схемы. Составляющая E_s определяется как математическое ожидание числа логических переходов (из 1 в 0 или из 0 в 1) за один период синхронизации и называется в литературе переключательной активностью. Рассеивание энергии микросхемой в целом можно оценить суммой мощностей, потребляемых всеми ее узлами [2, 7]:

$$P_s = \frac{1}{2} V_{dd}^2 f_{clk} \sum_{i=1}^{n} E_i C_i,$$

где n – число узлов в схеме; C_i – емкостная нагрузка i–го узла; E_i – переключательная активность i–го узла схемы.

Из этой формулы видно, что на уровне логического проектирования можно решать задачу энергосбережения, управляя числом переключений узлов схемы. Если в схеме нет переключений значений сигналов на ее входных и внутренних полюсах, то отсутствуют и затраты динамической энергии. Значения параметров V_{dd} и f_{clk} определяются используемой КМОП-технологией, поэтому минимизация динамической мощности P_s на уровне логического проектирования сводится к минимизации переключаемой емкости E_iC_i. На

логическом уровне можно управлять числом переключений узлов схемы, а емкостную нагрузку C_i узла оценивать коэффициентами разветвления их выходов.

Подходы к оценке энергопотребления логических схем. Известны два подхода [4, 8, 9, 10] к оценке энергопотребления схем: на основе динамического анализа и на основе статического анализа. Первый подход предполагает моделирование исследуемой схемы на множестве возможных сценариев ее функционирования (упорядоченных последовательностей наборов значений сигналов на ее входах) [11]. Наиболее известным и достаточно эффективным средством моделирования электрических схем является SPICE (Simulation Program with Integrated Circuit Emphasis) – программа моделирования с ориентацией на интегральные схемы (разработка Калифорнийского университета Беркли) [12]. В процессе моделирования схемы на заданной последовательности наборов значений входных сигналов в форме диаграмм входных напряжений вычисляются усредненные формы диаграмм выходных напряжений и токов узлов схемы, по которым вычисляется среднее энергопотребление схемы с учетом технологии ее изготовления. Однако результаты моделирования заданной электрической схемы существенно зависят от выбора тестируемой последовательности: от того какие и сколько наборов значений входных сигналов используется, в какой последовательности они подаются, какова форма сигналов и т.д. Методы оценки энергопотребления схем, основанные на моделировании, не годятся для использования в процессе проектирования схемы, когда ни она, ни ее окружение еще не спроектированы. В этом случае практически ничего не известно о возможных входных воздействиях на схему, что делает использование методов оценки энергопотребления на основе на моделирования практически невозможными для сложных схем [13].

В основе статических методов оценки энергопотребления схем лежит вероятностный подход к определению переключательной активности сигналов, он основан на вероятностных характеристиках входных сигналов и функционально-структурных свойствах исследуемой схемы. Эти методы обеспечивают компактное описание последовательности наборов входных сигналов. Подход предполагает задание вероятности переключения сигналов на каждом из входов схемы. В основе подхода лежит распространение вероятностной информации о смене значений сигналов от входов к выходам схемы [11]. Вероятностный подход позволяет 1) компактно определить возможные последовательности входных воздействий на схему и оценить энергопотребление схемы, исходя из взаимного влияния сигналов в последовательные моменты времени, и 2) избежать повторения циклов моделирования схемы на задаваемом большом множестве наборов значений входных сигналов.

Оценки переключательной активности сигналов, используемые при оптимизации функциональных описаний. Эти оценки выведены в предположении, что все полюсы схемы независимы (пространственная независимость) и значения сигналов на каждом из полюсов в двух соседних циклах синхронизации также независимы (временна́я независимость); В предлагаемых методах оптимизации двухуровневых представлений логических схем используется два подхода к оценке переключательной активности сигналов: вероятностная и энтропийная.

Первая оценка выведена для случая нулевой задержки сигналов узлами схемы. Различают [2, 4] вероятность p_i^1 (p_i^0) появления сигнала 1 (0) на некотором i-м полюсе и вероятность E_i смены сигнала на полюсе. Первая вероятность p_i^1 называется сигнальной (обозначается далее как p_i) и определяется средней долей тактов, на которых сигнал на i-м полюсе имеет значение 1. Вторая вероятность называется переключательной активностью i-го полюса (обозначается далее через E_i) и определяется средней долей тактов, на которых сигнал на нем меняет свое значение (относительно предшествующего такт). В предположении, что $p_i^1, p_i^0 < 1$ переключательная активность вычисляется через сигнальную вероятность p_i

$$E_i = 2p_i^1 p_i^0 = 2p_i(1 - p_i). \quad (1)$$

Вероятность p_y появления сигнала 1 на выходе y элемента зависит от реализуемой им функции. Для инвертора, функций И и ИЛИ при сделанных выше предположениях, сигнальные вероятности вычисляются по формулам:

$$p_{y^{\leftarrow}} = 1-p;$$
$$p_{y^{\wedge}} = \prod_{i=1}^{n(e)} p_i; \qquad (2)$$
$$p_{y^{\vee}} = 1 - \prod_{i=1}^{n(e)} (1-p_i).$$

Энтропийная оценка отличие от переключательной активности адекватно отражает внутренние задержки схемы, в ее основе лежат понятия равновесной вероятности и интенсивности переключений [8], которые определяют поведение сигнала не только в устойчивом состоянии в течение такта, но и во времени. Равновесная вероятность P_x сигнала $x(t)$ равна средней доле времени такта, в течение которого сигнал имеет значение 1. Интенсивность переключений A_x определяется как среднее число переключений сигнала $x(t)$ в единицу времени.

Показано [11], что для случая пространственной независимости входов x_i элемента, интенсивность переключений сигнала на выходе y вычисляется через равновесную вероятность P_x сигнала

$$A_y = \sum_{i=1}^{n} P(\frac{dy}{dx_i}) A_{xi}$$

Булев дифференциал. dy/dx функции y, зависящей от x, определяется как:

$$\frac{\partial y}{\partial x} \equiv y|_{x=1} \oplus y|_{x=0} = y(x) \oplus y(\bar{x}).$$

Если через $\lceil_1$ (и $\lceil_0$) обозначить среднее время между переключениями входного сигнала $x(t)$ в 1 (и 0), то среднее значение времени между изменениями значения сигнала $x(t)$ равно $\lceil = ½(\lceil_0 + \lceil_1)$, а вероятность переключения входного сигнала $x(t)$ в 1 равна $p(x) = \lceil_1/(\lceil_0 + \lceil_1)$, а интенсивность переключений A_x сигнала $x(t)$ определяется как $A_x = 2/(\lceil_0 + \lceil_1)$. Учитывая это соотношение между сигнальной вероятностью и интенсивностью переключений, формулу (5) для оценки энергопотребления можно упростить следующим образом [14]:

$$A(y_j) = \sum_{i=1}^{n_j} p(\frac{\partial y_i}{\partial x_j}) p(x_i) \qquad (3)$$

Учет энергопотребления при оптимизации двухуровневых представлений схем. Проблема минимизации двухуровневых представлений логических схем (или минимизации булевых функций в классе ДНФ) состоит в поиске минимальной по некоторому критерию системы ДНФ. Все методы минимизации ДНФ [15] основаны на поиске всех или некоторого подмножества простых импликант и выборе среди них приведенного подмножества, представляющего собой некоторую систему безызбыточных ДНФ. В качестве критериев качества минимизации обычно используется суммарное число элементарных конъюнкций во всех ДНФ системы или общее число литералов всех конъюнкций. Эти критерии качества минимизации системы ДНФ нацелены на сокращение площади СБИС и, как правило, не зависят от технологии изготовления проектируемых

схем. Специфика задачи минимизации по критерию энергопотребления состоит в том, что целевая система ДНФ должна быть минимальной не только по числу конъюнкций или литералов, но и по сумме некоторых весовых характеристик входящих в нее конъюнкций, которые отражают переключательную активность сигналов на входах и выходе элемента И, реализующего конъюнкцию. Эта оценка, исходя из сигнальных вероятностей входов элемента И, вычисляется по формулам (1), (2) и (3).

Трудность в нахождении минимального решения по энергопотреблению состоит в том, что часто таковыми являются решения, соответствующие не простым импликантам (а импликантам, содержащим избыточные литералы). В то время как, получение минимальных решений по площади, занимаемой микросхемой на кристалле, основано на поиске только простых импликант. В процессе оптимизации функциональных описаний устройств площадь и энергопотребление схемы в процессе логического синтеза тесно связаны в том смысле, что уменьшение площади схемы позволяет, как правило, также снизить ее энергопотребление, а увеличение площади, наоборот, ведет также к увеличению энергопотребления. Из этих соображений в процессе минимизации необходим компромисс критериев минимизации энергопотребления и площади. Компромиссным решением является минимизация, в процессе которой оценка получаемых текущих безызбычных ДНФ на предмет принятия их в качестве решения производится на основе ранжированного критерия: сначала площадь, потом энергопотребление, или наоборот. В первом случае находятся несколько решений с минимумом площади, из которых выбирается решение с минимумом энергопотребления. Во втором случае критерии отбора применяются в противоположном порядке.

Практически все методы минимизации двухуровневых представлений булевых функций основаны на разделении множества искомых простых импликант на три подмножества: существенных, несущественных и условно существенных. Первые должны быть включены в любое безызбыточное решение, вторые не должны включаться ни в одно решение, а из третьих выбирается некоторое безызбыточное подмножество, покрывающее все интервалы единичных областей задания минимизируемых функций, которые оказались не покрытыми существенными импликантами. Методы минимизации различаются способами построения простых импликант из числа условно существенных и критериями, которым они должны удовлетворять для включения их в решение. Различают методы: 1) последовательного построения простых импликант, включаемых в решение (например, метод конкурирующих интервалов [15]), путем укрупнения интервала булева пространства аргументов, представляющего импликанту, за счет включения в него непокрытых элементарных конъюнкций; 2) последовательного расширения интервала булева пространства аргументов, представляющего сначала конъюнкцию исходной ДНФ, а в перспективе простую импликанту, покрывающую эту конъюнкцию, за счет исключения некоторых литералов, входящих в нее (например, методы реализуемые ESPRESSO [16]).

Простейший метод минимизации может использовать лишь одну операцию расширения условно существенных импликант. Наиболее просто с целью учета энергосбережения модифицируются методы минимизации, в которых в качестве кандидатов в искомое решение строится сразу несколько простых импликант, или методы, в которых найденное решение модифицируется методом последовательных улучшений.

Модификация методов минимизации. Для того чтобы направить минимизацию к получению энергосберегающего решения, необходимо в процессе получения простых импликант и безызбыточных покрытий вычислять и учитывать переключательные активности всех простых импликант. Для пояснения процедур, направляющих процесс минимизации на получение нужного результата, рассматриваются операции, входящие практически во все методы минимизации.

Операция расширения интервала выполняется путем исключения его литералов. При расширении интервала принимаются во внимание две цели: уменьшить сложность этого

интервала, т.е. максимально расширить его, и покрыть с его помощью (полностью или частично) как можно больше непокрытых еще интервалов. При минимизации энергопотребления желательно 1) исключать не любые, а наиболее активные литералы; 2) покрывать не любые, а наиболее энергоемкие интервалы. В силу первого утверждения, на предмет исключения проверяются сначала литералы с большей переключательной активностью. В силу второго утверждения, важен порядок расширения интервалов: слишком раннее расширение интервала может препятствовать ситуации, когда некий другой интервал покроет этот рассматриваемый. Чтобы снизить энергопотребление, оценивается энергетический вклад каждого интервала, и энергоемкие (с большой переключательной активностью) интервалы расширяются последними, в надежде на то, что некоторые другие интервалы расширятся и покроют их.

Операция нахождения безызбыточного покрытия заключается в приведении текущего покрытия ДНФ к безызбыточному виду. При поиске безызбыточного множества простых импликант выбирается минимальное число наименее активных (с меньшей переключательной активностью) импликант. При этом из безызбыточных множеств простых импликант, представляющих собой покрытие исходного множества интервалов единичных областей минимизируемых булевых функций, отбираются минимальные по мощности или по суммарному числу литералов всех импликант. Каждое из отобранных множеств оценивается суммой переключательных активностей (или интенсивностей переключений) входящих в него импликант, и в качестве решения выбирается безызбыточное множество с минимальной оценкой переключательной активности.

В предлагаемых модификациях методов оптимизации двухуровневых представлений схем приоритет отдается площади, и среди равных по этому критерию выбирается вариант, лучший по энергопотреблению. Модификации известных методов минимизации булевых функций в классе ДНФ заключались в добавлении в них эвристик, направляющих процесс минимизации к получению систем ДНФ, реализуемых схемами с наименьшим энергопотреблением. За основной критерий минимизации принимается число литералов результирующей системы ДНФ, за второй – энегопотребление, оцениваемое суммарной переключательной активностью первого уровня двухуровневой схемы, представляющего множество конъюнкций:

$$P_{rs} = \sum_{r=1}^{m} \left(E(k_r) + \sum_{i=1}^{n(k_r)} E(x_i) \right),$$

где $E(k_r)$ и $E(x_i)$ – переключательные активности выхода элемента И, реализующего r–ю конъюнкцию, и i-го входного полюса; $n(k_r)$ – число литералов импликанты k_r. m – число импликант минимизированной формы системы булевых функций.

Модифицированная программа минимизации ESPRESSO реализует приближенный метод минимизации систем как полностью, так и частично определенных булевых функций. Результатом минимизации является безызбыточная система ДНФ. Сигнальные вероятности для входов двухуровневой схемы задаются в виде отдельного файла. Выбор метода минимизации происходит автоматически, в зависимости от исходных данных и дополнительных настроек.

В модифицированном методе ESPRESSO так же, как и в ESPRESSO-MV [16] (рисунок 1) в процессе минимизации итеративно выполняются операции: ESSENTIAL выделения существенных импликант; EXPAND расширения покрываемого интервала; REDUCE сокращения интервала; RESHAPE модификации текущего покрытия и IRREDUNDANT получения безызбыточного покрытия.

Операция ESSENTIAL находит существенные импликанты, которые входят в любое безызбыточное покрытие, поэтому соответствующие им простые импликанты сразу переносятся в искомую безызбыточную ДНФ в качестве ее обязательной части. Операция EXPAND производит расширение интервала булева пространства аргументов, представляющего некоторую конъюнкцию или импликанту. В процессе выполнения этой

операции анализируются литералы рассматриваемой конъюнкции на предмет возможности их исключения. Операция REDUCE сокращения интервала булева пространства является обратной расширению EXPAND. С ее помощью интервал булева пространства анализируется на возможность его сокращения путем добавления в соответствующую ему конъюнкцию некоторых литералов. Операция RESHAPE модифицирует текущее покрытие ДНФ путем замены пары смежных конъюнкций на пару новых смежных конъюнкций. При этом один интервал расширяется, другой сокращается. Операция IRREDUNDANT приводит текущее покрытие ДНФ к безызбыточному виду. Циклическое выполнение процедур REDUCE, EXPAND, IRREDUNDANT повторяется до тех пор, пока решение не стабилизируется, что означает, что оно достигло некоторого локального минимума. Для сбрасывания полученного решения с этого минимума используются более изощренные, чем REDUCE, процедуры.

В режиме минимизации энергопотребления (если заданы вероятности переключений) перед выполнением операции EXPAND интервалы упорядочиваются в порядке возрастания энергопотребления. С точки зрения снижения оценки энергопотребления важен порядок расширения интервалов: слишком раннее расширение интервала может исключить возможность его покрытия неким другим интервалом. Для этого, энергоемкие интервалы берутся для расширения последними, в надежде, что некоторые другие интервалы расширятся и покроют их. При самом выполнении операции EXPAND делаются попытки исключать не любые, а наиболее переключаемые литералы. Литералы рассматриваются в порядке убывания переключательной активности, т.е. сначала делается попытка исключить литералы с большей переключательной активностью.

При выполнении операции сокращении интервала REDUCE с учетом минимизации энергопотребления делаются попытки включать в него наименее переключаемые литералы. При построении безызбыточного множества простых импликант с помощью операции REDUCE предлагается выбирать минимальное число наименее энергоемких импликант (с меньшей интенсивностью переключений выхода). Каждое из формируемых множеств простых импликант, представляющих возможные безызбыточные ДНФ, оценивается суммой интенсивностей переключений входящих в него импликант, и в качестве решения выбирается безызбыточное множество с минимальной оценкой переключения.

Для установления влияния предлагаемых модификаций методов минимизации на энергопотребление получаемых схем были проведены экспериментальные исследования программных реализаций методов минимизации без учета и с учетом энергопотребления [17]. Исследования проводились на одних и тех же тестовых примерах из набора. В таблице 1 приведены средние значения тока I^1, I^2, потребляемого схемами в миллиамперах (мА), – площади s^1, s^2 КМОП-схем, построенных по минимизированным системам ДНФ и сложности исходных систем ДНФ: n, m и k – числа переменных, функций и конъюнкций исходных систем ДНФ. Жирным шрифтом выделены варианты, выигравшие по энергопотреблению.

Таблица 1. Результаты моделирования схем

Пример	Система ДНФ			Первой схемы		Второй схемы	
	n	m	k	s^1	I^1	s^2	I^2
b12	15	9	431	33575	0,47293	35176	**0,47201**
in0	15	11	138	141141	**1,09385**	157802	1,28238
life	9	1	512	18341	0,30817	18922	**0,26176**
mlp4	8	8	256	155425	0,59325	158020	**0,58931**
root	8	5	256	41789	0,38138	41158	**0,38136**
tms	8	16	30	56838	**0,62831**	60297	0,67691
z9sym	9	1	420	38396	0,51769	43616	**0,49248**

Заключение. Показано, что минимизация булевых функций с учетом переключательной активности сигналов позволяет получить уменьшение энергопотребления схем без увеличения их сложности.

Список литературы

[1] Рабаи Ж.М., Чандракасан А., Николич Б. Цифровые интегральные схемы. Методология проектирования. М.: ООО «И.Д. Вильямс», 2007. – 912 с.

[2] Roy K., Prasad S.C. Low Power CMOS VLSI Circuit Design. New York: John Wiley and Sons Inc., 2000. – 376 p.

[3] Белоус, А.И., Космическая электроника / А.И.Белоус, В.А. Солодуха, С.В. Шведов. Кн. 1. М: Техносфера, 2015, 696 с.

[4] Benini L., De Micheli G. Logic Synthesis for Low Power // Logic Synthesis and Verification (eds. S. Hassoun, T. Sasao, R.K. Brayton). Boston, Dardrecht, London : Kluwer Academic Publishers, 2002. P. 197–223.

[5] Chandrakasan A.P. Low Power CMOS Digital Design // IEEE Journal on Solid-State Circuits. – 1992. – Vol. 27. – No. 4. – P. 473–482.

[6] Holt G., Tyagi A. Minimizing Interconnect Energy through Integrated Low-Power Placement and Combinational Logic Synthesis // In ISPD, California USA. – 1997. – P. 48–53.

[7] Pedram M. Power Minimization in IC Design: Principles and Applications // ACM Transactions Design Automation Electronic Systems. – 1996. – Vol.1. – P. 3–56.

[8] Najm F.N. Transition density: A new measure of activity of digital circuits IEEE Trans. on Computer-Aided Design of Integr. Circuit and Systems. – 1993. – Vol. 12. – No 2. – P. 310–323.

[9] Buch P., Lin S., Nagasamy V., Kuh E.S. Techniques for fast circuit simulation applied to power estimation of CMOS circuits // Proc. of the Intern. Symposium on Low Power Design. – Dana Point, CA, April 23–26, 1995.

[10] Ghosh A., Devadas S., Keutzer K., White J. Estimation of average switching activity in combinational and sequential circuits // 29th ACM / IEEE Design Automation Conference, Tech. Dig. – June 1992. – P. 253–259.

[11] Najm F.N. A survey of Power Estimation Techniques in VLSI Circuits // IEEE Trans. on VLSI. – 1994. – No 12. – P. 446–455.

[12] Nagel L. SPICE2: a computer program to simulate semiconductor circuits. – Memo ERL-M520. Dept. Elect. and Computer Science, University of California at Berkeley. – 1975.

[13] Rajgopal S., Mehta G. Experiences with simulation-based schematic level current estimation // Proc. of the Intern. Workshop on Low Power Design. – 1994, April. – P. 9–14.

[14] Bahar R.I., Somenzi F. Boolean techniques for low power driven re-synthesis // Computer-Aided Design, 1995. ICCAD-95. Digest of Technical Papers., 1995 IEEE/ACM International Conference – 1995. – P. 428 – 432.

[15] Закревский А.Д., Поттосин Ю.В., Черемисинова Л.Д. Основы логического проектирования. Книга 2. Оптимизация в булевом пространстве. - Минск, ОИПИ НАН Беларуси, 2004. – 240 с.

[16] Brayton R.K., Hachtel G.D., McMullen C., Sangiovanni-Vincentelli A.L. Logic minimization algorithms for VLSI synthesis. – Boston, Massachusetts: Kluwer Academic Publishers, 1984. – 193 p.

[17] Черемисинова Л.Д., Черемисинов Д.И. Оптимизация двухуровневых логических схем с учетом энергопотребления // Проблемы разработки перспективных микро- и наноэлектронных систем – 2010. Сб.трудов / под общ.ред. акад. РАН А.Л. Стемпковского. – М.: ИППМ РАН. – 2010. – С. 14–19.

LOW POWER DRIVEN OPTIMIZATION OF TWO-LEVEL REPRESENTATIONS OF LOGICAL CIRCUITS

L.D. Cheremisinova

Professor at the Department of Economic Informatics, Belarusian State University of Informatics and Radioelectronics, Doctor of Technical Sciences, Professor, *l.cheremisinova@bsuir.by*

Abstract. The paper examines the problem of optimizing two-level AND-OR logic representations of logic circuits based on power consumption criterion. Power consumption estimates are presented, the use of which in the logic design process allows to reduce the power consumption of their implementation on a CMOS VLSI chip. Modifications to existing methods for minimizing Boolean functions in the class of disjunctive normal forms are proposed by adding heuristics that guide the minimization process toward obtaining representations implemented by circuits with lower power consumption.

Keywords: logic design, CMOS VLSI, power dissipation, minimization of Boolean functions.

УДК 796.012:004.8

АНАЛИЗ БИОМЕХАНИЧЕСКИХ ПРИЗНАКОВ ДВИЖЕНИЙ ДЛЯ ПРОГНОЗИРОВАНИЯ ДВИГАТЕЛЬНОЙ ОДАРЕННОСТИ ДЕТЕЙ







Г.Г. Григорьева
Старший преподаватель кафедры педагогики и психологии ФГБОУ ВО «ЧГИФКиС» ga_et@inbox.ru

Н.С. Давыдова
Доцент кафедры инфокоммуникационных технологий БГУИР, кандидат технических наук, доцент davydova-ns@bsuir.by

В.Е. Васюк
Доцент Высшей школы спортивной педагогики СПбПУ, кандидат педагогических наук

М.В. Давыдов
Доцент, кандидат технических наук, первый проректор БГУИР

К.И. Татарко
Доцент кафедры социальной и организационной психологии БГУ, кандидат психолог. наук, доцент

Аннотация. В статье рассматривается задача прогнозирования двигательной одаренности детей на основе мультимодальных биомеханических данных, полученных при анализе беговых локомоций. Использованы кинематические и динамические параметры бега, зарегистрированные с применением системы безмаркерного захвата движения и динамометрической платформы. На выборке из 26 детей среднего школьного возраста сформирован комплексный датасет, включающий 30 биомеханических параметров и экспертные оценки. Для отбора информативных показателей разработана параметрическая модель бега, позволившая структурировать кинематические и динамические характеристики движения. На основе коэффициента разделения выделены 14 наиболее значимых признаков, после чего выполнено прогнозирование уровня двигательной одаренности с использованием метода Random Forest Classifier. Наилучшие результаты классификации оставили accuracy 0,96. Показано, что предложенный подход позволяет не только прогнозировать уровень двигательной одаренности, но и выделять биомеханические параметры, наиболее значимые для спортивного отбора и педагогической диагностики.

Ключевые слова: двигательная одаренность, беговые локомоции, мультимодальные данные, биомеханический анализ, машинное обучение, параметрическая модель, отбор признаков.

Введение. Ранняя диагностика двигательной одаренности детей остается одной из ключевых задач спортивного отбора, поскольку эффективность подготовки спортивного резерва во многом зависит от своевременного выявления перспективных двигательных способностей [1–4]. В литературе двигательная одаренность рассматривается как интегративное качество, включающее физические, координационные и функциональные предпосылки к успешной спортивной деятельности [1, 2]. При этом традиционные подходы к отбору по-прежнему во многом опираются на педагогические наблюдения, тестовые задания и экспертные оценки, что ограничивает объективность и воспроизводимость результата [3, 4].

Перспективным направлением решения данной задачи является биомеханический анализ беговых локомоций. Бег относится к информативным формам естественной двигательной активности, в которой проявляются ритмо-темповая организация движения, межзвенная координация и характер опорного взаимодействия [5–8]. Современные системы безмаркерного захвата движения и силовые платформы позволяют регистрировать

кинематические и динамические характеристики с высокой точностью, формируя основу для количественной оценки двигательного потенциала детей [6–9].

Однако большое число регистрируемых параметров требует отбора информативных признаков и интерпретации результатов моделирования. Для практического использования интеллектуальных систем в раннем спортивном отборе важно не только получить высокую точность классификации, но и определить, какие именно биомеханические показатели в наибольшей степени влияют на итоговое решение модели [10–12]. Цель исследования – разработать подход к анализу биомеханических признаков беговых локомоций для прогнозирования двигательной одаренности детей.

Материалы и методы исследования. В исследовании биомеханических характеристик бега принимали участие 26 детей среднего школьного возраста (10,33±1,15 лет), не имеющих узкой спортивной специализации.

Для регистрации кинематических данных применялась система безмаркерного захвата движения Simi Reality Motion Systems, а для регистрации динамических параметров – стационарная динамометрическая платформа Bertec [8, 9]. Схема экспериментальной установки представлена на рисунке 1.

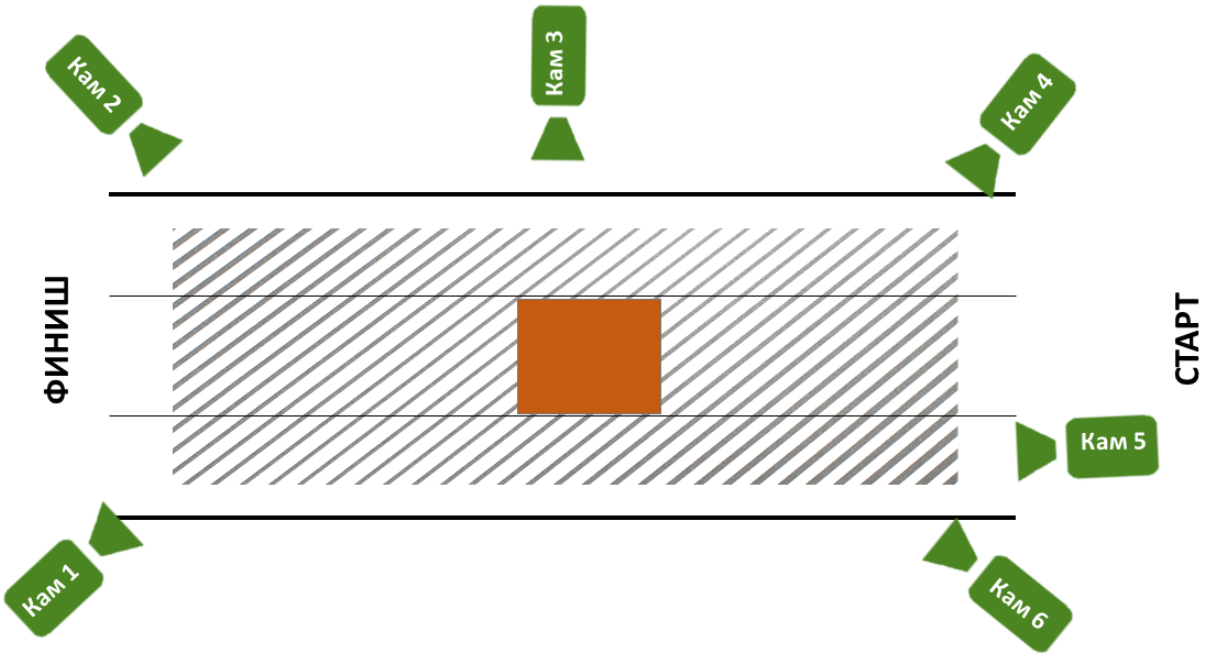


Рисунок 1. Схема экспериментальной установки с расположением камер системы безмаркерного захвата движения и силовой платформы

Уровень двигательной одаренности определяется на основе экспертной оценки, выполненной специалистами в области спортивной педагогики по совокупности параметров техники бега и проявлений двигательного потенциала. Экспертные заключения использовались как целевые метки при построении обучающей выборки и сопоставлялись с количественными биомеханическими данными [11].

Параметрическая модель бега. Для отбора биомеханических признаков двигательной одаренности была разработана параметрическая модель бега, основанная на представлении беговой локомоции как целостной системы взаимосвязанных кинематических и динамических характеристик. Модель учитывает фазовую структуру бегового шага, особенности координации, ритмо-темповую организацию движения и

параметры взаимодействия с опорой. Такой подход позволил перейти от анализа отдельных показателей к системному описанию двигательного действия, отражающему как внешнюю структуру бега, так и механизмы его реализации.

При выборе параметров для детей особое внимание уделялось показателям наиболее чувствительным к проявлениям двигательного потенциала и в то же время устойчивым к возрастной вариативности и отсутствию узкой спортивной специализации.

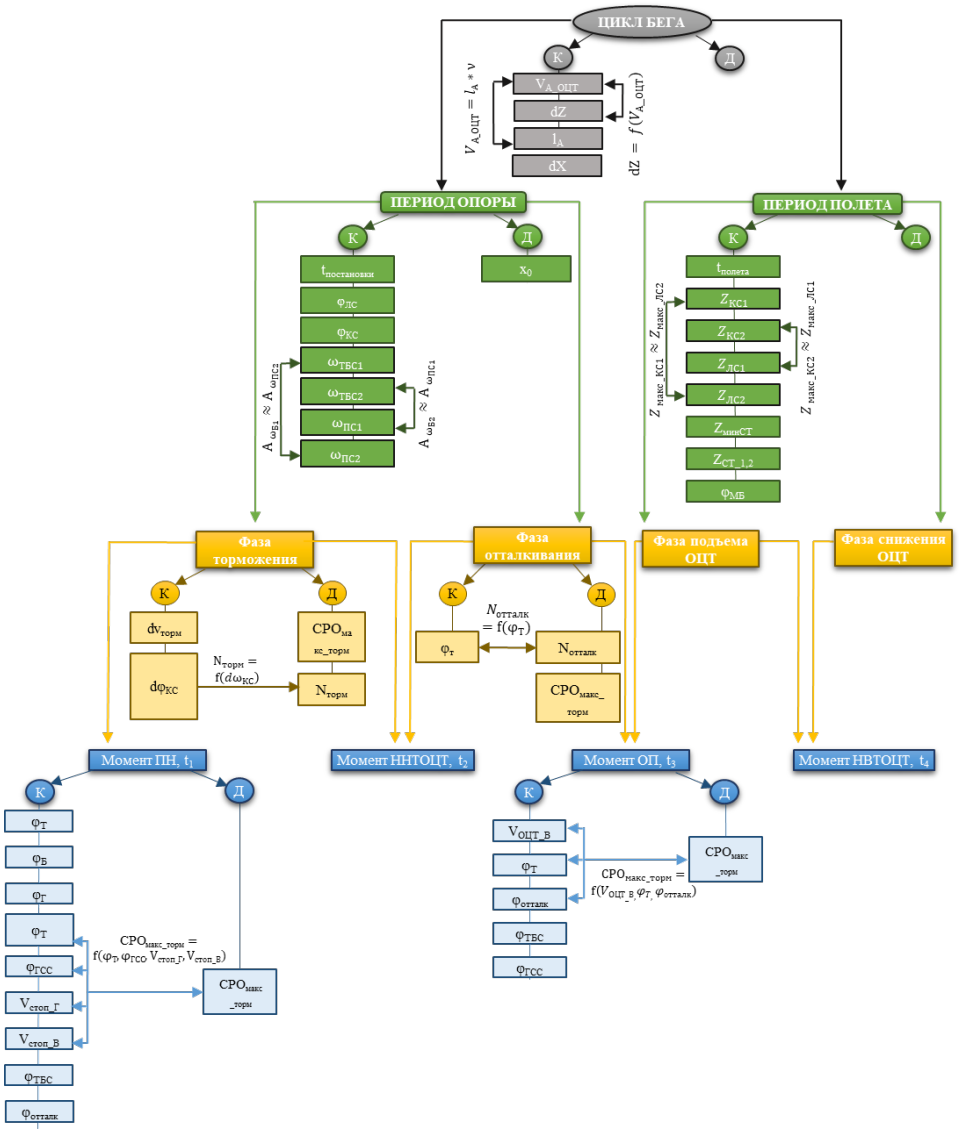


Рисунок 2. Параметрическая модель бега, используемая для выделения информативных биомеханических признаков

Формирование мультимодального признакового пространства. В структуру признакового пространства вошли три группы показателей: пространственно-временные, кинематические и динамические характеристики. Пространственно-временные параметры отражали общую организацию движения и включали длину и частоту шагов, длительность опорной и полетной фаз, а также темп бега. Кинематические признаки характеризовали движение звеньев тела, суставные углы, положение корпуса и межзвенную координацию. Динамические признаки описывали силовые особенности движения и параметры реакции опоры [5–9]. Интеграция этих групп показателей позволила сформировать целостный биомеханический профиль ребенка (рисунок 3), отражающий как внешнюю структуру движения, так и механизмы его реализации.

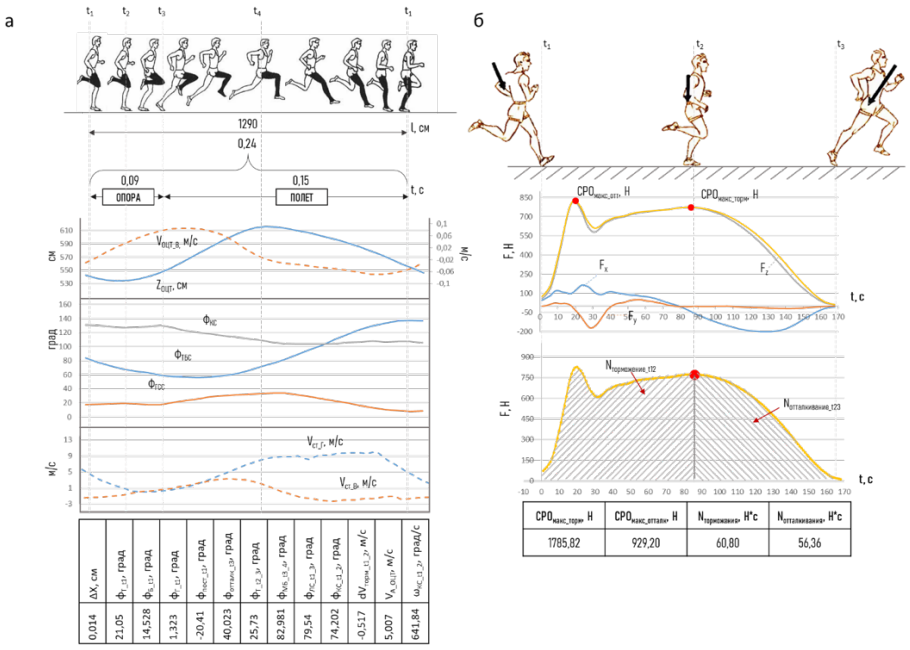


Рисунок 3. Пример визуализации индивидуального профиля двигательной одаренности по: а) кинематическим параметрам б) динамическим параметрам

В результате биомеханических исследований были сформированы две репрезентативные выборки: кинематическая (371 запись, 26 параметров) и комплексная (122 записи, 30 биомеханических параметров), объединенные с экспертными данными, полученными на основе 4 экспертов, для последующего прогнозирования уровня двигательной одаренности детей [11].

Отбор информативных признаков и построение прогностической модели. Для прогнозирования двигательной одаренности использовался метод машинного обучения Random Forest Classifier. Алгоритм основан на ансамбле решающих деревьев, обучаемых на случайных подвыборках данных и признаков [10]. Такой подход обеспечивает устойчивость классификации, снижает риск переобучения и позволяет выявить наиболее значимые биомеханические признаки. После формирования признакового пространства был выполнен отбор наиболее информативных признаков. Алгоритм, основанный на коэффициенте разделения, позволил выделить 14 наиболее значимых параметров из 30 исходных (таблица 1). Наибольший вклад в классификацию внесли параметры отталкивания (Кр = 3,93), временные характеристики фазы постановки ноги (Кр = 3,30) и динамические показатели взаимодействия с опорой [11].

Таблица 1. Коэффициенты разделения биомеханических параметров бега

Параметр	**Обозначения**	**Коэффициент разделения, $К_Р$**
Импульсом силы в фазе отталкивания	N_отталкивания	3.929391
Максимальное значение силы реакции опоры в фазе отталкивания	CPO_макс_отталк	3.652024
Время полета	Δt_полета	3.304497
Импульс силы в фазе торможения	N_торможения	3.267094
Максимальное значение силы реакции опоры в фазе торможения	CPO_макс_торм	2.408135
Угол постановки ноги на опору в момент постановки ноги на опору	φ_пост_t1	1.546394
Угол наклона голени опорной ноги (относительно вертикали) в момент постановки ноги на опору	φ_Г _t1	1.375293
Длина шага	l_A	1.262468

Продолжение таблицы 1

Параметр	**Обозначения**	**Коэффициент разделения, K_P**
Вертикальная посадочная скорость стопы опорной ноги в момент постановки ноги на опору	V_ст_В_t1	1.149943
Боковое смещение общего центра тяжести	ΔX	0.937635
Вертикальная скорость общего центра тяжести в момент отрыва ноги от опоры	V_ОЦТ_В_t3	0.911222
Угол наклона туловища (относительно вертикали) в момент постановки ноги на опору	φ_Т_t1	0.902614
Угол тазобедренного сустава в момент постановки ноги на опору	φ_ТБС_t1	0.872191
Угол голеностопного сустава в момент постановки ноги на опору	φ_ГСС_t1	0.805097

Итоговые показатели качества классификации представлены в таблице 2. Доля правильных классификаций на тестовой выборке составила 96%, что подтверждает высокую прогностическую способность предложенного подхода [11].

Таблица 2. Показатели качества модели Random Forest Classifier

Модель	Accuracy	Precision (1 класс)	Recall (1 класс)	F1 (1 класс)	Precision (2 класс)	Recall (2 класс)	F1 (2 класс)
Random Forest Classifier	0,96	1,00	0,92	0,96	0,94	1,00	0,97

Интерпретация вклада биомеханических признаков. Интерпретируемый анализ показал, что наибольшее значение для прогнозирования двигательной одаренности имеют признаки, отражающие ритмо-темповую структуру бега, межзвенную координацию и параметры опорного взаимодействия. Это свидетельствует о том, что двигательная одаренность проявляется не в отдельных изолированных характеристиках, а в целостной организации бегового движения [5, 7, 11]. Пространственно-временные показатели характеризуют упорядоченность и стабильность локомоции. Кинематические признаки отражают качество координационной организации движения и рациональность работы звеньев тела. Динамические параметры дополняют оценку за счет анализа силовой реализации шага. Практическая ценность интерпретации состоит в том, что специалист получает не только итоговый прогноз, но и объяснение факторов, лежащих в его основе, что делает модель пригодной для систем поддержки принятия решений при раннем спортивном отборе [10–11].

Заключение. В работе предложен подход к прогнозированию двигательной одаренности детей на основе мультимодальных биомеханических данных. Показано, что объединение пространственно-временных, кинематических и динамических характеристик бега позволяет сформировать информативное признаковое пространство для применения методов машинного обучения. Отбор признаков позволил выделить 14 наиболее значимых биомеханических показателей, а модель Random Forest Classifier продемонстрировала accuracy 0,96 на тестовой выборке. Полученные результаты подтверждают возможность использования объективных параметров бега в системах раннего спортивного отбора и педагогической диагностики [11].

Список литературы:

[1] Williams A.M., Reilly T. Talent identification and development in soccer. Journal of Sports Sciences. 2000;18(9):657–667.

[2] Gallahue D.L., Ozmun J.C. Understanding Motor Development: Infants, Children, Adolescents, Adults. 6th ed. New York: McGraw-Hill; 2006.

[3] Губа В.П. Теория и методика современных спортивных исследований. Москва: Спорт; 2016.

[4] Leite N., Lorenzo Calvo A., Cumming S., Gonçalves B., Calleja-Gonzalez J. Editorial: Talent Identification and Development in Sports Performance. Frontiers in Psychology. 2021;12.

[5] Novacheck T.F. The biomechanics of running. Gait & Posture. 1998;7:77–95.

[6] Oeveren B.T., et al. The biomechanics of running and running styles: a synthesis. Sports Biomechanics. 2024;23(4):516–554.

[7] Xiang L., et al. Rethinking running biomechanics: a critical review of ground reaction forces, tibial bone loading, and the role of wearable sensors. Frontiers in Bioengineering and Biotechnology. 2024;12.

[8] Nakano N., et al. Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. Frontiers in Sports and Active Living. 2020;2:50.

[9] Bertec Force Plates [Electronic resource]. Available at: https://www.bertec.com/products/force-plates.

[10] Breiman L. Random forests. Machine Learning. 2001;45(1):5–32.

[11] Григорьева Г.Г. Метод и техническая система биомеханической оценки двигательной одаренности детей на основе анализа беговых локомоций: дис. ... канд. техн. наук. Минск; 2025.

Авторский вклад

Григорьева Гарсия Гаврильевна – постановка задачи, сбор и обработка экспериментальных данных, подготовка текста статьи.

Давыдова Надежда Сергеевна – разработка методологии исследования, научное консультирование, редактирование текста.

Васюк Валерий Евстафьевич – концептуализация исследования, интерпретация результатов, итоговое научное редактирование.

Давыдов Максим Викторович – разработка прогностической модели, реализация алгоритмов машинного обучения, анализ результатов исследования.

Татарко Кристина Игоревна - обзор литературы и аргументация выбора методов исследования, форматирование статьи.

ANALYSIS OF BIOMECHANICAL MOVEMENT FEATURES FOR PREDICTING MOTOR GIFTEDNESS IN CHILDREN

G.G. Grigorieva
Senior Lecturer, Department of Pedagogy and Psychology, Churapcha State Institute of Physical Culture and Sports Churapcha, Russian Federation ga_et@inbox.ru

N.S. Davydova
Associate Professor, Department of Infocommunication Technologies, Belarusian State University of Informatics and Radioelectronics Candidate of Technical Sciences, Associate Professor Minsk, Republic of Belarus davydova-ns@bsuir.by

V.E. Vasyuk
Associate Professor, Higher School of Sports Pedagogy, Peter the Great St. Petersburg Polytechnic University Candidate of Pedagogical Sciences St. Petersburg, Russian Federation

M.V. Davydov
Associate Professor, Candidate of Technical Sciences, First Vice-Rector, Belarusian State University of Informatics and Radioelectronics Minsk, Republic of Belarus

K.I. Tatarko
Associate Professor, Department of Social and Organizational Psychology, Belarusian State University Candidate of Psychological Sciences, Associate Professor Minsk, Republic of Belarus

Abstract. The paper addresses the problem of predicting motor giftedness in children using multimodal biomechanical data obtained from running locomotion analysis. Kinematic and dynamic running parameters recorded by a markerless motion capture system and a force platform were used in the study. Based on a sample of 26 middle-school children, a comprehensive dataset including 30 biomechanical parameters and expert assessments was formed. To select informative indicators, a parametric running model was developed, which made it possible to structure the spatiotemporal, kinematic, and dynamic characteristics of movement. Using the separation coefficient, 14 most significant features were identified, after which the level of motor giftedness was predicted using the Random Forest Classifier method. The best classification results reached an accuracy of 0.96. The proposed approach makes it possible not only to predict the level of motor giftedness, but also to identify the biomechanical parameters that are most significant for sports selection and pedagogical diagnostics.

Keywords: motor giftedness, running locomotion, multimodal data, biomechanical analysis, machine learning, parametric model, feature selection.

УДК 004.8+004.048

МОДЕЛИ И МЕТОДЫ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ ДЕТЕКТИРОВАНИЯ ОБЪЕКТОВ В ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ ОБРАБОТКИ ВИЗУАЛЬНЫХ ДАННЫХ



А.А. Навроцкий
Доцент кафедры информационных технологий автоматизированных систем БГУИР, кандидат физико-математических наук, доцент
dekfitu@bsuir.by



Л.Ю. Шилин
Профессор кафедры информационных технологий автоматизированных систем БГУИР, кандидат технических наук, доктор технических наук
shilin@bsuir.by



Д.Д. Хмыз
Ассистент кафедры вычислительных методов и программирования БГУИР, магистрант кафедры ИТАС
d.khmyz@bsuir.by



К.Д. Сморщок
Инженер-программист, ООО «Эфкон диджитал»
s.kirill150282@mail.ru

А.А. Навроцкий
Окончил Минский радиотехнический институт. Область научных интересов связана с разработкой систем технического зрения и технологиями программирования.

Л.Ю. Шилин
Окончил Минский радиотехнический институт, специальность «Автоматика и телемеханика». Область научных интересов связана с анализом и синтезом цифровых систем автоматизированного управления.

Д.Д. Хмыз
Окончила Белорусский национальный технический университет. Область научных интересов связана с технологиями программирования и разработкой систем технического зрения.

К.Д. Сморщок
Область научных интересов связана с технологиями программирования.

Аннотация. В работе представлено исследование и практическая реализация нейросетевой системы детектирования объектов транспортной инфраструктуры на основе современной архитектуры Ultralytics YOLO26. Рассмотрены методы глубокого обучения для обработки визуальных данных в задачах реального времени, особенности оптимизации параметров сети с использованием модифицированного алгоритма MuSGD, а также комплексная оценка качества модели по метрикам IoU, Precision, Recall, F1 и mAP. Полученные результаты (mAP@0.5 = 0.768, Precision = 0.93, Recall = 0.84) подтверждают эффективность предложенного подхода для применения в интеллектуальных информационных системах мониторинга.

Ключевые слова: глубокое обучение, детектирование объектов, обработка визуальных данных, YOLO, интеллектуальные информационные системы, оптимизация.

Введение. Анализ визуальной информации с помощью методов глубокого обучения становится ключевым направлением в развитии интеллектуальных систем. Высокая скорость и точность распознавания объектов в реальном времени открывают широкие возможности для автоматизации мониторинга, управления и поддержки принятия решений.

При этом важным фактором является оптимизация нейросетевых архитектур с учётом ограничений вычислительных ресурсов и требований к минимальной задержке обработки.

Основная часть. Современные алгоритмы детектирования объектов активно развиваются, выходя за рамки традиционных схем и позволяя достигать новых уровней производительности и стабильности. Современные информационные технологии в сфере обработки данных активно используют методы глубокого обучения для анализа визуальной информации.

В задачах интеллектуального мониторинга, автоматизированного управления и систем поддержки принятия решений особое значение приобретает корректное и оперативное детектирование объектов на изображениях и видеопотоках. Такие системы функционируют в условиях ограниченных вычислительных ресурсов и требований к реальному времени обработки, что определяет необходимость использования эффективных архитектур нейронных сетей.

Эволюция методов object detection прошла путь от двухступенчатых детекторов (Faster R-CNN) [1] к одношаговым архитектурам семейства YOLO [2], обеспечивающим баланс между точностью и скоростью обработки.

Современные end-to-end модели, включая трансформерные подходы (DETR) [3], минимизируют необходимость постобработки и этапа подавления немаксимумов (NMS), что снижает вычислительную нагрузку и повышает стабильность инференса.

В настоящей работе применена архитектура Ultralytics YOLO26 [4], относящаяся к одношаговым детекторам нового поколения с головой типа one-to-one.

Исключение NMS позволяет обеспечить предсказуемое время обработки кадра и уменьшить латентность, что критично для информационных систем реального времени.

Обучение модели выполнялось с использованием стратегии трансферного обучения на основе предобученных весов датасета COCO [5].

Такой подход позволяет использовать универсальные визуальные признаки низкого и среднего уровня (контуры, текстуры, геометрические структуры), адаптируя их к специализированной предметной области.

Оптимизация параметров сети осуществлялась с применением гибридного алгоритма MuSGD. Классический стохастический градиентный спуск реализуется по формуле:

$$wt+1=wt-\eta L(wt), \quad (1)$$

где –скорость обучения,

L(wt)–градиент функции потерь.

Модификация MuSGD включает нормализацию градиента с учётом оценки его дисперсии:

$$vt=\mu vt-1+L(wt)st+\varepsilon. \quad (2)$$

Такой подход стабилизирует процесс обучения, уменьшает осцилляции и ускоряет достижение минимума функции потерь, что особенно важно при высокой размерности параметрического пространства.

Для оценки качества модели использовался комплекс метрик, традиционно применяемых в задачах детектирования объектов.

Коэффициент перекрытия IoU (Intersection over Union) определяется как отношение площади пересечения предсказанной и истинной рамок к площади их объединения:

$$IoU=AreaintersectionAreaunion. \quad (3)$$

Точность (Precision) рассчитывается как:

$$\text{Precision}=\frac{TP}{TP+FP}, \tag{4}$$

а полнота (Recall):

$$\text{Recall}=\frac{TP}{TP+FN}. \tag{5}$$

Гармоническое среднее этих показателей определяется через F1-метрику:

$$F1=\frac{2PR}{P+R}. \tag{6}$$

Интегральной характеристикой выступает средняя точность (mAP), вычисляемая как среднее значение Average Precision по всем классам.

В результате обучения получены следующие показатели качества:

mAP@0.5 = 0.768, Precision = 0.93, Recall = 0.84, F1 = 0.62.

Высокое значение Precision свидетельствует о низкой вероятности ложных срабатываний, а Recall = 0.84 подтверждает способность модели обнаруживать большинство целевых объектов. Значение mAP@0.5 характеризует устойчивость модели при варьировании порога уверенности.

Кривая Precision–Recall для модели YOLO26 представлена на рисунке 1.

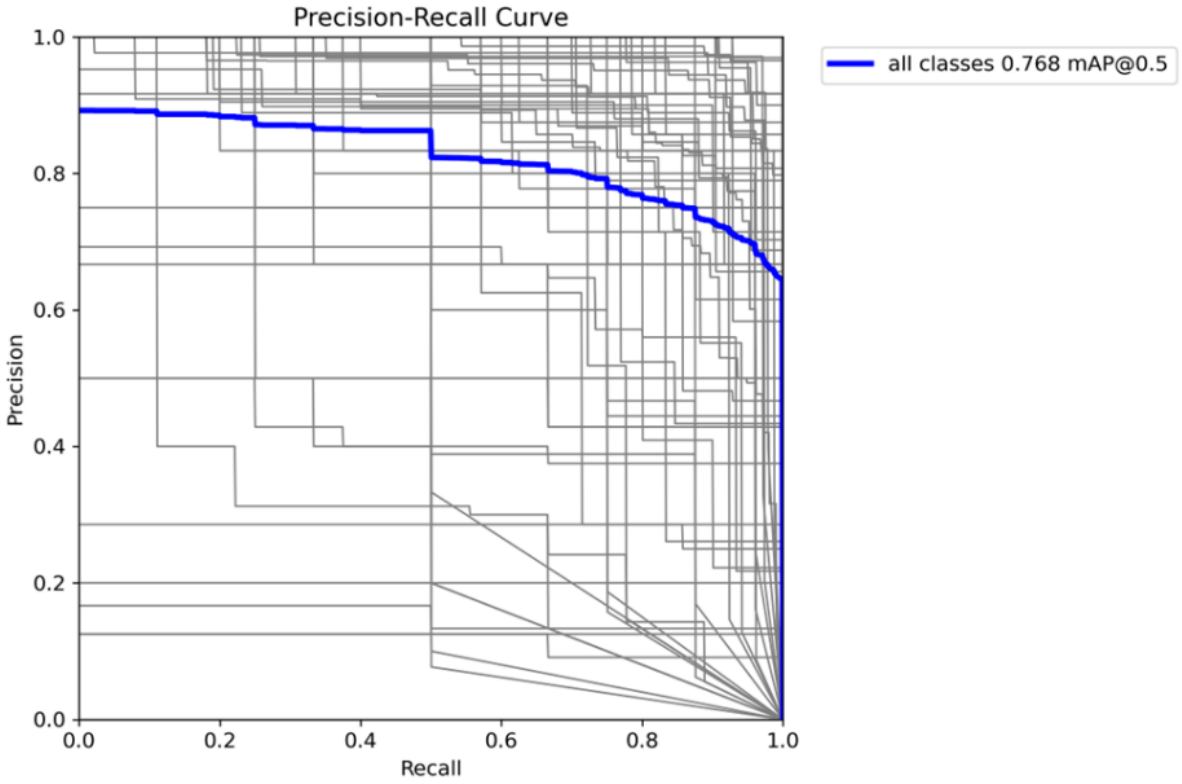


Рисунок 1 Кривая Precision–Recall для модели YOLO26

Особое внимание уделялось анализу дисбаланса классов (long-tail распределения), характерного для реальных датасетов [6]. Наличие редких категорий объектов оказывает влияние на снижение F1-метрики, однако интегральные показатели mAP демонстрируют достаточную обобщающую способность модели.

Заключение. Проведённое исследование демонстрирует эффективность современных методов глубокого обучения для задач автоматизированного детектирования объектов в интеллектуальных информационных системах.

Использование архитектуры YOLO26 и оптимизатора MuSGD позволяет обеспечить баланс между точностью и вычислительной эффективностью, что делает предложенный подход перспективным для внедрения в системы реального времени.

Список литературы

[1] Ren S., He K., Girshick R., Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks // IEEE TPAMI. – 2017.

[2] Redmon J., Farhadi A. YOLOv3: An Incremental Improvement // arXiv:1804.02767. – 2018.

[3] Carion N. et al. End-to-End Object Detection with Transformers // ECCV. – 2020.

[4] Ultralytics. Ultralytics YOLO Documentation. – 2024.

[5] Lin T.-Y. et al. Microsoft COCO: Common Objects in Context // ECCV. – 2014.

[6] Zhang Y. et al. Deep Long-Tail Learning: A Survey // IEEE TPAMI. – 2023.

Авторский вклад

Анатолий Александрович Навроцкий – постановка задачи исследования, формулирование основных алгоритмов, разработка структуры статьи.

Леонид Юрьевич Шилин – постановка задачи исследования и формулировка гипотез.

Дарья Дмитриевна Хмыз – разработка алгоритмов работы системы, сбор и систематизация фактических данных, написание статьи.

Кирилл Дмитриевич Сморщок – обучение интеллектуальной системы обработки.

DEEP LEARNING MODELS AND METHODS FOR OBJECT DETECTION IN INTELLIGENT VISUAL DATA PROCESSING SYSTEMS

A.A. Navrotsky
Associate Professor, Department of Information Technologies for Automated Systems, BSUIR, PhD in Physics and Mathematics, Associate Professor

L.Yu. Shilin
Professor, Department of Information Technologies and Automated Systems, BSUI), Candidate of Technical Sciences, Doctor of Technical Sciences

D.D. Khmyz
Assistant Professor, Department of Computational Methods and Programming, BSUIR, Master's Student, Department of ITAS

K.D. Smorshchok
Student, group 320602, BSUIR

Abstract. The paper presents a study and practical implementation of a neural network–based object detection system for transport infrastructure using the modern Ultralytics YOLO26 architecture. Methods of deep learning for visual data processing in real-time applications are considered, along with network parameter optimization using the modified MuSGD algorithm. A comprehensive evaluation of model performance is conducted using IoU, Precision, Recall, F1-score, and mAP metrics. The obtained results (mAP@0.5 = 0.768, Precision = 0.93, Recall = 0.84) confirm the effectiveness of the proposed approach for deployment in intelligent information monitoring systems.

Keywords: deep learning, object detection, visual data processing, YOLO, intelligent information systems, optimization.

УДК 004.932

ФОРМИРОВАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ДЛЯ НЕЙРОННЫХ СЕТЕЙ ПРИ ПРОСТРАНСТВЕННО-ВРЕМЕННОМ ОБНАРУЖЕНИИ ДЫМА НА ВИДЕО





Е.Р. Адамовский
Старший преподаватель кафедры вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой
e.adamovsky@psu.by

Р.П. Богуш
Заведующий кафедрой вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой, доктор технических наук, профессор
r.bogush@psu.by

Е.Р. Адамовский
Окончил Полоцкий государственный университет имени Евфросинии Полоцкой. Область научных интересов :компьютерное зрение, машинное обучение, обработка изображений и видео.

Р.П. Богуш
Область научных интересов: обработка изображений и видео, интеллектуальные системы, машинное обучение.

Аннотация. В работе представлена технология формирования набора данных для обучения и тестирования моделей искусственных нейронных сетей, выполняющих пространственно-временной анализ кадров при детектировании дыма на видеопоследовательности. Предложена модификация формата представления последовательностей изображений и метаданных для данной задачи, обеспечивающая совместимость с форматом COCO. Отличием предложенного подхода является также формирование синтетической выборки путем совмещения системы частиц и фона с автоматизированным аннотированием объектов.

Ключевые слова: детектирование дыма, компьютерное зрение, синтезированные изображения, машинное обучение.

Введение. Развитие искусственного интеллекта и увеличение производительности вычислительных устройств позволяет на практике применять искусственные нейронные сети (ИНС) для задач, которые ранее решались другими способами менее эффективно. Так, с целью детектирования пожара на ранней стадии и, в частности, дыма, являющегося одним из его признаков, используются физические датчики, которые обладают рядом недостатков и ограниченной областью применения, например, они неэффективны на открытых пространствах. При наличии высокопроизводительного аппаратного обеспечения, анализ видеозаписей может выполняться с помощью моделей ИНС. Их преимущество заключается в способности автоматически выделять из анализируемых данных наиболее релевантные признаки, необходимые для эффективного детектирования дыма в различных условиях съемки.

Известно, что важным фактором при обучении моделей ИНС является объем и разнообразие обучающей выборки. В таблице 1 приведены датасеты, которые содержат видеозаписи или последовательности кадров с областями дыма и их характеристики: количество видеозаписей N, их разрешение, средняя длительность $t_{ср}$, кадровая частота

(frames per second, FPS), а также содержание. Указаны способы аннотирования: классификация, локализация (ограничивающие рамки объектов); frame-level (аннотации на уровне отдельных кадров), video-level (аннотации на уровне видеозаписей).

Таблица 1. Наборы данных с областями дыма для обучения моделей ИНС

Датасет	***N***	**Разрешение, пк**	$t_{ср}$, **с**	**FPS**	**Аннотирование**	**Содержание**
FiSmo (FireVid) [1]	27	от 320×240 до 600×336	120	от 15 до 30	frame-level классификация	записи пожаров
RESCUER [1]	61	от 320×240 до Full HD	20	от 15 до 30		учения пожарных
VisiFire [2]	33	от 320×240 до 720×576	90	от 7 до 30	-	записи пожаров
D-Fire [3]	100	от HD до FullHD	30	2	-	городские пейзажи, записи пожаров
Mivia Smoke [4]	149	292×240	800	от 8 до 17	video-level классификация	дым, облака, солнце
FireSense [5]	49	от 320×240 до 640×360	35	от 10 до 30	video-level классификация	записи пожаров
FURG [6]	23	до HD	40	30	frame-level локализация	записи пожаров
HPWREN [7]	470	до 4K	-	-	frame-level классификация	кадры системы отслеживания пожаров
RISE [8]	12 567	180×180	360	~0,1	video-level классификация	записи промышленных выбросов
KMU FIRE & SMOKE [9]	38	320×240	115	24	video-level классификация	записи пожаров
M4SFWD [10]	28	от 776×452 до 1480×684	95	30	frame-level локализация	синтезированный дым, лесные пожары

Анализ таблицы 1 свидетельствует, что значительная часть датасетов не содержит покадровых аннотаций. За исключением HPWREN, разрешение видеозаписей не превышает 1920 пк по ширине, а в большинстве случаев составляет меньшие значения, порядка 320×240 пк и ниже, до 180×180 пк в RISE. Почти все датасеты основаны на реальных видеозаписях, но также представлен M4SFWD с синтезированными изображениями дыма и фона. В FiSmo и FURG дым объединяют с пламенем в общий класс «пожар» (fire). Форматы аннотирования в датасетах различны, и требуют приведения к единой форме с целью возможности их совместного использования.

Для рассмотренных датасетов произведена оценка разнообразия сцен, при этом отдельной сценой считалась уникальная комбинация места, времени и условий съемки, которая визуально отличалась от других видеозаписей в датасете. Анализ показал, что крупные наборы характеризуются низким разнообразием, например, в Mivia Smoke видеозаписи сняты в одном месте, в RISE представлено всего три локации, а HPWREN содержит сцены природных пожаров в одном регионе с однообразным ландшафтом. В наборах с менее чем 100 видеозаписями от 50% до 80% сцен являются уникальными, при этом выборка может быть ограничена собственной тематикой, например, в RESCUER представлены только видеозаписи тушения промышленных объектов. Суммарное

количество кадров на датасет, с учетом длительности видеозаписей и их кадровой частоты, составляет десятки и сотни тысяч, но малое количество уникальных сцен в датасетах нивелирует большой объем данных, поскольку важную роль при обучении моделей ИНС имеет разнообразие обучающих примеров. Таким образом, сложность задачи формирования обучающей выборки с объектами дыма обусловлена рядом причин.

1. Дефицит видеозаписей и наборов данных. Известно, что для достижения высокой обобщающей способности моделей ИНС, тренировочных образцов должно быть больше, чем обучаемых параметров, в современных архитектурах составляющих десятки миллионов. Количество и качество доступных в открытых источниках видеозаписей, содержащих области дыма, не обеспечивают необходимый объем обучающей выборки. Большинство видеозаписей имеют низкое разрешение и получены в условиях отсутствия жесткой фиксации камеры, что затрудняет точное аннотирование областей интереса. Пожар часто фиксируется уже на поздних стадиях, когда дым занимает большую площадь кадра. В открытом доступе обнаружено ограниченное количество датасетов, которые характеризуются небольшим объемом и низким уровнем уникальности сцен, а также имеют различные форматы представления. На практике высокие требования к объему обучающей выборки могут быть снижены за счет аугментирования данных и трансферного обучения. Так, в [11] показано, что предобученные модели с 30-40 млн параметров успешно обучаются на датасете с высоким разнообразием (2731 уникальная сцена) из 12 586 кадров с пламенем и дымом. В [12] модель на основе Deformable DETR (37 млн параметров) была обучена на датасете из 10 250 изображений с дымом, и достигла точности 49,7% mAP. В [13] модель на основе RT-DETR (36 млн параметров) была обучена на датасете из 12 564 изображений, и достигла точности 96,1% AP_{50}. Как правило, для извлечения первичных признаков применяются сверточные нейронные сети, предобученные на датасетах ImageNet: ImageNet-21k (14 млн изображений, 21 000 классов) или ImageNet-1k (1,3 млн изображений, 1 000 классов). Для моделей, выполняющих пространственно-временной анализ, часто используется датасет Kinetics-400 (246 тысяч видеозаписей, 400 классов). Однако следует учитывать, что, при разработке новых архитектур, использование предобученных моделей невозможно или в значительной степени ограничено.

2. Визуальное представление. Дым является полупрозрачным объектом без четкой формы и границ, что затрудняет его локализацию. Движение дыма является ключевым признаком, позволяющим отличать его от объектов, обладающих схожим визуальным представлением, но другими динамическими свойствами: туман, облака или пыль. Как следствие, обучающая выборка должна предоставлять возможность для отслеживания динамики областей дыма во времени. Решением является объединение кадров в связанные последовательности (серии), где последний кадр является опорным, аннотирован информацией о координатах ограничивающих рамок и содержит ссылки на предыдущие кадры серии в порядке их следования. Таким образом, формирование обучающей выборки для моделей ИНС, решающих задачу детектирования дыма на видеопоследовательности на основе пространственных и временных его особенностей, требует компиляции большого количества уникальных записей и аннотирования исходных данных в едином формате, который позволяет хранить информацию о сериях кадров.

Разработка формата представления данных. Анализ показывает, что для аннотирования объектов на видеозаписях не существует единого общепринятого формата, а имеющиеся реализации специализированы, например, только для трекинга. В исследованиях применяются модификации с учетом специфики решаемых задач на основе распространенных форматов для аннотирования отдельных изображений: COCO (формат JSON), Pascal VOC (формат XML), YOLO (формат TXT), которые не эффективны для видеопоследовательностей. Поэтому предлагается модификация формата COCO, как показано на рисунке 1.

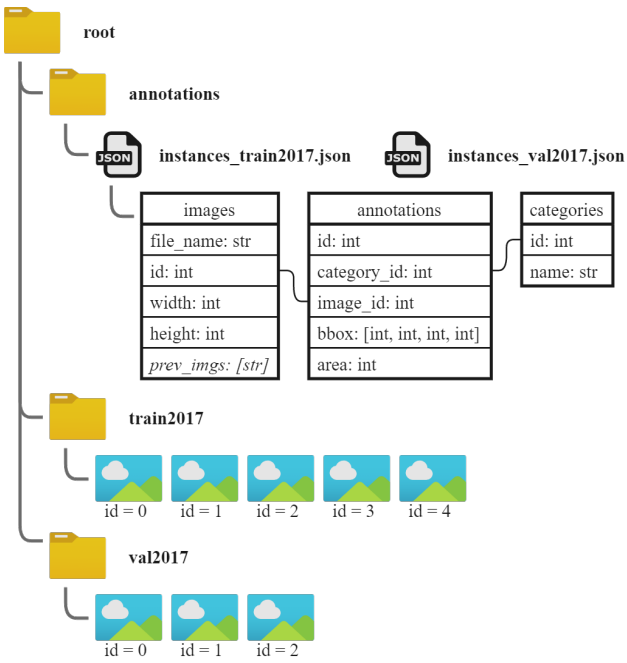


Рисунок 1. Структура формата представления кадров и метаданных в обучающем наборе

Формат предполагает наличие директории «annotations» для хранения двух JSON-файлов с аннотациями для тренировочной и валидационной выборки, «instances_train2017.json» и «instances_val2017.json» соответственно. Файлы аннотаций содержат разделы: «images» для перечисления кадров в соответствующей директории «train2017» или «val2017», «annotations» для аннотирования кадров, «categories» для определения списка детектируемых классов. В раздел «images» включено поле «prev_imgs», которое содержит относительные пути к предыдущим кадрам серии в порядке убывания их порядкового номера в серии. Предложенный подход допускает перекрытие между сериями кадров и различный размер серии T, включая вариант $T<T_m$, где T_m – предполагаемый фиксированный размер серии, который может быть обусловлен архитектурой конкретной модели ИНС. Предлагаемая стратегия заключается в циклическом заполнении серии до размера T_m. При $T = 1$ выполняется удаление всех аннотаций текущей серии, что обусловлено отсутствием в ней временных признаков объектов дыма. Современные программные реализации моделей ИНС содержат модули загрузки данных, которые процедурно, в рамках конвейера обработки, добавляют к изображениям аугментации, что позволяет снизить занимаемый датасетом объем памяти на носителе информации. В качестве примера может быть рассмотрен модуль загрузки COCO в реализации архитектуры DETR [14], который требует лишь небольшой модификации для возможности обработки серий кадров. Следовательно, предварительная аугментация датасета не требуется, однако может включать отражение, изменение яркости и контраста, масштабирование, поворот, наложение шума и размытие.

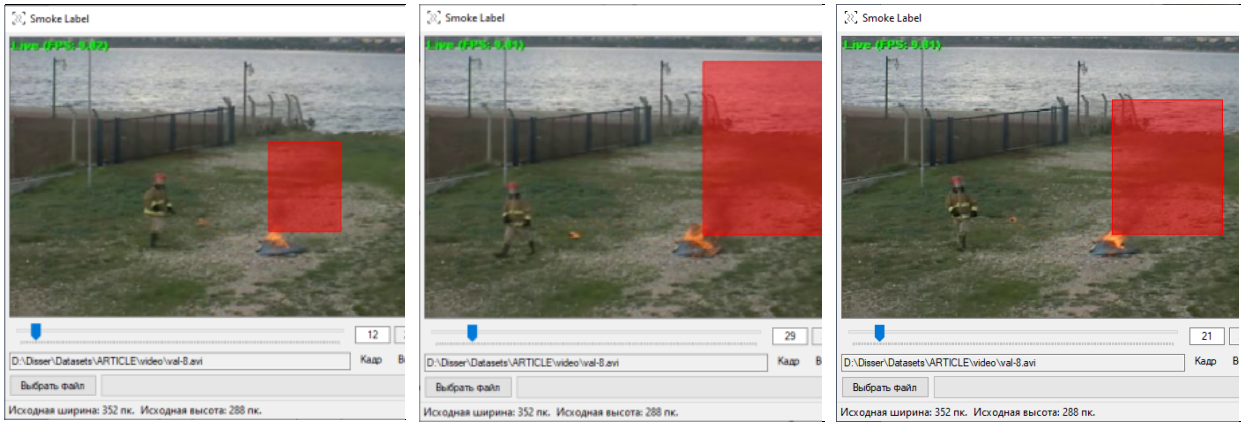
Генерация обучающей выборки. Покадровое аннотирование видеозаписей оператором имеет высокую трудоемкость, поэтому таким способом может быть подготовлен лишь ограниченный объем данных. В ряде исследований предлагается синтезировать сцены с объектами дыма при помощи инструментов трехмерного моделирования и визуализации: Unity и Unreal Engine [10], NVIDIA Omniverse [15], Blender [16]. Окружение в моделируемых сценах предлагается формировать на основе трехмерных объектов [10], реальных фотографий [17] и кадров из видеоигр с фотореалистичной графикой [16]. Совмещение дыма и фона может производиться сразу, при генерации синтезированного дыма, либо отложено, используя альфа-канал полученных изображений или технологию хромакей (green screen) [18, 19].

Генерация синтезированных данных позволяет сформировать разнообразную выборку и автоматизировать процесс аннотирования. Однако эффективность подхода ограничена визуальными отличиями синтезированного дыма от настоящего, который является сложной дисперсной системой [17], т.е. актуальна проблема смещения распределения данных (domain shift). Следовательно, доля синтезированных кадров в обучающей выборке должна быть ограничена. Предлагаемые соотношения реальных и синтезированных кадров варьируются от 3:1 до 1:1 [15, 20].Таким образом, решением проблемы дефицита тренировочных выборок является подготовка гибридного датасета, который содержит видеозаписи с реальным дымом и синтезированные образцы, а часть кадров не содержит объектов дыма. Негативную выборку следует разделить: сцены произвольного содержимого и сцены с объектами, похожими на дым, включая облака, туман, пар и смог. Генерация синтезированных данных выполнена с помощью игрового движка Unity Engine 5 и инструмента системы частиц (Particle System), параметры которого подбирались для каждой серии кадров так, чтобы дым выглядел правдоподобно, имел небольшой размер и плотность, имитируя начальную стадию возгорания. Управляющий скрипт изменял следующие параметры: форму эмиттера дыма, его плотность, скорость и направление частиц, влияние ветра, яркость и направление освещения. Для каждой серии фиксировалось определенное количество кадров, в которых отражалась динамика развития пожара от момента появления объекта дыма. Предполагается, что модели ИНС, реализующие пространственно-временной анализ, при обучении на таких данных будут вынуждены учитывать временной контекст, необходимый для разрешения неопределенности, вызванной недостаточной информативностью пространственных признаков. Автоматизированная разметка в Unity Engine выполнялась путем проекции координат частиц с непрозрачностью выше порогового значения в экранное пространство и вычислением ограничивающей рамки на основе минимальных и максимальных значений координат по обеим осям. Фоновые изображения для каждой серии загружались из подготовленного набора данных, собранного на основе изображений из сети Интернет с произвольным содержимым: пейзажи, архитектурные объекты, интерьеры. На рисунке 2 показаны примеры с синтезированными объектами дыма, изображения кадрированы для лучшего восприятия.



Рисунок 2. Автоматически аннотированные кадры с синтезированным дымом на основе системы частиц в Unity Engine

Для аннотирования видеозаписей оператором разработано программное средство в виде десктопного приложения «Smoke Label» для операционной системы Windows. Приложение предоставляет функционал разметки произвольного количества объектов класса «дым» (smoke) на видеозаписях, используя ключевые кадры и графический интерфейс для создания и редактирования ограничивающих рамок. Отличие от аналогичных open-source инструментов, включая CVAT, Label Studio и LabelMe, заключается в том, что «Smoke Label» позволяет формировать выборки в предложенном формате. Создание обучающей выборки происходило следующим образом. Видеозаписи загружались в «Smoke Label» и автоматически разбивались на отдельные видеокадры. На выбранных кадрах создавались опорные рамки вокруг объектов дыма (рисунок 3а, б), а затем производилось автоматическое вычисление рамок в промежуточных кадрах путем линейной интерполяции координат и размеров опорных рамок (рисунок 3в).



а) б) в)

Рисунок 3. Аннотирование объекта дыма в приложении «Smoke Label»: а) построение рамки для кадра №12; б) построение рамки для кадра №29; в) интерполированная рамка в кадре №21

На рисунке 4 показаны примеры аннотированных с помощью «Smoke Label» кадров из датасета с реальными объектами дыма, которые отмечены красными прямоугольниками.



Рисунок 4. Аннотированные кадры с реальными объектами дыма

На рисунке 5 показаны примеры кадров из датасета, которые содержат объекты, по отдельным визуальным признакам обладающие сходством с дымом.

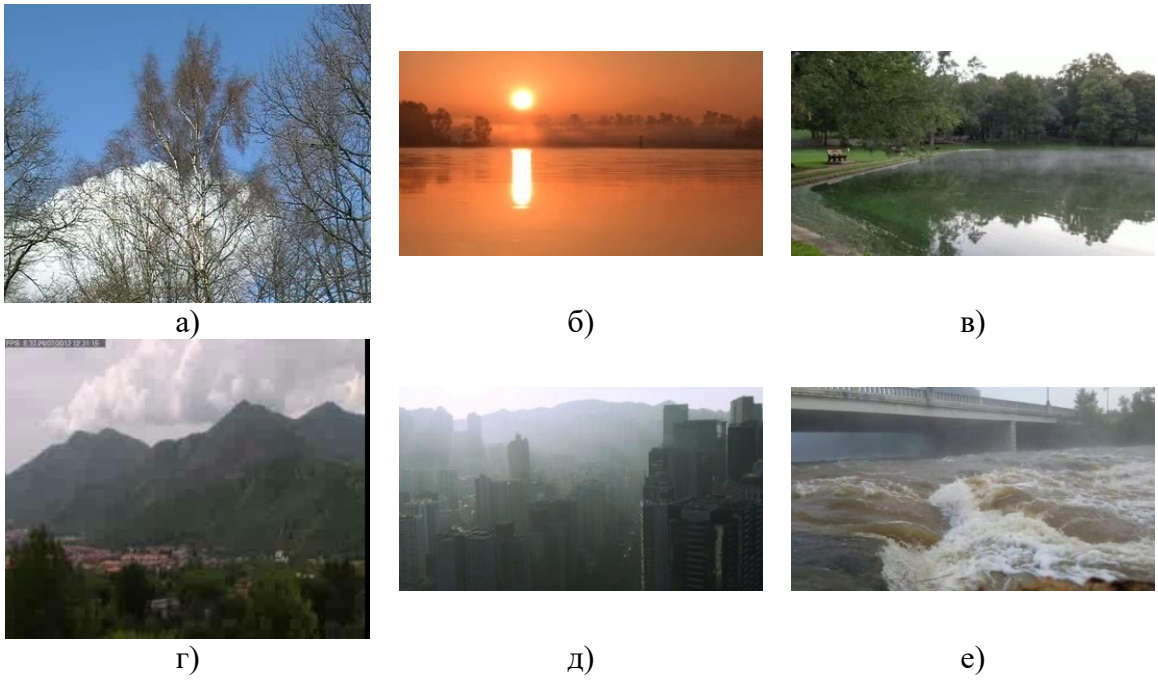


Рисунок 5. Примеры кадров с похожими на дым объектами: а) качающиеся на ветру деревья;
б) солнечные блики; в) туман; г) облака; д) смог; е) водный поток и пар

Итоговый набор данных сформирован при помощи разработанных на языке Python наборов скриптов с заданными настройками, включая шаг выборки и максимальное количество кадров в серии, на основе датасетов, полученных ранее, как показано на рисунке 6.

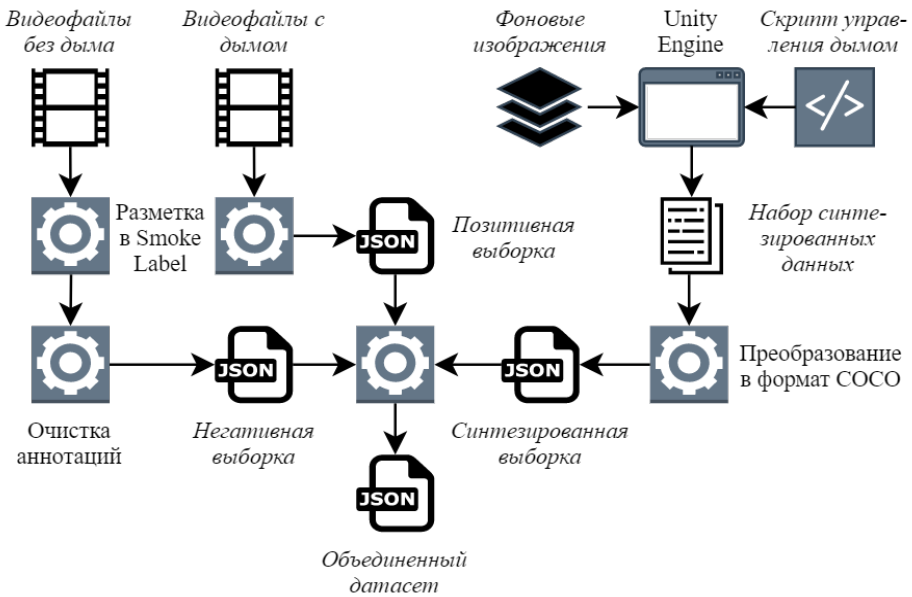


Рисунок 6. Сценарий формирования обучающей выборки

Тренировочная часть датасета (train) состоит из 36 360 серий (4 769 уникальных сцен), валидационная часть (val) состоит из 4 350 серий (475 сцен), всего 40 710 серий (5244 сцены). 9 000 серий (100 сцен), т.е. 22% от объема датасета, сформированы на основе рассмотренных в таблице 1 наборов, остальные выборки получены из собранных вручную

видеозаписей или синтезированы. Ограниченное использование материалов из существующих датасетов обусловлено тем, что для каждого набора видеозаписи были сгруппированы по представленным в них сценам, и для каждой группы формировалась только одна выборка, за счет чего достигалось максимальное разнообразие данных без доминирования отдельных сцен. Формируемые выборки имели сопоставимый друг с другом объем (от 50 до 300 серий) с целью равномерной репрезентации всех возможных условий съемки объектов дыма.

Следует отметить, что в сформированном датасете количество серий эквивалентно числу кадров, поскольку каждый кадр гарантированно является опорным в одной из представленных серий, что продемонстрировано на рисунке 7.

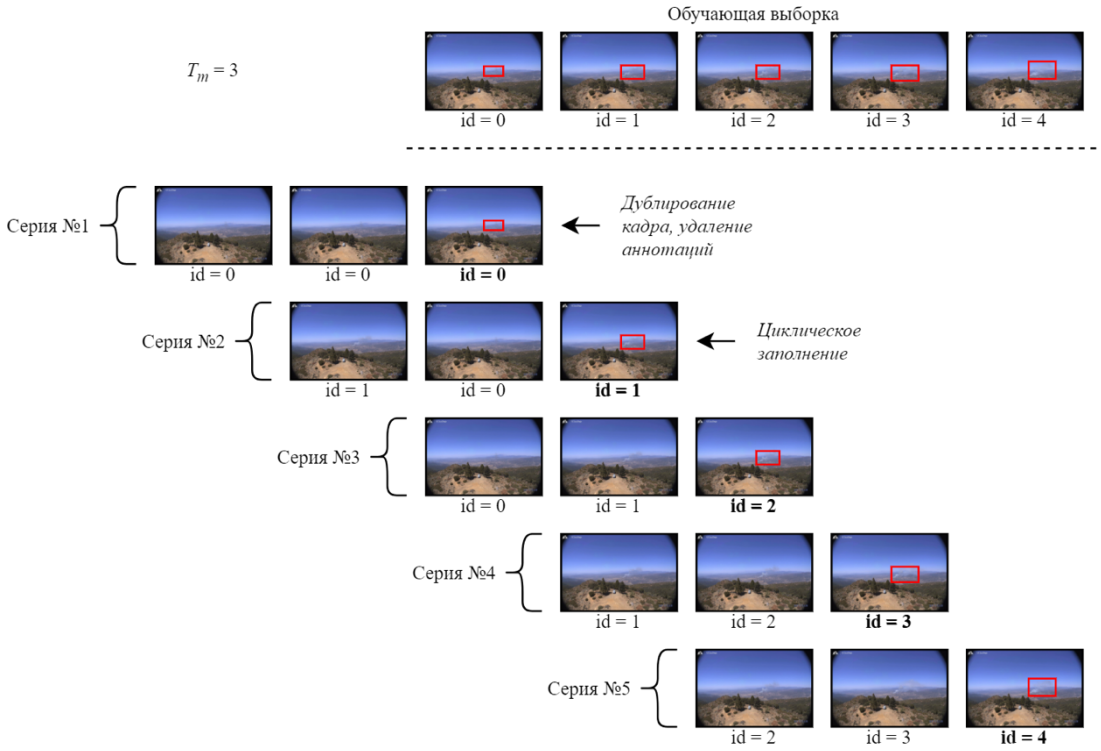


Рисунок 7. Общая схема формирования серий кадров

В тренировочных сериях 83% кадров содержат дым, выполняется соотношение реальных и синтезированных данных 3,3:1. Соотношение позитивных и негативных выборок составляет 3:1. Разделение тренировочной и валидационной выборки выполнено в пропорции 90/10, причем в валидационный набор не включены кадры с синтезированными объектами дыма. Объем занимаемой памяти обучающей выборки составил 5 ГБ.

Состав созданного датасета представлен в таблице 2.

Таблица 2. Состав датасета

№	Train		Val		Описание
	Серии	Сцены	Серии	Сцены	
1	19400	334	3600	96	Серии на основе видеозаписей с реальным дымом
2	5960	35	380	9	Серии на основе видеозаписей с похожими на дым объектами
3	4000	4000	370	370	Серии на основе видеозаписей без дыма и похожих объектов
4	7000	400	0	0	Серии с синтезированными объектами дыма

Заключение. Разработана технология формирования набора данных для обучения и тестирования моделей искусственных нейронных сетей, которые выполняют задачу детектирования объектов дыма на видеопоследовательности. Технология формирования обучающей выборки включает: сбор видеозаписей с дымом и объектами, обладающими схожими визуальными признаками; аннотирование видеозаписей с помощью специального программного средства «Smoke Label»; формирование синтетических объектов дыма на основе системы частиц в игровом движке Unity Engine; объединение полученных выборок. На основе разработанной технологии создан датасет в модифицированном формате COCO, состоящий из 40 710 серий кадров, включая выборки с реальным и синтезированным дымом, который может быть использован для обучения моделей искусственных нейронных сетей, реализующих пространственно-временной анализ видеокадров.

Список литературы

[1] Cazzolato, M. T. FiSmo: A Compilation of Datasets from Emergency Situations for Fire and Smoke Analysis / M. T. Cazzolato, L. P. S. Avalhais, D. Y. T. Chino [et al.] // SBBD 2017 – Proceedings of Satellite Events: Dataset Showcase Workshop (DSW). – São Paulo : SBC, 2017. – URL: https://sbbd.org.br/2017/wp-content/uploads/sites/3/2017/10/proceedings-satellite-events-sbbd-2017.pdf. – Дата доступа: 26.03.2026.

[2] VisiFire [Электронный ресурс] // Bilkent University, Department of Electrical and Electronics Engineering. – URL: http://signal.ee.bilkent.edu.tr/VisiFire/Demo. – Дата доступа: 26.03.2026.

[3] D-Fire Videos [Электронный ресурс]. – Режим доступа: https://clk.li/BhNw. – Дата доступа: 26.03.2026.

[4] Fiore, E. M. et al. A video dataset for smoke detection based on deep learning / E. M. Fiore, G. Fenu, M. Marras, R. Pellicanò // Journal of Visual Communication and Image Representation. – 2020. – Vol. 72. – P. 102857.

[5] chrisfilo. FireSense: Dataset for Fire and Smoke Detection [Электронный ресурс] // Kaggle. – URL: https://www.kaggle.com/datasets/chrisfilo/firesense. – Дата доступа: 26.03.2026.

[6] Hogan, I. Fire Warn: Fire Hazards Detection Using Deep Learning Models / I. Hogan, D. Qiao, R. Luo [et al.] // 2021 IEEE Third International Conference on Cognitive Machine Intelligence (CogMI) : proceedings. – Atlanta, GA, USA : IEEE, 2021. – PP. 1-10. – DOI: 10.1109/CogMI52975.2021.00010.

[7] Yazdi, A. Nemo: An Open-Source Transformer-Supercharged Benchmark for Fine-Grained Wildfire Smoke Detection / A. Yazdi, H. Qin, C. B. Jordan [et al.] // Remote Sensing. – 2022. – № 14(16). – Art. 3979. – DOI: 10.3390/rs14163979.

[8] Hsu, Y. C. Project RISE: Recognizing Industrial Smoke Emissions / Y. C. Hsu, T. H. Huang, T. Y. Hu [et al.] // Proceedings of the AAAI Conference on Artificial Intelligence. – 2021. – № 35(17). – PP. 14813-14821. – DOI: 10.1609/aaai.v35i17.17777.

[9] Gupta, T. Early Wildfire Smoke Detection in Videos / T. Gupta, H. Liu, B. Bhanu // 2020 International Conference on Pattern Recognition (ICPR) : proceedings. – 2021. – PP. 8523-8530. – DOI: 10.1109/ICPR48806.2021.9413231.

[10] Wang, G. Multiple scenarios, multiple weather conditions, multiple lighting conditions and multiple wildfire objects: Synthetic Forest Wildfire Dataset (M4SFWD) [Электронный ресурс] / G. Wang // IEEE Dataport. – 2024. – DOI: 10.21227/m9kz-bw61. – Дата доступа: 26.03.2026.

[11] Han, X. Benchmarking Multi-Scene Fire and Smoke Detection / X. Han, N. Pu, Z. Feng [et al.] // Chinese Conference on Pattern Recognition and Computer Vision (PRCV) : proceedings. – Singapore : Springer Nature Singapore, 2024. – PP. 203-218. – DOI: 10.1007/978-981-97-8795-1_14.

[12] Huang, J. A Small-Target Forest Fire Smoke Detection Model Based on Deformable Transformer for End-to-End Object Detection / J. Huang, J. Zhou, H. Yang [et al.] // Forests. – 2023. – № 14(1). – Art. 162. – DOI: 10.3390/f14010162.

[13] Yang, L. Real-Time Smoke Detection in Surveillance Videos Using an Enhanced RT-DETR Framework with Triplet Attention and HS-FPN / L. Yang, Y. Cheng, F. Xu [et al.] // Fire. – 2024. – № 7(11). – Art. 387. – DOI: 10.3390/fire7110387.

[14] Carion, N. End-to-End Object Detection with Transformers / N. Carion, F. Massa, G. Synnaeve [et al.] // European Conference on Computer Vision (ECCV) : proceedings. – Cham : Springer, 2020. – PP. 213-229. – DOI: 10.1007/978-3-030-58452-8_13.

[15] Seefried, E. Balancing Quality and Quantity: The Impact of Synthetic Data on Smoke Detection Accuracy in Computer Vision / E. Seefried, C. Jung, J. Fitzgerald [et al.] // CVPR 2024 Workshop SynData4CV : proceedings. – 2024. – URL: https://openreview.net/forum?id=ckhhCcGq9n. – Дата доступа: 26.03.2026.

[16] Mao, J. Wildfire Smoke Classification Based on Synthetic Images and Pixel- and Feature-Level Domain Adaptation / J. Mao, C. Zheng, J. Yin [et al.] // Sensors. – 2021. – № 21(23). – Art. 7785. – DOI: 10.3390/s21237785.

[17] Xu, G. Deep Domain Adaptation Based Video Smoke Detection using Synthetic Smoke Images / G. Xu, Y. Zhang, Q. Zhang [et al.] // Fire Safety Journal. – 2017. – Vol. 93. – DOI: 10.1016/j.firesaf.2017.08.004.

[18] Chen, S.-H. Real-Time Video Smoke Detection Based on Deep Domain Adaptation for Injection Molding Machines / S.-H. Chen, J.-H. Jang, M.-J. Youh [et al.] // Mathematics. – 2023. – № 11(17). – Art. 3728. – DOI: 10.3390/math11173728.

[19] Zhang, Q. Wildland Forest Fire Smoke Detection Based on Faster R-CNN using Synthetic Smoke Images / Q. Zhang, G. Lin, Y. Zhang [et al.] // Procedia Engineering. – 2018. – Vol. 211. – PP. 441-446. – DOI: 10.1016/j.proeng.2017.12.034.

[20] Seefried, E. Smoke+: A Video Dataset for Automated Fine-Grained Assessment of Smoke Opacity : thesis for the Candidate's degree in Technical Sciences / E. Seefried. – Fort Collins : Colorado State University, 2024.

Авторский вклад

Адамовский Егор Русланович – анализ предметной области, формирование обучающей выборки, проектирование специального программного обеспечения.

Богуш Рихард Петрович – постановка задачи исследования, анализ полученных результатов, корректировка рукописи статьи.

TRAINING DATASET FORMATION FOR NEURAL NETWORKS IN SPATIO-TEMPORAL SMOKE DETECTION FROM VIDEO

Y.R. Adamovskiy
Senior Lecturer at the Department of Computer Systems and Networks, Polotsk State University

R.P. Bohush
Doctor of Technical Sciences, Professor, Head of Computer Systems and Networks department of Polotsk State University

Abstract. This paper presents a dataset formation technology for training and testing neural network models performing spatio-temporal analysis of frames for smoke detection on video sequences. A modification of the image sequence and metadata representation format is proposed for this task, ensuring compatibility with the COCO format. A distinctive feature of the proposed approach is the formation of a synthetic dataset by combining a particle system with background images and automated object annotation.

Keywords: smoke detection, computer vision, synthesized images, machine learning.

УДК 004.8

ПРИМЕНЕНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ РАСПОЗНАВАНИЯ АКТИВНОСТИ И УЛУЧШЕНИЯ КАЧЕСТВА ПОЗИЦИОНИРОВАНИЯ В ПОМЕЩЕНИИ



П. А. Шлык
БГУИР, г. Минск, Республика Беларусь
star-war-sandakan@mail.ru

П.А. Шлык
Окончил Белорусский государственный универсистет информатики и радиоэлектронники.

Аннотация. Исследуются теоретические предпосылки применения методов машинного обучения к задачам распознавания типа активности человека (Human Activity Recognition, HAR) и повышения качества позиционирования в помещении в условиях гетерогенного аппаратного обеспечения: различных

конфигураций сенсоров и характеристик среды. Рассматриваются теоретические основания для интеграции HAR-классификатора с байесовскими фильтрами, для применения нейронных сетей для сглаживания шума среды RSSI-сигнала, а также для обеспечения межустройственной переносимости моделей посредством алгоритмической калибровки и методов адаптации домена. Сформулированы гипотезы, подлежащие экспериментальной проверке.

Ключевые слова: HAR, машинное обучение, позиционирование

Введение. Классические аналитические методы позиционирования (трилатерация с использованием фильтра Калмана) не используют весь объём доступной информации: данные инерциальных датчиков (акселерометр, гироскоп), контекст активности, исторические траектории. Методы машинного обучения потенциально позволяют извлечь дополнительную ценность из потока сенсорных данных.

Существующие коммерческие решения (IndoorAtlas, Navigine) используют элементы машинного обучения, однако построение комплексной системы, объединяющей контекстно-зависимую навигацию, робастную обработку сигналов и методы межустройственного переноса моделей, остаётся малоисследованной областью.

HAR: распознавание типа активности. Задача распознавания типа активности (HAR) в данном контексте формулируется как классификация многомерных временных рядов. На вход модели подается скользящее окно сенсорных измерений $W=\{si\}i=t-Tt$ на интервале длительностью T, где каждый вектор $si=ax,ay,az,x,y,zT$ содержит показания акселерометра и гироскопа. В качестве рабочего диапазона предлагается T2,5 с, что соответствует характерным временным масштабам циклов базовых двигательных активностей. Целью является сопоставление данного окна с одним из дискретных классов физической активности $y\in$ {стоит, идёт, бежит, сидит, поднимается, спускается}.

Рассматриваемые архитектуры. В рамках исследования для классификации временных рядов рассматриваются две базовые нейросетевые архитектуры, выбор которых обусловлен доступными датчиками:

Архитектура 1 (Многослойная рекуррентная сеть LSTM): Ориентирована на устройства с полным набором инерциальных датчиков (акселерометр и гироскоп). Структурно модель представляет собой каскад рекуррентных слоев с применением методов регуляризации. Выбор данной архитектуры основывается на способности сетей с долгой краткосрочной памятью эффективно анализировать многомерные временные ряды, фиксируя характерные закономерности движения с учетом пространственной ориентации устройства, которую обеспечивает гироскоп [12].

Архитектура 2 (Гибридная нейронная сеть 1D-CNN-LSTM) [1]: Исследуется преимущественно для устройств с ограниченным аппаратным обеспечением (только трехосевой акселерометр), где данные подвержены высокому уровню позиционного шума из-за неизвестной ориентации устройства. Архитектура включает одномерный сверточный слой с последующей подвыборкой (pooling) и рекуррентным блоком. Теоретическое обоснование гибридного подхода заключается в том, что сверточный слой выступает в роли робастного фильтра, автоматически извлекая инвариантные к шуму локальные признаки (форму сигнала), тем самым снижая размерность и сложность задачи для последующего анализа временной динамики слоями LSTM.

Проблема гетерогенности устройств. Фундаментальной проблемой использования единой обученной модели на нескольких устройствах являются различия в самих датчиках, ориентации датчиков на печатной плате, частотах дискретизации (20-200 Гц), динамическом диапазоне и уровне шума. В литературе зафиксировано, что модели, обученные на данных одного устройства, демонстрируют значительное падение точности на других [7].

Существенная часть указанных различий может быть устранена алгоритмически, без привлечения методов машинного обучения. Предлагается калибровочный подход: модель обучается на данных в некоторой канонической системе координат, а при развертывании

на конкретном устройстве применяется матрица аффинного преобразования, приводящая показания датчиков к каноническому виду. Параметры преобразования могут быть определены на основании спецификаций производителя (ориентация осей, масштабный коэффициент). Аналогично, различия в частотах дискретизации устраняются алгоритмической передискретизацией к единой частоте. Ограничение данного подхода - необходимость знания аппаратной спецификации целевого устройства, что, выполнимо для большинства серийных моделей.

Для остаточных различий, не поддающихся алгоритмической компенсации (индивидуальные шумовые характеристики, нелинейности датчиков), рассматриваются методы адаптации домена. В частности, подходы по выравниванию распределений признаков между устройствами и частотно-ориентированная адаптация спектральных характеристик [7, 8]. Вопрос о том, в какой мере алгоритмическая калибровка снижает потребность в ресурсоёмких методах адаптации домена, подлежит экспериментальной проверке.

Гипотеза интеграции HAR с фильтром позиционирования. Результат HAR-классификации может быть использован для адаптивного переключения модели движения в байесовском фильтре. Теоретическим обоснованием данного подхода служит тот факт, что при разных типах активности характеристики движения качественно различаются:

Таблица 1. Модель движения для определённого типа активности

Активность	**Предполагаемая модель движения**
Стоит / сидит	Модель покоя, минимальный шум процесса
Идёт	Модель константной скорости (CV) [10]
Бежит	Модель константного ускорения (CA) [10]
Подъём по лестнице	Расширение вектора состояния z-компонентой

Данный подход концептуально близок к структуре фильтра с взаимодействующими множественными моделями (Interacting Multiple Model, IMM), в котором переключение между моделями движения управляется внешним HAR-классификатором [13]. Гипотеза состоит в том, что такая интеграция снижает ошибку позиционирования при смене режима движения по сравнению с фильтром с фиксированной моделью. Следует отметить, что использование скользящего окна длительностью T вносит соответствующую задержку в определение типа активности, что может влиять на точность фильтра при резких сменах режима движения.

Устранение влияния среды на сигнал. RSSI-сигнал в помещениях подвержен систематическим искажениям (≈4-8 дБ): многолучевое распространение, затухание при прохождении через препятствия, интерференция [2]. Классические методы фильтрации (медианный фильтр, скользящее среднее, фильтр Калмана) способны устранять высокочастотный шум, однако они не моделируют сложную пространственную структуру помех, специфичную для геометрии конкретного помещения.

Нейросетевое моделирование среды. В качестве основного подхода рассматривается использование шумоподавляющего автоэнкодера (1D-Denoising Autoencoder) для моделирования поведения сигнала: [14]

- Вход: окно шумных RSSI-измерений rt-w,...,rt.
- Выход: оценка «чистого» уровня RSSI.
- Функция потерь: минимизация среднеквадратичной ошибки (MSE) между предсказанием и эталонным расстоянием.

Задача архитектуры заключается не в прямой низкочастотной фильтрации, а в построении неявной нелинейной модели искажений сигнала. Сеть, обучаясь на наборе данных конкретной локации, извлекает закономерности (многолучевое замирание, мертвые зоны), характерные именно для этой среды. Гипотеза состоит в том, что обученная модель

среды способна реконструировать истинное расстояние с меньшей ошибкой, чем универсальные аналитические фильтры.

Позиционирование на основе радиокарты. Метод радиокарты (fingerprinting) предполагает сопоставление вектора значений RSSI от N точек доступа с двумерными координатами. Для аппроксимации этой функции рассматриваются глубокие нейронные сети (DNN) и алгоритмы градиентного бустинга [6]. Теоретическое преимущество данного подхода заключается в неявном учёте сложной геометрии помещения без построения аналитической модели распространения сигнала. Рассматривается гибридная схема, где метод радиокарты используется для грубой инициализации координат, а байесовский фильтр - для их сглаживания в реальном времени.

Подходы к формированию датасета. Качество и репрезентативность обучающей выборки во многом определяют итоговую точность и обобщающую способность моделей машинного обучения [9]. Использование существующих открытых наборов данных для разработки комплексной системы позиционирования и распознавания активности сопряжено с рядом ограничений. Во-первых, публичные HAR-датасеты содержат показания инерциальных датчиков, но не включают синхронных RSSI-измерений и истинных координат [3]. В свою очередь, базы данных для локализации в помещении не содержат высокочастотных IMU-потоков и разметки контекста активности [4].

Во-вторых, как было отмечено ранее, модель искажений сигнала жестко привязана к геометрии конкретного помещения, что делает нецелесообразным перенос пространственных RSSI-данных из сторонних источников. Кроме того, большинство открытых датасетов собраны с использованием ограниченного набора устройств при фиксированном положении смартфона (например, только на поясе или в кармане), что затрудняет обучение моделей, устойчивых к аппаратной гетерогенности и произвольной ориентации.

В связи с этим возникает необходимость формирования собственного набора данных, собранного in situ, а также применения алгоритмических методов для искусственного расширения его вариативности. Ниже рассматриваются подходы к решению этих задач.

Синтетический датасет из системы локализации. Использование существующей системы локализации (Indoor Positioning System, IPS) позволяет получать траектории движения с временными метками [5]. Дифференцирование этих координат позволяет оценить скорость пользователя, что может служить основой для автоматической разметки типа активности. Синхронизация таких пространственных данных с показаниями инерциальных датчиков потенциально обеспечивает формирование обогащенного датасета без необходимости ручной разметки.

Контрастное обучение. В рамках данного подхода из одного временного окна сигнала путем различных аугментаций формируются положительные пары. Процесс обучения кодировщика направлен на максимизацию согласованности представлений внутри таких пар и минимизацию сходства между независимыми окнами, что достигается за счет использования функции потерь InfoNCE [15].

Применимость контрастного обучения к задачам классификации временных рядов сенсорных данных показана в [16]. Ожидаемым результатом является формирование признаковых представлений, инвариантных к характеристикам конкретного устройства и его положению.

Методы аугментации сенсорных данных. Для искусственного обогащения обучающей выборки и повышения устойчивости моделей применяются методы аугментации временных рядов. Рассматриваемы подходы включают:

- добавление высокочастотного шума (jittering);
- нелинейное искажение временной шкалы (time warping);
- случайное пространственное вращение (rotation) векторов ускорения и угловой скорости;

- масштабирование амплитуды сигнала (magnitude scaling);
- перестановку временных сегментов в рамках одного окна (permutation).

Вопрос об оптимальном наборе методов аугментации для задачи классификации активности подлежит экспериментальному исследованию.

Гипотезы для экспериментальной проверки. H1: Интеграция HAR-классификатора с UKF (Unscented Kalman Filter), реализующим адаптивное переключение моделей движения (константной скорости, константного ускорения, покоя), снижает среднюю ошибку позиционирования по сравнению с UKF с фиксированной моделью [11].

H2: HAR-модель на основе модуля ускорения (Архитектура 2) обеспечивает приемлемую точность классификации базовых активностей на устройствах без гироскопа, достаточную для управления моделью движения в фильтре.

H3: 1D-Denoising Autoencoder для RSSI превосходит медианный фильтр и скользящее среднее по метрикам MAE/RMSE оценки расстояния в условиях многолучевого распространения.

H4: Алгоритмическая калибровка данных сенсоров посредством матричных преобразований обеспечивает переносимость HAR-модели между устройствами с приемлемой потерей точности, существенно снижая необходимость применения ресурсоёмких методов адаптации домена по сравнению с прямым переносом (direct transfer).

H5: Синтетическая разметка HAR-датасета через данные системы локализации (IPS) пригодна для обучения моделей при условии, что ошибка позиционирования IPS не превышает порогового значения, определяемого экспериментально (предположительно, порядка характерного размера шага - 0.5 - 1 м).

H6: Использование глубоких нейронных сетей для позиционирования на основе метода радиокарты обеспечивает более точную начальную оценку координат по сравнению с аналитическими моделями распространения сигнала в условиях сложной геометрии помещения.

H7: Предварительное контрастное обучение на неразмеченных данных сенсоров позволяет извлекать признаки, инвариантные к типу устройства, что повышает точность HAR-классификатора при переносе на новые устройства без необходимости сбора дополнительных размеченных данных.

Заключение. Основным результатом проведенного теоретического анализа является обоснование перехода от изолированных алгоритмов позиционирования и распознавания активности к единой контекстно-зависимой системе обработки.

Показано, что методы машинного обучения потенциально способны эффективно дополнять классические аналитические модели: нейросетевые архитектуры берут на себя извлечение робастных признаков из зашумленных данных сред с многолучевым распространением сигнала (моделирование RSSI) и гетерогенных сенсоров, в то время как байесовский фильтр обеспечивает математически строгую агрегацию этих данных в реальном времени.

Предложенный гибридный подход, а также методология алгоритмической калибровки и контрастного обучения для обеспечения межустройственной переносимости формируют теоретический фундамент для создания робастных систем локализации.

В качестве направления для дальнейших исследований рассматривается экспериментальная валидация выдвинутых гипотез в условиях реальных промышленных объектов с высоким уровнем зашумленности радиоэфира и непредсказуемой ориентацией пользовательских устройств.

Список литературы

[1] Ordóñez F. J., Roggen D. Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition // Sensors. 2016. Vol. 16(1). P. 115.

[2] Zafari F. et al. A Survey of Indoor Localization Systems and Technologies // IEEE Communications Surveys \& Tutorials. 2019. Vol. 21(3). Pp. 2568-2599.

[3] Anguita D. et al. A Public Domain Dataset for Human Activity Recognition Using Smartphones // ESANN. 2013.

[4] Torres-Sospedra J. et al. UJIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization // IPIN. 2014.

[5] Kang W., Han Y. SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization // IEEE Sensors Journal. 2015.

[6] Hsieh C. et al. Deep Learning for Indoor Localization with Wi-Fi Fingerprinting // IEEE Access. 2019.

[7] Ragab M. et al. AdaTime: A Benchmarking Suite for Domain Adaptation on Time Series Data // ACM Trans. Knowledge Discovery from Data. 2023.

[8] He R. et al. Domain Adaptation for Time Series Under Feature and Label Shifts (Raincoat) // ICML. 2023.

[9] Sambasivan N. et al. "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI // CHI '21. 2021.

[10] Li X. R., Jilkov V. P. Survey of maneuvering target tracking. Part I. Dynamic models // IEEE Transactions on Aerospace and Electronic Systems. 2003. Vol. 39(4). P. 1333-1364.

[11] Wan E. A., Van Der Merwe R. The unscented Kalman filter for nonlinear estimation // Proc. IEEE Symposium on Adaptive Systems for Signal Processing, Communications, and Control (ASSPCC). 2000. P. 153-158.

[12] Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9(8). P. 1735 - 1780.

[13] Blom H. A. P., Bar-Shalom Y. The Interacting Multiple Model Algorithm for Systems with Markovian Switching Coefficients // IEEE Transactions on Automatic Control. 1988. Vol. 33(8). P. 780 - 783.

[14] Vincent P. et al. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion // Journal of Machine Learning Research. 2010. Vol. 11. P. 3371 - 3408.

[15] van den Oord A., Li Y., Vinyals O. Representation Learning with Contrastive Predictive Coding // arXiv preprint arXiv:1807.03748. 2018.

[16] Eldele E. et al. Time-Series Representation Learning via Temporal and Contextual Contrasting // Proc. IJCAI. 2021. P. 2352 - 2359.

Авторский вклад

Шлык Павел Алексеевич - формирование статьи.

THEORETICAL FOUNDATIONS OF ML FOR ACTIVITY RECOGNITION

P. A. Shlyk

BSUIR, Minsk, Belarus

star-war-sandakan@mail.ru

Abstract. The research examines the theoretical foundations for applying machine learning methods to human activity recognition tasks (Human Activity Recognition, HAR) and improving indoor positioning accuracy under heterogeneous hardware conditions, including various sensor configurations and environmental characteristics.

It addresses the theoretical basis for integrating a HAR classifier with Bayesian filters, using neural networks to smooth RSSI signal environmental noise, and ensuring inter-device model portability through algorithmic calibration and domain adaptation techniques.

Hypotheses have been formulated for experimental verification.

Keywords: HAR, ML, indoor positioning

УДК 004.8:004.62:316.4

ДИНАМИЧЕСКИЕ ЦИФРОВЫЕ ДВОЙНИКИ КАК ИНСТРУМЕНТ СНИЖЕНИЯ СОЦИАЛЬНЫХ РИСКОВ БОЛЬШИХ ДАННЫХ В УСЛОВИЯХ ИНДУСТРИИ 4.0





В.Г. Евдокимов
Аспирант кафедры информационных технологий автоматизированных систем БГУИР
vigandvdk@gmail.com

А.А. Навроцкий
Декан факультета информационных технологий и управления БГУИР, кандидат физико-математических наук, доцент
navrotsky@bsuir.by

А.А. Навроцкий
Окончил Минский радиотехнический институт. Область научных интересов включает программные системы, системы компьютерного зрения и СВЧ-устройства.

В.Г. Евдокимов
Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов включает оптимизацию алгоритмов искусственного интеллекта, применение ИИ в условиях неполноты данных и разработку адаптивных автономных интеллектуальных систем.

Аннотация. В статье развивается концепция применения динамических цифровых двойников как инструмента согласования решений в сложных социально-экономических системах, функционирующих в условиях интенсивной цифровизации. Обосновывается, что основные риски Big Data в государственном и корпоративном управлении связаны не только с объёмом и скоростью обработки данных, но прежде всего с усредняющим характером алгоритмических моделей, фрагментарностью цифровизации и слабой согласованностью микро- и макроуровней принятия решений. Предлагается подход, при котором государство, предприятия, домохозяйства, индивиды и инфраструктурные контуры рассматриваются как взаимосвязанные источники уникальных и динамически изменяющихся данных. Их интеграция в систему многоуровневых цифровых двойников позволяет перейти от статического профилирования к адаптивному управлению, учитывающему реальные потребности, ограничения, траектории состояний и структуру межузловых связей. Показано, что наибольший эффект такой подход даёт не при локальном, а при сквозном внедрении, охватывающем жизненный цикл производства, распределения, потребления и государственного регулирования. Сделан вывод, что динамический цифровой двойник может быть интерпретирован не только как технологическая модель объекта, но и как инструмент объективизации распределения ресурсов и поддержания социально-экономического равновесия.

Ключевые слова: Big Data, цифровой двойник, динамический цифровой двойник, Индустрия 4.0, Индустрия 5.0, государственное управление, управление на основе данных, микроэкономика, макроэкономика, распределение ресурсов, алгоритмическая справедливость.

Введение. Цифровизация экономики и государственного управления сопровождается не просто ростом объёмов данных, а качественным изменением самой логики управленческого воздействия. Данные постепенно превращаются в самостоятельный фактор координации, прогнозирования и распределения ресурсов. В российской стратегической повестке этот подход закреплён на нормативном уровне: данные в

цифровой форме рассматриваются как значимый фактор повышения эффективности производства, управления и оказания услуг [1].

Однако накопление больших массивов данных само по себе не устраняет системные дисбалансы. Напротив, в ряде случаев Big Data усиливают структурные перекосы, поскольку управленческие решения начинают строиться на основе агрегированных профилей, усреднённых групп и исторических корреляций. В таком режиме отдельный субъект - человек, домохозяйство, предприятие, муниципалитет - фактически редуцируется до статистического образа, тогда как его актуальное состояние, ограничения и динамика изменения потребностей учитываются лишь частично. Поэтому рост объёма данных не гарантирует содержательной адекватности решения.

Проблема особенно заметна в сферах распределения ограниченных ресурсов: бюджетных средств, энергии, производственных мощностей, логистических возможностей, медицинских и образовательных услуг, продовольствия и трудовых ресурсов. Ошибка в интерпретации потребностей одного уровня системы способна породить каскадный дисбаланс на других уровнях. Вследствие этого локальная оптимизация перестаёт совпадать с общественной эффективностью.

В научной литературе цифровые двойники первоначально рассматривались преимущественно как инструмент моделирования технических объектов и производственных процессов, однако впоследствии область их применения расширилась и стала охватывать инфраструктуру, городскую среду и организационные системы [2-5]. Цель настоящей статьи состоит в том, чтобы предложить интерпретацию цифрового двойника как динамической модели согласования микро- и макроэкономических состояний. Исходный тезис заключается в следующем: каждый участник общественно-экономической системы является носителем уникальных данных; при включении этих данных в динамическую архитектуру цифровых двойников появляется возможность объективнее оценивать потребности и ограничения как конкретного узла, так и всей цепи управления в целом.

Теоретический контекст: от больших данных к управлению на основе связанных состояний. В исследованиях, посвящённых Big Data, давно отмечено, что проблема заключается не только в объёме, скорости и разнообразии данных, но и в способах их включения в управленческую практику. Для публичного сектора это означает переход к управлению, основанному на данных. Такая модель потенциально сокращает временной лаг между событием и реакцией системы управления, но одновременно повышает требования к качеству данных, архитектуре их интеграции и режимам доверия между участниками цифрового взаимодействия [6].

На практике цифровое управление часто строится по принципу платформенной централизации. Данные собираются из множества источников, агрегируются, стандартизируются, после чего используются для прогнозирования, ранжирования и автоматизации решений. Такой подход эффективен для задач массовой обработки, но имеет ряд уязвимостей.

Во-первых, такие решения основаны на статистической логике усреднения. В типовых ситуациях они могут демонстрировать достаточно высокую результативность, однако значительно хуже учитывают индивидуальные отклонения, нетипичные сочетания ограничений и редкие сценарии.

Во-вторых, для них характерна ориентация на локальную оптимизацию, при которой отдельная отрасль, ведомство или цифровой сервис стремятся повысить собственную эффективность без должного учёта последствий для связанных контуров.

В-третьих, подобные системы способны не только фиксировать, но и воспроизводить предвзятость, содержащуюся в исходных данных и сложившихся институциональных практиках. Этические аспекты алгоритмизации, включая непрозрачность процедур,

возможность дискриминационных эффектов и размывание ответственности, подробно рассматриваются в современной литературе [7–9].

В терминах системного управления это означает, что проблема сводится не столько к недостатку объёма данных, сколько к отсутствию такой модели, которая позволяла бы отражать изменяющиеся состояния объектов и структуру их взаимосвязей. Именно здесь концепция динамического цифрового двойника приобретает особую аналитическую ценность. Классическое понимание цифровых двойников формировалось в инженерном и производственном контексте как виртуальная репрезентация физического объекта или процесса, синхронизируемая с реальными данными и позволяющая прогнозировать его поведение, тестировать сценарии и оптимизировать жизненный цикл [2-4]. Для социально-экономических систем этого понимания уже недостаточно. Человек, предприятие или регион не являются техническими объектами в узком смысле; их состояние формируется под влиянием институтов, ресурсов, поведения других участников и внешних шоков. Следовательно, цифровой двойник в таких системах должен отражать не только параметры объекта, но и его позицию в распределённой структуре зависимостей. Иными словами, речь должна идти о динамическом цифровом двойнике, способном описывать изменение потребностей, ограничений и связей во времени.

Ограничения точечной цифровизации. Основная слабость современной цифровой трансформации заключается в том, что она часто реализуется по секторальному принципу. Развиваются цифровое производство, умное сельское хозяйство, цифровое здравоохранение, системы мониторинга занятости, городские платформы, цифровые бюджеты. Каждое из этих направлений имеет собственную логику, собственный массив данных и собственные показатели эффективности. Однако реальная социально-экономическая система не делится на изолированные домены. Потребление продовольствия связано с доходами, доходы - с занятостью и производственной структурой, производственная структура - с ресурсной и энергетической обеспеченностью, а всё вместе - с бюджетной, инфраструктурной и демографической динамикой.

Если цифровой двойник внедряется только на одном участке, он действительно способен повысить локальную точность решений. Но именно за счёт этого возникает риск перекоса: модель начинает улучшать тот сегмент, для которого она сконструирована, не видя системных последствий. Цифровой двойник агропромышленного комплекса, не связанный с динамикой потребительского спроса, логистики, цен, доходов населения и энергетических ограничений, может оптимизировать производство и одновременно усиливать диспропорции в распределении.

Содержательно это означает наличие четырёх взаимосвязанных проблем.

Первая проблема - редукция уникальности субъекта. Большие данные в традиционной платформенной архитектуре чаще фиксируют цифровой след, чем строят модель изменяющегося состояния.

Вторая проблема - разрыв микро- и макроуровня. Макроэкономические решения принимаются на основе агрегированных показателей, а микрорешения - на основе локальных профилей. Между ними часто отсутствует механизм согласования в режиме, близком к реальному времени.

Третья проблема - асимметрия участия. Те, кто генерирует данные, как правило, не контролируют способы их интерпретации и использования. Это усиливает недоверие и снижает легитимность управленческих решений.

Четвёртая проблема - фрагментарность цифровой архитектуры. Отдельные отрасли цифровизуются быстрее других, что не устраняет, а иногда даже усиливает общую дисбалансированность системы.

Таким образом, задача состоит не в том, чтобы собрать ещё больше данных, а в том, чтобы переопределить единицу анализа. Такой единицей должен стать не «профиль» и не

«сегмент», а динамический цифровой двойник субъекта, встроенный в сеть других двойников.

Динамический цифровой двойник как инструмент согласования. Предлагаемый подход основывается на следующем положении: любой элемент цепи управления - индивид, домохозяйство, предприятие, муниципалитет, отрасль, государственный институт, инфраструктурный контур - является источником уникальных данных, которые не сводятся к общему массиву без потери содержательной специфики. Однако если данные не просто собираются, а включаются в динамическую модель объекта, то уникальность перестаёт быть источником шума и становится основанием для более объективного решения.

В этом смысле динамический цифровой двойник выступает как универсальный уравнитель. Он не устраняет различия между субъектами, а делает их управленчески сопоставимыми. Сопоставимость достигается за счёт того, что каждый субъект описывается в единой логике: через структуру потребностей, доступных ресурсов, ограничений, связей и траектории изменения состояния. Тогда индивид, предприятие и государственный орган перестают быть несоизмеримыми объектами разных классов и могут анализироваться как узлы одной распределённой системы.

С практической точки зрения это означает переход от статической аналитики к адаптивному распределению. Например, потребности домохозяйства нельзя корректно оценить только по доходу или потребительскому профилю; необходимо учитывать состав семьи, территориальную доступность услуг, энергетические расходы, занятость, долговую нагрузку, сезонность и состояние здоровья. Аналогично предприятие нельзя описывать только объёмом выпуска или финансовой отчётностью: значимы зависимость от поставщиков, кадровая устойчивость, энергетическая обеспеченность, состояние оборудования, логистические риски и рыночная волатильность.

Если государственный уровень представлен цифровыми двойниками бюджета, ресурсов, социальной нагрузки, нормативных ограничений и инфраструктурных возможностей, а частный сектор и граждане поставляют данные о фактическом состоянии производства, потребления и потребностей, то формируется динамическая структура цифрового двойника микро- и макроэкономики. Её ценность состоит не в тотальном наблюдении, а в возможности согласовывать решения по всей цепи. Без государства невозможно зафиксировать общесистемные рамки, без бизнеса и граждан - достаточную детализацию фактического состояния системы. Односторонняя модель в любом случае будет неполной.

Формализация модели. Для концептуального описания введём систему взаимосвязанных цифровых двойников (формула 1):

$$DTt=\{DTiindt, DTifirmt, DTiregt, DTistatet, DTiinft\}, \quad (1)$$

где DTiindt - двойники индивидов и домохозяйств, DTifirmt - двойники предприятий, DTiregt - двойники региональных систем, DTistatet - государственный макродвойник, DTiinft - двойники инфраструктурных и ресурсных контуров.

Состояние каждого узла представим вектором (формула 2):

$$Sat= \langle Nat, Rat, Cat, Lat, Dat \rangle, \quad (2)$$

где Nat - вектор потребностей, Rat - вектор располагаемых ресурсов, Cat - совокупность ограничений, Lat - структура связей с другими узлами, Dat - параметры динамики состояния.

Тогда задача согласованного распределения ресурсов может быть записана как минимизация суммарного расхождения между потребностями и фактическим обеспечением при учёте межузловых дисбалансов: (формула 3):

$$t= a=1Aa*\|Nat- Raalloct\|+ \lambda* a=1Ab=1\ b\neq a\ Aab(t)\ , \quad (3)$$

где a - коэффициент значимости узла, Raalloct - объём фактически выделенных ресурсов, ab(t) - функция межузлового дисбаланса, - коэффициент системной связанности.

Динамика обновления цифрового двойника может быть представлена оператором перехода (формула 4):

$$Sa(t+1)=F(Sat, Xat, Uat, Eat), \quad (4)$$

где Xat - поток новых данных о состоянии узла, Uat - управленческое воздействие, Eat - внешние шоки и экзогенные эффекты.

Для оценки качества распределения введём показатель объективности решения (формула 5):

$$Ot=1-a=1A\|Nat- Raalloct\|\ a=1A\|Nat\|+ \varepsilon, \quad (5)$$

где $\varepsilon>0$ - малая константа.

Представленный аппарат задаёт состав системы, структуру состояния узла, целевую функцию адаптивного распределения, механизм обновления и критерий оценки результата.

Архитектура симбиоза цифровых двойников микро- и макроэкономики. Предлагаемая модель строится как многоуровневая архитектура.

На микроуровне формируются цифровые двойники индивидов, домохозяйств, предприятий, конкретных производственных объектов, сервисных платформ, ферм, логистических узлов и иных непосредственных участников экономического оборота. Здесь фиксируются фактические потребности, ресурсы, ограничения, циклы потребления, производственные нагрузки и поведенческие изменения.

На мезоуровне располагаются двойники муниципалитетов, регионов, отраслей, кластеров и цепочек поставок. Они агрегируют данные нижнего уровня, но не сводят их к простым средним значениям; задача этого уровня - выявлять структурные зависимости, территориальные перекосы и отраслевые узкие места.

На макроуровне находятся цифровые двойники государственного бюджета, социальной нагрузки, занятости, инфраструктуры, ресурсных балансов и иных параметров национальной экономики. Этот уровень задаёт рамки допустимых и приоритетных решений.

Связь между уровнями должна быть двусторонней. Сверху вниз система передаёт ограничения, нормативы, приоритеты и ресурсные лимиты. Снизу вверх поступают данные о реальном состоянии субъектов, спросе, дефицитах, избыточности, изменении нагрузок и нарушениях равновесия. В результате возникает динамический контур согласования, в котором каждый уровень корректирует другой.

Такой подход особенно важен для сфер, где жизненный цикл решения распределён между множеством разнородных узлов. Например, продовольственная система включает производство, поставщиков сырья, транспорт, хранение, энергоснабжение, розничную торговлю, доходы населения и бюджетные механизмы поддержки. Если цифровые двойники охватывают весь цикл, появляется возможность управлять не отдельными событиями, а соотношением состояний во всей цепи.

Преимущества и пределы применимости. Главное преимущество динамических цифровых двойников в предложенной трактовке состоит в том, что они позволяют заменить логику усреднения логикой согласования. Речь не идёт об отказе от агрегирования как такового; речь идёт о том, чтобы агрегирование не уничтожало состояние конкретных узлов, а сохраняло его в управленчески значимой форме.

Из этого вытекает ряд практических следствий.

Во-первых, повышается точность оценки потребностей. Система работает не с абстрактной категорией «потребитель», «предприятие» или «регион», а с изменяющейся структурой конкретного объекта.

Во-вторых, снижается риск алгоритмической несправедливости, поскольку решение опирается не только на историческую типологизацию, но и на текущее состояние субъекта [7; 9].

В-третьих, повышается прозрачность распределительных решений. Если решение может быть обосновано через параметры цифрового двойника и его место в системе взаимосвязей, оно приобретает более чёткое рациональное основание.

В-четвёртых, повышается устойчивость управления. Система получает возможность выявлять диспропорции на ранних стадиях, до того, как они проявятся в форме острого дефицита, перегрузки инфраструктуры или роста социального напряжения.

Вместе с тем следует учитывать и ограничения данной модели. Применение динамических цифровых двойников требует развитой инфраструктуры сбора и синхронизации данных, сопоставимости их форматов, устойчивого правового режима обращения с данными и достаточного уровня институционального доверия. При отсутствии этих условий модель может либо остаться декларативной, либо использоваться преимущественно как инструмент асимметричного контроля.

Заключение. Проведённый анализ показывает, что динамические цифровые двойники могут рассматриваться как перспективный инструмент снижения социальных рисков Big Data в том случае, если они понимаются не как изолированные цифровые копии отдельных объектов, а как взаимосвязанные модели состояний, включённые в единую систему управления. В таком понимании цифровой двойник выступает не только как средство мониторинга, но и как инструмент повышения обоснованности решений.

Ключевой результат предложенного подхода состоит в смещении акцента от работы с большими данными как совокупностью массивов информации к работе с динамически изменяющимися состояниями субъектов и структурой их взаимосвязей. Это позволяет в большей степени учитывать потребности индивидов, предприятий и территорий, соотносить локальные решения с макроэкономическими ограничениями и снижать диспропорции, возникающие в условиях фрагментарной цифровизации.

Тем самым подтверждается исходная гипотеза статьи: интеграция динамических цифровых двойников государства, частного сектора, граждан и инфраструктурных контуров способна сформировать более адаптивную систему управления и распределения ресурсов, тогда как их внедрение лишь в отдельных сферах создаёт риск системного дисбаланса.

Список литературы

[1] Указ Президента Российской Федерации от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017-2030 годы».

[2] Grieves M., Vickers J. Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems // *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* / ed. by F.-J. Kahlen, S. Flumerfelt, A. Alves. Cham: Springer, 2017. P. 85-113.

[3] Tao F., Zhang H., Liu A., Nee A. Y. C. Digital twin in industry: state-of-the-art // *IEEE Transactions on Industrial Informatics*. 2019. Vol. 15, no. 4. P. 2405-2415.

[4] Fuller A., Fan Z., Day C., Barlow C. Digital twin: enabling technologies, challenges and open research // *IEEE Access*. 2020. Vol. 8. P. 108952-108971.

[5] Kritzinger W., Karner M., Traar G., Henjes J., Sihn W. Digital twin in manufacturing: a categorical literature review and classification // *IFAC-PapersOnLine*. 2018. Vol. 51, no. 11. P. 1016-1022.

[6] Janssen M., Brous P., Estevez E., Barbosa L. S., Janowski T. Data governance: organizing data for trustworthy artificial intelligence // *Government Information Quarterly*. 2020. Vol. 37, no. 3. Art. 101493.

[7] *Mittelstadt B. D., Allo P., Taddeo M., Wachter S., Floridi L. The ethics of algorithms: mapping the debate //* Big Data & Society. *2016. Vol. 3, no. 2.*

[8] *Floridi L., Cowls J. A unified framework of five principles for AI in society //* Harvard Data Science Review. *2019. Vol. 1, no. 1.*

[5] *Batty M. Digital twins //* Environment and Planning B: Urban Analytics and City Science. *2018. Vol. 45, no. 5. P. 817–820.*

Авторский вклад

Навроцкий Анатолий Александрович – руководство исследованием и постановка задачи. Общее руководство научным проектом, формулировка научной проблемы.

Евдокимов Виталий Геннадьевич – разработка концепции, алгоритмического обеспечения и анализ практического применения.

DYNAMIC DIGITAL TWINS AS A TOOL FOR REDUCING THE SOCIAL RISKS OF BIG DATA IN THE CONTEXT OF INDUSTRY 4.0

V.G. Evdokimov
Postgraduate student of the Department of Information Technologies of Automated Systems, BSUIR

A.A. Navrotsky
Dean of the Faculty of Information Technology and Control, BSUIR, PhD in Physics and Mathematics, Associate Professor

Abstract. This article develops the concept of using dynamic digital twins as a tool for coordinating decisions in complex socio-economic systems operating under conditions of intensive digitalization. It is argued that the main risks of Big Data in public and corporate governance are associated not only with the volume and speed of data processing, but above all with the averaging nature of algorithmic models, the fragmented character of digitalization, and the weak coordination between micro- and macro-level decision-making. The article proposes an approach in which the state, enterprises, households, individuals, and infrastructure systems are viewed as interconnected sources of unique and dynamically changing data. Their integration into a system of multilevel digital twins makes it possible to move from static profiling to adaptive governance that takes into account actual needs, constraints, state trajectories, and the structure of inter-node relationships. It is shown that this approach yields the greatest effect not when implemented locally, but when deployed end-to-end across the entire lifecycle of production, distribution, consumption, and public regulation. It is concluded that a dynamic digital twin can be interpreted not only as a technological model of an object, but also as a tool for objectifying resource allocation and maintaining socio-economic balance.

Keywords: Big Data, digital twin, dynamic digital twin, Industry 4.0, Industry 5.0, public governance, data-driven governance, microeconomics, macroeconomics, resource allocation, algorithmic fairness.

УДК 004.8:001.891

РОЛЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В СОВРЕМЕННОМ НАУЧНОМ ПОЗНАНИИ



Е.В. Степанец
Преподаватель кафедры тактической и общевоенной подготовки военного факультета в БГУИР
e.stepanets@bsuir.by

Е.В. Степанец

Окончил учреждение образования «Военная академия Республики Беларусь». Область научных интересов связана с отдельными аспектами цифровой трансформации современных наук и философских взглядов на цифровизацию как процесс. Правовое регулирование отдельных вопросов цифровизации и информатизации.

Аннотация. В статье исследуется фундаментальное влияние информационно-коммуникационных технологий и искусственного интеллекта на трансформацию современного научного познания. Рассматривается переход от традиционной эпистемологии к цифровой реальности, в которой алгоритмы выступают в роли «когнитивных протезов», а классические лабораторные методы дополняются компьютерным моделированием. Особое внимание уделяется социальным и этическим вызовам цифровизации науки: угрозе утраты доказательности из-за непрозрачности нейросетей, размытию личной ответственности ученого, предвзятости алгоритмов и изменениям в форматах научной коммуникации. В качестве главного ответа на эти вызовы обосновывается необходимость формирования у исследователей «гуманитарного иммунитета» – глубокого синтеза технической грамотности и философско-этической подготовки. Автор подчеркивает, что только интеграция инженерного и гуманитарного образования позволит сохранить автономию субъекта познания и обеспечить безопасное развитие цивилизации в эпоху информационно-коммуникационных технологий.

Ключевые слова: цифровая трансформация науки, ИКТ, эпистемология, компьютерное моделирование, машинное обучение, научная коммуникация, гуманитарный иммунитет, философия, этика.

Введение. История науки – это история совершенствования наших органов чувств. Если Галилео Галилей и Антони ван Левенгук использовали оптику для расширения границ видимого, то современный исследователь использует алгоритмы для расширения границ мыслимого. Сегодня ученый все реже взаимодействует с физическим объектом напрямую. Между ними пролегает «цифровая дистанция»: серверы, облачные вычисления и сложные интерфейсы.

Это не просто техническое обновление. Мы наблюдаем трансформацию самого способа производства знания. Информационно-коммуникационные технологии (далее – ИКТ) перестали быть просто «линейкой» или «весами» – они превратились в «когнитивный протез», влияющий на формирование итогов научного исследования.

Немаловажным будет затронуть в заявленной теме и позицию кандидата юридических наук А.С. Игнатенко, которой в своей работе по вопросам цифровизации отметил, что «…единственный акт законодательства, который на сегодня содержит определение понятиям: «цифровизация», «цифровое развитие», является СТБ 2583–2020 «Цифровая трансформация. Термины и определения», введенный в действие с 1 марта 2021 г. постановлением Государственного комитета по стандартизации Республики Беларусь от 8 декабря 2020 г. № 95.». «На наш взгляд, устоявшегося и всецело сформированного определения понятия «цифровизация», как и правового разграничения с понятием «информатизация», в национальном законодательстве Республики Беларусь в области защиты государства, так и в иных не наблюдается.» [1, с. 390].

Традиционная эпистемология строилась на дихотомии: активный субъект (ученый) и пассивный объект (природа). Инструмент считался «прозрачным». Однако в цифровую эпоху эта прозрачность исчезает.

Эпистемологические основания применения ИКТ в научном познании. Современный исследователь работает в связке с алгоритмом. Возьмем сферу биоинформатики. Секвенирование генома генерирует такие объемы данных, которые человеческий мозг не способен охватить в принципе. Программное обеспечение не просто «показывает» данные, оно их конструирует. Цифровизация стирает автономию субъекта познания – он превращается в элемент сетевой системы, где пределы гипотез определяются вычислительной емкостью.

«Не менее проблемным является вопрос о том, какой тип личности нужен современному обществу. Идеалы и образцы прошлого вдохновляют сегодня немногих. В перспективе постчеловека фактом является ослабление потребности у нынешней молодежи к внутренним усилиям, развитию воли к достижениям, требующих сосредоточенной работы над собой. Все чаще ставка делается на «Google-мышление» и «помощь» Сети. Шанс самосохраниться видится в синтезе естественного с искусственным. Тем самым, возможно, открывается новый план эволюции человека. Но пока этот процесс носит спонтанный характер и имеет непредсказуемую траекторию. Эмпирически же можно

констатировать факт появления нового кумира – Искусственного интеллекта (ИИ).» [2, с. 294].

В своей работе кандидат юридических наук А.С. Игнатенко также признавал, что «информатизация» и «цифровизация» как правовые понятия соотносятся как общее и частное. Иными словами, информатизация как всеобъемлющий процесс включает в себя цифровизацию. Основой к разграничению понятий необходимо принимать техническую составляющую (технология), которая непосредственно взаимодействует с информацией [1, с. 395].

Цифровая трансформация породила новый тип эксперимента – *in silico* (компьютерное моделирование, симуляция эксперимента). Компьютерное моделирование позволяет моделировать процессы, которые невозможно воспроизвести в лаборатории: от столкновения галактик до долгосрочных климатических изменений.

Моделирование предоставляет исследователю способность управлять переменными. Можно остановить время, изменить гравитацию или повернуть вспять химическую реакцию.

«Создав триаду «модель – алгоритм – программа», исследователь получает в руки универсальный, гибкий и недорогой инструмент, который отлаживается и тестируется в «пробных» вычислительных экспериментах. После того, как адекватность (достаточное соответствие) триады исходному объекту имеется, с моделью проводятся разные «опыты», дающие все требуемые свойства и характеристики объекта. Важно, что процесс моделирования сопровождается улучшением и уточнением всех звеньев триады. Такой метод познания сочетает в себе преимущества и теории, и эксперимента. Действительно, работа не с самим объектом (явлением или процессом), а с компьютерной моделью дает возможность относительно быстро, без существенных затрат исследовать его свойства и поведение в любых мыслимых ситуациях. Это составляет преимущества теории. В то же время вычислительные эксперименты с моделями объектов позволяют, опираясь на мощь вычислительных методов и компьютеров, глубоко и полно изучать объекты, что недоступно чисто теоретическим подходам. Это уже составляет преимущества эксперимента.» [3, с. 9 – 10].

Однако, наряду с преимуществами, имеется ряд недостатков компьютерного моделирования:

1. Заблуждение о том, что моделирование может качественно обнаруживать новые явления, так как должно быть подтверждение в реальных условиях и в реальных экспериментах.
2. Модельный анализ уменьшает возможные объяснения. Из объекта моделирования можно «выжать» только то, что входит в рамки модели.
3. Чрезмерное упрощение ради скорости вычислений может отсечь ключевые факторы.

Методологические инновации и социальные аспекты цифровой трансформации науки. Самым сложным вызовом для академического сообщества стало внедрение машинного обучения. Нейронные сети часто выдают сверхточные предсказания (например, в диагностике заболеваний по МРТ), но не могут объяснить, как они пришли к такому выводу.

Этот разрыв ставит под угрозу главный идеал науки – доказательность. Если ученый не понимает логику системы, он превращается из исследователя в «верующего» пользователя. Г.И. Малыхина и В.И. Чуешов в своей работе подчеркивают важность сохранения критического подхода к использованию ИКТ, как один из главный вызов для современного образования.

«Трудно переоценить значение органичного диалектического единства парадигм образования как цели и средства в условиях современной цифровой трансформации жизни общества и человека. Цифровая трансформация открывает обучающимся доступ к самым

современным образовательным платформам, стирает с помощью цифры некоторые аспекты социального неравенства, а также нивелирует различия в месте и положении человека и общества в системе современного разделения общественного труда. В цифровом мире существенно возрастает потребность в философской подготовке не только представителей поколения Z, но и идущего ему на смену поколения. В набирающей темп стихии современной цифровой реальности эти поколения нередко должны будут действовать без опоры на помощь родителей и педагогов, формируя значительно раньше последних собственную устойчивую шкалу ценностей. Такую шкалу, в которой бы интерес к ИКТ не перечеркивал потребности всестороннего цельного развития человека, шкалу как такую социально-гуманитарную платформу личности, на которой спорт, музыка, искусство, философия и т.п. не исключали бы, а дополняли и направляли потребности и интересы развития Интернета вещей и Интернета идей» [4, с. 162].

Цифровая трансформация науки – это не только изменения в методах и философских основаниях, но и серьезные социальные изменения. Меняется то, как ученые общаются друг с другом, как организуется научная работа, как оцениваются результаты, какие этические проблемы возникают.Традиционная научная коммуникация строилась вокруг печатных журналов и личных встреч на конференциях. Ученый писал статью, отправлял ее в журнал, проходил рецензирование (этот процесс мог занимать месяцы и даже годы), статья публиковалась и становилась доступной подписчикам журнала. Сегодня эта система быстро меняется. Электронные журналы ускорили публикацию, но главное изменение связано с появлением препринт-серверов – платформ, где ученые могут выкладывать свои работы до их публикации в журналах, сразу делая их доступными для всего научного сообщества. Это ускоряет распространение знаний, позволяет получать обратную связь на ранних стадиях, снижает зависимость от издательских монополий. Развиваются социальные сети для ученых (ResearchGate, Academia.edu), где можно обсуждать работы, задавать вопросы авторам, находить коллег для совместных проектов. Формируются новые формы публикации – данные и код становятся такими же важными результатами исследования, как и статьи. Появляются платформы для обмена данными и программным обеспечением, что повышает прозрачность и воспроизводимость исследований.

ИКТ позволяют создавать новые формы организации науки –распределенные исследовательские коллективы. Ученые из разных стран могут работать над общей задачей, не покидая своих институтов. Это особенно важно для крупных проектов, требующих объединения ресурсов и компетенций. Такие проекты не только ускоряют обработку данных, но и привлекают широкую публику к участию в научных исследованиях. Цифровизация науки порождает новые этические проблемы. Одна из них – это вопрос о приватности при использовании больших данных. Многие научные исследования используют данные о людях – медицинские записи, данные социальных сетей, информацию о перемещениях. Другая проблема – это вопрос об ответственности за результаты, полученные с помощью алгоритмов. Если алгоритм машинного обучения допускает ошибку, кто несет ответственность – разработчик алгоритма, исследователь, использовавший его, или организация, в которой проводилось исследование? Традиционная наука исходила из принципа личной ответственности исследователя, но в условиях, когда значительная часть работы делегируется компьютеру, этот принцип требует переосмысления. Важной этической проблемой является и вопрос о предвзятости алгоритмов. Если алгоритм обучается на данных, которые содержат систематические искажения, то он будет воспроизводить и усиливать эти искажения. Исследователи, использующие методы машинного обучения, должны осознавать эту проблему и предпринимать шаги для ее выявления и устранения.

Г.И. Малыхина и В.И. Чуешов в своей работе о гуманитарной безопасности подчеркивают, что цифровая трансформация всех сфер жизни, включая науку, требует осмысления рисков и угроз, связанных с цифровизацией.

«… в наши дни важнейшими средствами устранения угроз в сфере отечественной гуманитарной безопасности оказываются: обеспечение достойного уровня притока научных и научно-педагогических кадров в экономику страны; увеличение притока в образование и науку молодежи; обеспечение достойного уровня финансирования науки, здравоохранения, образования и др. Среди последних важно обратить внимание на угрозы, связанные с особенностями развития информационно-коммуникативной и интеллектуальной сфер жизнедеятельности белорусского общества» [5, с. 259 – 260].

Интеграция гуманитарного знания в подготовку инженеров будущего. Цифровизация – это не только про код, это про смыслы. Когда вычислительные задачи забирает машина, за человеком остается самое сложное: задавать вопросы и нести этическую ответственность.

Г.И. Малыхина и А.А. Лисенкова в работе об интеграции технического и гуманитарного образования отмечают, что подготовка специалистов, способных эффективно работать в условиях цифровой трансформации, требует преодоления разрыва между техническим и гуманитарным знанием.

«… интеграция гуманитарного образования в подготовку инженерных кадров приносит множество выгод для общества, создавая более гибких, нравственно и социально ориентированных профессионалов. Этот подход, развивая технические навыки, учитывает социокультурные и этические особенности, и формирует целостное понимание проблем и задач, стоящих перед инженером. В итоге, интеграция технических и гуманитарных знаний становится не просто стратегическим выбором, но и необходимостью для формирования образованных, творческих специалистов, способных не только создавать передовые технологии, но и применять их с уважением к особенностям социокультурного контекста. Это единство знаний становится ключом к настоящему и будущему, где технологии и человеческое развитие тесно переплетаются, обеспечивая устойчивое и сбалансированное развитие общества и государства.» [6, с. 145 – 146]. Подготовка ученого будущего должна базироваться на синтезе технической грамотности и глубокой гуманитарной культуры. Без понимания контекста, этики и философии науки исследователь рискует стать заложником собственных инструментов. В некотором смысле можно сказать, что интеграция гуманитарного знания – это «иммунная система» науки в эпоху ИКТ.

Заключение. Таким образом, главный вызов современности, в контексте роли информационных технологий в современном научном познании, заключается в сохранении автономии субъекта. Чтобы не стать заложником собственных инструментов, ученый будущего должен обладать «гуманитарным иммунитетом» – синтезом технической грамотности и глубокой философско-этической подготовки. Только интеграция технического и гуманитарного знания позволит использовать весь спектр возможностей ИКТ, сохраняя за человеком роль того, кто несет ответственность за вектор развития цивилизации.

Список литературы

[1] Игнатенко, А. С. Правовое обеспечение цифрового развития института защиты Республики Беларусь / А. С. Игнатенко // Актуальные проблемы достижения целей устойчивого развития в условиях цифровой трансформации государства и права в Республике Беларусь : материалы Респ. науч.-практ. конф., Минск, 18–19 нояб. 2022 г. / БГУ, Юрид. фак., Каф. конституционного права ; редкол.: Г. А. Василевич (отв. ред.) [и др.]. – Минск, 2022. – С. 389–396.

[2] Малыхина, Г. И., Миськевич, В. И. Гуманитарная культура в контексте вызовов цифровой эпохи [Электронный ресурс] / Малыхина, Г. И., Миськевич, В. И. – Режим доступа: https://libeldoc.bsuir.by/handle/123456789/53659 – Дата доступа: 30.03.2026.

[3] Паничев В.В., Соловьев Н.А. Компьютерное моделирование [Электронный ресурс] / Паничев В.В., Соловьев Н.А. – Режим доступа: https://elib.osu.ru/bitstream/123456789/8694/1/2698_20110926.pdf. – Дата доступа: 30.03.2026.

[4] Малыхина Г.И., Чуешов В.И. Социально-гуманитарное образование как средство и цель в цифровую эпоху [Электронный ресурс] / Малыхина Г.И., Чуешов В.И. – Режим доступа: https://libeldoc.bsuir.by/handle/123456789/41915 – Дата доступа: 30.03.2026.

[5] Малыхина Г.И., Чуешов В.И. Гуманитарная безопасность в структуре национальной безопасности Республики Беларусь [Электронный ресурс] / Малыхина Г.И., Чуешов В.И. – Режим доступа: https://libeldoc.bsuir.by/handle/123456789/46476 – Дата доступа: 30.03.2026.

[6] Лисенкова А.А., Малыхина Г.И. Интеграция технического и гуманитарного: новый подход к образованию инженеров будущего [Электронный ресурс] / Лисенкова А.А., Малыхина Г.И. – Режим доступа: https://libeldoc.bsuir.by/handle/123456789/54963 – Дата доступа: 30.03.2026.

Авторский вклад

Степанец Ефим Владимирович – постановка задачи исследования, разработка методологии исследования, руководство исследованием.

THE ROLE OF INFORMATION TECHNOLOGIES IN MODERN SCIENTIFIC KNOWLEDGE

E.V. Stepanets
Lecturer in the Department of Tactical and General Military Training Military Faculty, BSUIR

Abstract. The article explores the fundamental influence of information and communication technologies and artificial intelligence on the transformation of modern scientific knowledge. It examines the transition from traditional epistemology to a digital reality, in which algorithms act as 'cognitive prostheses,' and classical laboratory methods are supplemented by computer modeling. Special attention is paid to the social and ethical challenges of the digitalization of science: the threat of losing evidentiality due to the opacity of neural networks, the blurring of personal responsibility of the scientist, algorithmic bias, and changes in the formats of scientific communication. As the main response to these challenges, the necessity of forming a 'humanitarian immunity' in researchers is justified – a deep synthesis of technical literacy and philosophical-ethical training. The author emphasizes that only the integration of engineering and humanitarian education will allow the preservation of the autonomy of the knowing subject and ensure the safe development of civilization in the era of information and communication technologies.

Keywords: digital transformation of science, ICT, epistemology, computer modeling, machine learning, scientific communication, humanistic immunity, philosophy, ethics

УДК 621.391:004.75

CLASSIC AND MODERN ENCRYPTION ALGORITHMS FOR BIG DATA PROCESSING



Y. Liashchevich
Postgraduate of the BSUIR



Cai Yuxin
Graduate of Luoyang Normal University

Y. Liashchevich

Graduated from BSUIR. Postgraduate in the Department of Electronic Engineering and Technology. His research interests include the effects of infrasound on biological tissue, infocommunication signals and systems.

Cai Yuxin

Graduate of Luoyang Normal University. The area of scientific interest is research into encryption and decryption technologies for communication signals based on new algorithms.

Abstract. The article discusses the using methods of research into encryption and decryption technologies for communication signals based on new algorithms in big data systems

Keywords: encryption, decryption, channel coding, burst interference.

Introduction. The modern world is living in an era of rapid development in wireless technologies and big data processing. Wireless systems such as 5G and 6G IoT, satellite communications transmit massive amounts of sensitive data, must ensure data transmission security [1]. As data transfer speeds increase, so does the volume of data. Routing and data storage devices must ensure maximum data protection from unauthorized access. Modern data encryption algorithms, are used for this purpose. Research Questions it is how effective are new encryption algorithms for communication signals? And can these algorithms resist emerging threats? What trade-offs exist between speed, security, and resource usage?

Basic encryption algorithms. Encryption is a technique for securing the information so only the authorized parties can interpret the data. What happens is that encryption provides transformations on the plaintext and transforms it into ciphertext which isn't human-readable. Not only in the digital era but ciphers have also been widely used throughout history. In the figure below, we present the general view of the classification of the ciphers. As we've seen, we categorize ciphers in terms of some characteristics that they have or according to their usage. They are mainly divided into two different types; classical and modern classes. The most common and used in digital are is modern class. It's because of its dynamic and static cryptographic techniques that pave the way for more detailed classes, symmetric and asymmetric ciphers (Fig. 1):

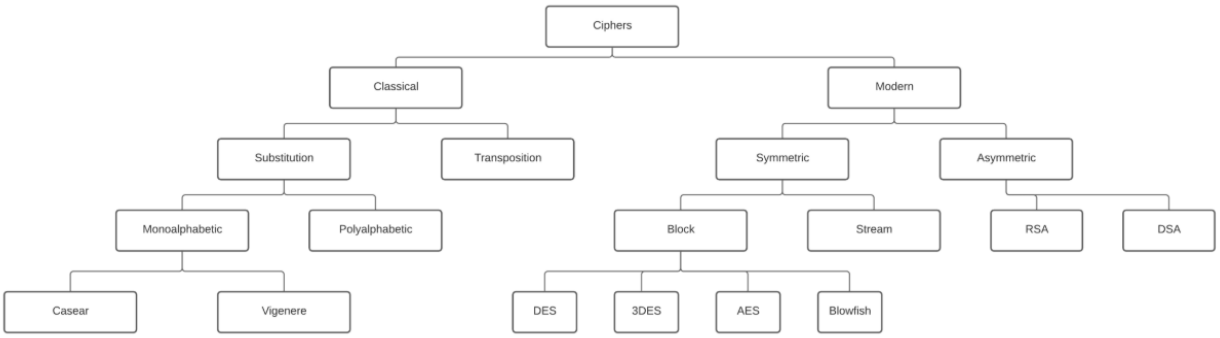


Figure 1. Classification of encryption algorithms

Symmetric encryption uses only one key to encrypt and decrypt data, while asymmetric encryption uses two different keys [2]. A public key is used for encryption, and a private key is used for decryption. Despite the differences between the public and private keys, they are related and are therefore called a key pair. DES, 3DES, AES, and Blowfish belong to the class of symmetric block ciphers.

Encryption Algorithms. DES is developed by IBM and based on a design by Horst Feistel. It was one of the widely used and publicly available cryptographic systems when it was first released. Even though its first debut is in the 70s, it was later adopted by the National Institute of Standards and Technology (NIST). It's a symmetric-key algorithm for the encryption of digital data. It has a block size of 64 bits and uses the Feistel network as a structure. It's slow and not used in the software.

Triple DES, 3DES, or TDES is officially the Triple Data Encryption Algorithm. It's a symmetric-key block cipher and it applies the DES algorithm three times to each block. It has a block size of 64 bits and a key length of 112 or 168 bits. It also uses the Feistel network since it's based on DES [2,3].

Due to the modern cryptology techniques and supercomputing, like the DES, 3DES has also some serious vulnerabilities. That's why the NIST has deprecated DES and 3DES for new applications in 2017 and for all applications by 2023. Blowfish is another symmetric-key encryption technique designed by Bruce Schneier in 1993 as an alternative to the DES encryption algorithm. Therefore, it is significantly faster than DES and provides a good encryption rate. Its key length is 446 bits, and way better than DES, and 3DES. Therefore, it's more difficult to crack the key of Blowfish. It also has a block size of 64 bits. It can be used in software as well. However, AES receives more attention today, and Schneier recommends Towfish as an alternative to Blowfish. It has a free license and is available for all uses. AES is another type of cipher that protects the data from malicious parties. It's currently one of the strongest encryption algorithms available. Since AES combines speed and security properly, it allows us to carry on with our online activities without any interruption. As AES uses the same key to both encrypt and decrypt data, it is also a symmetric type of encryption. There are three types of lengths of AES encryption keys 128, 192, and 256 bits. Each key length has different possible key combinations. It has a different structure than other encryption algorithms, it uses the substitution-permutation network.

Emerging Encryption Algorithms. Recent developments and emerging trends of cyber security

There are many recent developments in cyber security with the help of new algorithms, procedures and frameworks. Quantum Key Distribution (QKD) is a groundbreaking method in quantum cryptography that enables secure communication through the exchange of cryptographic keys [4]. Its security is rooted in the principles of quantum mechanics, particularly the no-cloning theorem and quantum entanglement, which make it virtually impossible for an eavesdropper to intercept or duplicate the key without being detected. Basic Principle:

QKD allows two parties (usually referred to as A and B) to securely exchange a cryptographic key over a potentially insecure channel. The security of QKD comes from quantum properties such as quantum superposition and entanglement.

- **Quantum Mechanics Behind QKD:**
- **No-Cloning Theorem**: It states that an arbitrary quantum state cannot be copied exactly, meaning any attempt to intercept a quantum key would disturb the state and be detectable.
- **Heisenberg Uncertainty Principle**: Measuring a quantum state disturbs it, so any eavesdropping attempts can be easily detected by the parties involved.

Types of QKD Protocols

There are several types of QKD protocols, but the most well-known ones include:

- **BB84 Protocol (1984)**:
- The first and most widely used QKD protocol, developed by Charles Bennett and Gilles Brassard.
- It uses **polarization** states of photons (or other quantum particles) to encode information. Alice sends photons in one of four states, and Bob measures them using randomly chosen bases.
- If no eavesdropping occurs, Alice and Bob will share a secure key.
- **E91 Protocol (1991)**:
- Based on **quantum entanglement**, where two particles (photons) are entangled in such a way that the measurement of one instantly affects the state of the other, even if they are far apart.
- This protocol uses **Bell's theorem** to ensure the authenticity and security of the key exchange.
- **Decoy State Protocols:**
- A modification to the basic QKD protocols, designed to mitigate the impact of photon number splitting (PNS) attacks, which can occur when an eavesdropper tries to intercept the photons in a weak coherent pulse.
- **Measurement-Device-Independent QKD (MDI-QKD):**
- Addresses the vulnerability of QKD systems to imperfections in detectors by enabling secure communication even when measurement devices are compromised.

Security of QKD

- **Quantum Security:**
- The key feature of QKD is its **unconditional security**. In theory, QKD can offer perfect security against any computational or classical cryptographic attack.
- Any eavesdropping or interference with the quantum channel results in observable errors that alert the communicating parties. This is because the quantum state will be disturbed, revealing the presence of an intruder.
- **Practical Security Concerns:**
- **Side-Channel Attacks**: Despite the inherent security of QKD, practical implementations can still be vulnerable to side-channel attacks, which exploit weaknesses in the physical hardware, such as detectors or transmission equipment.
- **Device Calibration & Trust**: The security of QKD systems relies heavily on trusted devices. If any device is compromised, the whole system's security can be compromised.

Post-Quantum Cryptography (PQC) involves developing cryptographic algorithms run on conventional computers that are secure against attacks from future, powerful quantum computers. These algorithms, such as lattice-based cryptography, replace current standards like RSA and ECC, which are vulnerable to Shor's algorithm [3,4]. PQC ensures long-term data security, preventing "harvest now, decrypt later" threats.

Categories of PQC Algorithms

PQC algorithms fall into several categories based on the type of problem they aim to solve and the type of cryptographic function they support.

- **Public-Key Cryptography** (asymmetric encryption):
- These are algorithms that rely on the difficulty of certain problems (e.g., integer factorization, discrete logarithms) which quantum computers can break but classical computers cannot efficiently solve.
- **Symmetric Cryptography**:
- These are used to encrypt and authenticate data using shared secrets. Quantum computers pose a smaller threat here (as Grover's algorithm only offers quadratic speedup, not an exponential one), but adjustments (like doubling key sizes) can maintain security.
- **Hash-Based Cryptography**:
- These use hash functions as the foundation for creating signatures and other forms of public-key cryptography. They are quantum-safe because quantum computers don't offer significant speedup over classical brute-force attacks on hash functions.

Candidate Algorithms for PQC (NIST Standardization)

The **NIST Post-Quantum Cryptography Standardization Process** aims to evaluate and standardize quantum-resistant algorithms. As of 2026, the final selections are still ongoing, but several promising candidates have emerged:

Public-Key Algorithms

- **Lattice-Based Cryptography**:
- Based on the hardness of lattice problems, which are believed to be resistant to quantum algorithms.
- Examples:
- **Kyber** (Encryption and Key Establishment)
- **NTRU** (Encryption and Key Exchange)
- **FrodoKEM** (Key Encapsulation)
- **NTS-KEM** (Key Encapsulation)
- **Code-Based Cryptography**:
- Based on error-correcting codes, specifically the difficulty of decoding random linear codes.
- Example: **McEliece** (Encryption)
- Known for large key sizes, but extremely resistant to quantum attacks.

- **Multivariate Polynomial Cryptography**:
- Relies on the difficulty of solving systems of multivariate polynomial equations over finite fields.
- Example: **Rainbow** (Digital Signatures)
- **Hash-Based Cryptography**:
- Based on the security of hash functions. While these are quantum-resistant, they generally have large signature sizes.
- Example: **XMSS** (Extended Merkle Signature Scheme)
- Used for signing messages with a smaller state, suitable for high-speed environments.
- **Isogeny-Based Cryptography**:
- Uses the mathematical structure of elliptic curves and their isogenies (maps between elliptic curves).
- Example: **SIKE** (Supersingular Isogeny Key Exchange)
- Considered a potential lightweight and efficient choice for key exchange.

Symmetric-Key Cryptography

- **AES (Advanced Encryption Standard)**:
- AES is already considered secure, but in a post-quantum world, it would require key lengths to be doubled (e.g., AES-256 instead of AES-128) to maintain security.
- **SHA-3 (Secure Hash Algorithm 3)**:
- SHA-3 is designed to be resistant to quantum attacks, as quantum computers do not provide significant speedup for hash functions.

Current State of PQC. NIST Standardization Process: The NIST post-quantum cryptography initiative is working towards finalizing the quantum-resistant standards. As of 2026, the finalists for encryption and key exchange are Kyber and NTRU, and FrodoKEM is a possible alternative for scenarios requiring more conservative security. For digital signatures, FALCON (lattice-based) and Rainbow (multivariate) are among the leading candidates.

Industry Adoption. While no large-scale commercial implementations of PQC are widespread yet, various companies and organizations are testing PQC algorithms in pilot projects. Some entities are using hybrid systems combining classical and quantum-safe algorithms to ensure security today and in the future.

Future of PQC.

- **Hybrid Cryptographic Systems**:
- During the transition period (while quantum computers are still under development), hybrid systems using both classical and post-quantum algorithms will likely be implemented for additional security.
- **Quantum Networks & Quantum Key Distribution (QKD)**:
- As quantum communication networks develop, PQC algorithms will be integrated into these systems, potentially along with QKD, to provide an extra layer of security.
- **AI and Machine Learning in Cryptography**:
- Machine learning and AI might help optimize PQC systems or analyze vulnerabilities in new PQC algorithms, driving further innovation.

Conclusion. In this article, we've briefly given some information about DES, 3DES, Blowfish, and AES, QKD, PQC. Also, we compared and evaluated them in terms of some metrics. Although we think that encryption algorithms protect the data both in theory and practice, keep in mind that there are various attacks such as malware, key search, brute force, phishing, side-channel, and more. These attacks aim to find a way to hack the system. To sum up, while every encryption algorithm has weaknesses and advantages, some of them can lose their reliability over time. So, we should be careful when we select cryptographic algorithms. We should both consider our application and the specifics of the encryption algorithm.

References

[1] Andrea Goldsmith - Wireless Communications. -Cambridge University Press. – 2005 pp. 204-207

[2] Singhal, Nidhi and Raina, J P S. “Comparative Analysis of AES and RC4 Algorithms for Better Utilization”, International Journal of Computer Trends and Technology, ISSN: 2231-280, July to Aug Issue 2011, pp. 177-181.

[3] Nadeem, Aamer; "A Performance Comparison of Data Encryption Algorithms", IEEE 2005.

[4] Abdulbast A. Abushgra Variations of QKD Protocols Based on Conventional System Measurements. - 2024

Author's contribution

The authors contributed equally to the writing of the article.

КЛАССИЧЕСКИЕ И СОВРЕМЕННЫЕ АЛГОРИТМЫ ШИФРОВАНИЯ ДЛЯ ОБРАБОТКИ БОЛЬШИХ ДАННЫХ

Е.И. Лещевич
Выпускник БГУИР, преподаватель в Лоянский Педагогический Университет

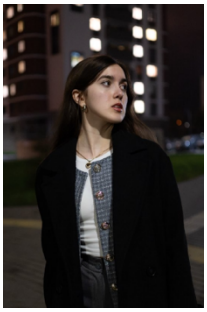
Цай Юйсинь
Выпускник в Лоянский Педагогический Университет

Аннотация. В статье рассматриваются методы исследования технологий шифрования и дешифрования коммуникационных сигналов на основе новых алгоритмов в системах обработки больших данных.

Ключевые слова: шифрование, дешифрование, канальное кодирование, пакетные помехи.

УДК 658.5.011

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИЙ BIG DATA И КОНЦЕПЦИИ BYOD ДЛЯ ОЦЕНКИ ВЛИЯНИЯ МОБИЛЬНЫХ УСТРОЙСТВ НА ПРОИЗВОДИТЕЛЬНОСТЬ ТРУДА В ОРГАНИЗАЦИЯХ



У.А. Пателина
Студент инженерно-экономического факультета БГУИР
ulya_P1999@mail.ru



О.Н. Шкор
Старший преподаватель кафедры экономики БГУИР, магистр экономических наук, доктор философии в области экономики
shkor@bsuir.by

У.А. Пателина

Родилась в 2007 году в Орше. В 2024 году закончила ГУО «Гимназия №1 г. Орши». В этом же году поступил в УО «БГУИР», была зачислена на бюджетную форму обучения по специальности «Цифровой маркетинг» инженерно-экономического факультета.

О.Н. Шкор

Родилась в Минске. Закончила БПИ в 1984 году. В 2001 г. защитила магистерскую диссертацию на тему: «Использование блочно-модульной системы обучения в профессиональной ориентации школьников» по специальности 08.00.05 «Экономика и управление народным хозяйством». В 2010 г. защитила докторскую диссертацию (PhD) на тему: «Экономическое обоснование формирования транспортной логистической системы Республики Беларусь» в Международной кадровой академии (Киев). С 2014 г. по настоящее время - заместитель заведующего кафедрой по научно-исследовательской работе студентов.

Аннотация. В статье исследуется влияние мобильных технологий и концепции BYOD на производительность труда с применением методов Big Data. Рассматриваются преимущества использования мобильных устройств в корпоративной среде, включая повышение гибкости и ускорение коммуникаций. Особое внимание уделяется возможностям аналитики больших данных для оценки эффективности трудовых процессов и оптимизации бизнес-решений. Делается вывод о том, что интеграция мобильных технологий и Big Data способствует повышению эффективности управления трудовыми ресурсами.

Ключевые слова: мобильные технологии, BYOD, Big Data, производительность труда, аналитика данных, цифровизация, корпоративные системы.

Введение. Цифровая трансформация экономики на современном этапе во многом обусловлена повсеместным внедрением мобильных решений и инструментов глубинной аналитики в корпоративную среду [1].

Интеграция смарт-устройств в повседневные рабочие процессы перестала быть вспомогательным элементом, превратившись в фундаментальную основу операционной гибкости и обеспечения непрерывного доступа к значимым массивам информации [2].

Параллельно с этим инструментарий Big Data открывает новые горизонты для обработки крупных массивов неструктурированной информации [3], позволяя выявлять скрытые корреляции, влияющие на итоговую эффективность организации.

В связи с этим особый интерес представляет анализ взаимовлияния распространения мобильных технологий и динамики производительности труда с применением современных методов обработки данных.

Трансформация трудовых процессов под влиянием мобильных технологий и концепции BYOD. Внедрение мобильных технологий влечет за собой кардинальный пересмотр традиционных подходов к организации труда [2].

Стираются жесткие привязки к стационарному рабочему месту, что приобретает критическое значение в эпоху доминирования удаленных форматов занятости и географически распределенных команд.

Модель BYOD, подразумевающая эксплуатацию личных гаджетов сотрудников для выполнения корпоративных задач [1], выступает драйвером ряда качественных изменений. Как свидетельствуют данные, подобный подход стимулирует:

- Рост автономности и пространственной мобильности персонала.
- Сокращение временных разрывов в принятии управленческих решений.
- Интенсификацию горизонтальных и вертикальных коммуникаций;
- Оптимизацию временных затрат за счет стирания границ между личным и рабочим временем (в контексте доступности).

В предшествующих исследованиях автора было подтверждено, что применение мобильных технологий коррелирует с ростом показателей производительности [1], что объясняется снижением барьеров доступа к корпоративным сервисам и ускорением циклов взаимодействия между сотрудниками.

Аналитический потенциал Big Data в оценке трудовой эффективности. Технологии Big Data позволяют перейти от субъективной оценки [3] эффективности труда к объективному анализу на основе данных.

В контексте мобильных технологий это включает:

- Анализ активности сотрудников в корпоративных системах.
- Обработку данных о времени выполнения задач.
- Оценку коммуникационной нагрузки.
- Выявление узких мест в бизнес-процессах.

С помощью инструментов Big Data организации могут [4]:

- Прогнозировать производительность сотрудников.
- Оптимизировать рабочие процессы.
- Выявлять неэффективные практики использования мобильных устройств.

В рамках анализа эффективности использования мобильных технологий целесообразно выделить основные типы данных, формируемых в рамках концепции BYOD и применяемых в бизнес-аналитике (таблица 1).

Таблица 1 . Категории данных

Категория данных	Источник генерации	Потенциал для оптимизации бизнеса
Геопространственные	GPS, Wi-Fi-трекинг	Оптимизация логистики, анализ перемещений сотрудников, планирование маршрутов [2]
Поведенческие	Паттерны использования приложений, временная активность	Анализ эффективности рабочего времени, выявление неэффективных действий [3]
Коммуникационные	Метаданные звонков и сообщений	Анализ взаимодействий в коллективе, улучшение коммуникаций [4]
Телеметрия устройств	Заряд устройства, состояние сети, производительность	Прогнозирование сбоев, оптимизация IT-инфраструктуры [5]

Проблемы и риски интеграции. Несмотря на преимущества, интеграция мобильных технологий и Big Data сопровождается рядом проблем [5]:

- Информационная безопасность. Использование личных устройств повышает риск утечки данных и требует внедрения дополнительных механизмов защиты.
- Рост нагрузки на IT-инфраструктуру. Обработка больших объёмов данных требует значительных вычислительных ресурсов.
- Конфиденциальность данных сотрудников. Сбор и анализ данных о деятельности сотрудников должен соответствовать требованиям законодательства.
- Необходимость квалифицированных специалистов. Для работы с Big Data требуются специалисты в области аналитики данных и Data Science.

Перспективы использования Big Data в управлении трудовыми ресурсами. Интеграция мобильных технологий и Big Data открывает новые возможности для управления персоналом [4]:

- Внедрение систем интеллектуального мониторинга эффективности.
- Персонализация рабочих процессов.
- Использование предиктивной аналитики.
- Автоматизация управленческих решений.

В будущем можно ожидать более тесной интеграции мобильных платформ с системами искусственного интеллекта, что позволит создать адаптивные модели управления трудовыми ресурсами.

Заключение. Проведённый анализ показал, что мобильные технологии и концепция BYOD оказывают значительное влияние на производительность труда сотрудников.

Использование методов Big Data позволяет существенно повысить эффективность оценки и управления трудовыми процессами.

В результате можно сделать следующие выводы:

- Мобильные технологии способствуют повышению производительности труда за счёт увеличения доступности информации и гибкости работы.
- Методы Big Data позволяют объективно оценивать эффективность использования мобильных устройств и оптимизировать бизнес-процессы.
- Интеграция мобильных технологий и аналитики данных требует комплексного подхода, включающего обеспечение безопасности и развитие IT-инфраструктуры.

Список литературы

[1] Пателина, У. А. Влияние мобильных устройств на повышение производительности труда на предприятиях [Электронный ресурс] / У. А. Пателина // Цифровая трансформация. – 2025. – С. 459-460. – Режим доступа: https://elib.bsu.by/bitstream/123456789/340611/1/Цифровая%20трансформация%202025.pdf. – Дата доступа: 13.03.2026.

[2] Бычкова, О. В. Влияние цифровизации на динамику производительности труда: систематизация подходов [Электронный ресурс] / О. В. Бычкова // Новое в экономической кибернетике. – 2025. – С. 204-207. – Режим доступа: https://dongu-nec.ru/index.php/nec/article/view/463. – Дата доступа: 13.03.2026.

[3] Зиниша, О. С. Технология Big Data в бизнесе: преимущества и пути совершенствования [Электронный ресурс] / О. С. Зиниша, Д. Г. Кочеян, М. А. Мокосеева // Colloquium-journal. – 2025. – С. 155-157. – Режим доступа: https://cyberleninka.ru/article/n/tehnologiya-big-data-v-biznese-preimuschestva-i-puti-sovershenstvovaniya.pdf. – Дата доступа: 13.03.2026.

[4] Шубина, В. И. Big Data: граница инноваций, развития и конкуренции [Электронный ресурс] / В. И. Шубина, Е. Л. Кузнецова // Концепт. – 2017. – С. 2-3. – Режим доступа: https://cyberleninka.ru/article/n/big-data-granitsa-innovatsiy-razvitiya-i-konkurentsii. – Дата доступа: 13.03.2026.

[5] Савельев, А. И. Проблемы применения законодательства о персональных данных в эпоху больших данных [Электронный ресурс] / А. И. Савельев // Журнал правовых исследований. – 2015. – С. 47-52. – Режим доступа: https://cyberleninka.ru/article/n/problemy-primeneniya-zakonodatelstva-o-personalnyh-dannyh-v-epohu-bolshih-dannyh-big-data.pdf. – Дата доступа: 13.03.2026.

[6] Алибулатова, М. А. Влияние использования цифровых технологий на производительность труда и эффективность бизнеса [Электронный ресурс] / М. А. Алибулатова // Экономика и менеджмент инновационных технологий. – 2025. – Режим доступа: https://ekonomika.snauka.ru/2025/04/24725. – Дата доступа: 13.03.2026.

Авторский вклад

У. А. Пателина – постановка задачи исследования, анализ научной литературы, обобщение результатов ранее проведённого исследования по влиянию мобильных технологий на производительность труда, разработка структуры работы, написание текста статьи, формулирование выводов.

О. Н. Шкор – научное руководство исследованием, консультирование по вопросам применения методов Big Data, корректировка содержания и научное редактирование статьи.

USING BIG DATA TECHNOLOGIES AND THE BYOD CONCEPT TO ASSESS THE IMPACT OF MOBILE DEVICES ON LABOR PRODUCTIVITY IN ORGANIZATIONS

U.A. Patselina
Student of engineering and economics at the BSUIR

O.N. Shkor
Senior Lecturer at the Department of Economics, BSUIR, Master of Economics, Doctor of Philosophy in Economics

Abstract. The article examines the impact of mobile technologies and the BYOD concept on labor productivity using Big Data methods. The advantages of using mobile devices in the corporate environment, including increased flexibility and accelerated communication, are considered. Special attention is paid to the potential of big data analytics for assessing the efficiency of work processes and optimizing business decisions. It is concluded that the integration of mobile technologies and Big Data contributes to increasing the efficiency of human resource management.

Keywords: mobile technologies, BYOD, Big Data, labor productivity, data analytics, digitalization, corporate systems.

УДК 004.032.26:519.177

ГРАФОВАЯ НЕЙРОННАЯ СЕТЬ НА ОСНОВЕ АДАПТИВНОГО УРАВНЕНИЯ РЕАКЦИИ-АДВЕКЦИИ-ДИФФУЗИИ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ МОЛЕКУЛЯРНЫХ ГРАФОВ



А.С. Довгий
Студент бакалавриата факультета прикладной математики и информатики БГУ
alexanderrdouhi@gmail.com



К.В. Василевский
Доцент кафедры компьютерных технологий и систем БГУ, кандидат физико-математических наук
VasilevskyK@bsu.by

А.С. Довгий

Обучается по специальности «Прикладная информатика» на кафедре компьютерных технологий и систем факультета прикладной математики и информатики белорусского государственного университета. Область научных интересов охватывает спектральную теорию графов, нейронные дифференциальные уравнения и анализ временных рядов.

К.В. Василевский

Окончил Белорусский государственный университет. Область научных интересов связана с дифференциальными уравнениями в частных производных, математическим моделированием, прогнозированием временных рядов методами глубокого обучения и компьютерной графикой.

Аннотация. Предложена архитектура GRADE (Graph Reaction-Advection-Diffusion Equation) – графовая нейронная сеть, моделирующая непрерывную динамику представлений вершин графа как решение полного уравнения реакции-адвекции-диффузии. Архитектура расширяет CDE-GRAND четырьмя основными нововведениями: (1) выпуклая комбинация и адвекции и диффузии вместо фиксированной диффузии с модулируемой конвекцией; (2) обучаемый коэффициент внимания в диффузионном члене; (3) гейтированный реакционный член; (4) обучаемое анизотропное RBF-ядро с обучаемым вектором весов. Экспериментальная валидация на задаче многозадачной бинарной классификации молекулярных графов ogbg-moltox21 (12 ассеев токсичности, 7831 молекула) демонстрирует ROC-AUC 0,745 для GRADE, что превосходит GCN (0,734), GRAND (0,725) и CDE-GRAND (0,725).

Ключевые слова: графовые нейронные сети, уравнение реакции-адвекции-диффузии, молекулярные графы, выпуклая комбинация, обучаемое внимание.

Введение. Предсказание свойств молекул по их структуре является фундаментальной задачей вычислительной химии, фармацевтики и токсикологии. Высокопроизводительный виртуальный скрининг позволяет оценивать биологическую активность и токсичность миллионов соединений без проведения дорогостоящих лабораторных экспериментов [1]. Молекула естественным образом представляется графом $G=(V,E)$, в котором вершины V соответствуют атомам, а рёбра E –химическим связям.

Графовые нейронные сети (ГНС) стали стандартным инструментом для решения задач на графах. Базовые архитектуры –GCN [2] и GAT [3] –основаны на парадигме передачи сообщений [4], при которой признаки вершин обновляются путём агрегирования информации от соседей. Математически этот процесс эквивалентен дискретной диффузии

на графе: на каждом слое представление вершины усредняется с представлениями её соседей, что хорошо работает при гомогенности –свойстве графа, при котором смежные вершины принадлежат одному классу.

Однако многие реальные графы по своей структуре гетерогенны: смежные вершины принадлежат различным классам [5]. Это наблюдается в сетях веб-страниц, биологических сетях, молекулярных структурах, сетях вопросов-ответов. На таких графах диффузионное усреднение приводит к потере дискриминативной информации –явлению, известному как пересглаживание (over-smoothing) [6]. С увеличением числа слоёв (шагов диффузии) все представления вершин сходятся к одному и тому же вектору, что делает классификацию невозможной.

Степень гетерогенности характеризуется глобальным коэффициентом гомогенности рёбер

$$h_{edge} = \frac{|\{(u,v) \in E : y_u = y_v\}|}{|E|}, \qquad (1)$$

где y_i – метка вершины $i \in V$. Задача особенно сложна при $h_{edge} < 0{,}5$ – выраженной гетерогенности рёбер графа, когда большинство связей соединяют вершины разных классов. В случае, если имеет место значительный дисбаланс классов, предпочтительнее использовать скорректированный коэффициент глобальной гомогенности рёбер

$$h_{adj} = \frac{h_{edge} - \sum_{k=1}^{C} D_k^2/(2|E|)^2}{1 - \sum_{k=1}^{C} D_k^2/(2|E|)^2}, \qquad (2)$$

где $D_k = \sum_{i:\, y_i = k} d_i$ – суммарная степень всех вершин класса k, d_i – степень вершины i. Для скорректированного коэффициента критерием выраженной гетерогенности рёбер графа служит условие $h_{adj} < 0$.

Перспективным направлением преодоления данных ограничений является представление ГНС как решения дифференциального уравнения на графе. В работе [7] показано, что выходы глубоких нейронных сетей могут быть представлены как решения обыкновенных дифференциальных уравнений (Neural ODE), что позволяет непрерывно параметризовать глубину сети и применять адаптивные численные методы. Архитектура GRAND [8] перенесла эту идею на графы, представив эволюцию признаков вершин как диффузионный процесс. CDE-GRAND [9] расширила модель конвекционным членом для учёта направленного переноса информации, опираясь на теоретическую связь между нейронными сетями и уравнением конвекции-диффузии [10, 11].

Однако уравнение конвекции-диффузии является лишь частным случаем более общего *уравнения реакции-адвекции-диффузии* [12], широко применяемого в математической физике, химической кинетике и биологии для описания процессов с одновременным переносом, рассеиванием и локальной трансформацией вещества. Реакционный член позволяет моделировать локальные нелинейные взаимодействия, которые невозможно выразить только через диффузию и конвекцию.

Введение реакционного члена мотивировано химической природой молекулярных графов. В отличие от социальных или информационных сетей, молекулы обладают локальными электронными эффектами, которые не сводятся к взаимодействию с соседями: формальный заряд, неподелённые электронные пары, резонансные структуры определяют реакционную способность атома *независимо* от непосредственного окружения. Диффузия и конвекция – принципиально межвершинные процессы – не способны выразить такие локальные трансформации. Реакционный член заполняет этот пробел, позволяя каждой вершине изменять своё представление без участия соседей.

Целью настоящей работы является разработка архитектуры GRADE (Graph Reaction-Advection-Diffusion Equation), реализующей полное уравнение реакции-адвекции-

диффузии на графе и развивающей адаптивный подход к регуляции этих процессов. GRADE вносит четыре нововведения: (1) выпуклую комбинацию диффузии и конвекции с коэффициентами $\alpha_i(t)$ и $1-\alpha_i(t)$, позволяющую модели плавно перераспределять ресурсы между процессами; (2) обучаемые коэффициенты диффузии через механизм масштабированного скалярного внимания; (3) гейтированный реакционный член, моделирующий локальные нелинейные трансформации; (4) обучаемое анизотропное RBF-ядро в адаптивном коэффициенте. Идея в основе определения диффузионного и конвекционного членов наследуется из CDE-GRAND, что обеспечивает прямую сопоставимость результатов. Однако их аналитическое задание наряду с механизмом баланса вкладов являются нововведением. Валидация проводится на задаче предсказания токсичности молекулярных графов – практически значимой задаче для фармацевтического скрининга.

Задача классификации молекулярных графов. Рассматривается задача многомерной бинарной классификации молекулярных графов. Дан конечный набор молекул $G = \{G_s\}_{s=1}^{S}$, каждая из которых представлена графом $G_s = (V_s, E_s)$ с $N_s = |V_s|$ вершинами (атомами) и $|E_s|$ рёбрами (связями). Каждой вершине $i \in V_s$ сопоставлен вектор категориальных признаков $f_i = \left(f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(K)}\right)$, где K – число типов признаков. Для каждой молекулы задан вектор меток $y_s \in \{0, 1, \emptyset\}^M$, где M – число тестов (ассеев токсичности), а $\emptyset$ обозначает отсутствующую метку.

Структура графа описывается матрицей смежности $A \in \{0,1\}^{N\times N}$ и матрицей степеней $D = diag(d_1, \dots, d_N) \in R^{N\times N}$, где $d_i = deg_{G_s}(i)$ – степень вершины $i \in V$ в графе G_s. Графовый лапласиан определяется как $L = D - A$. Окружение вершины i обозначается $N_i = \{ j \in V_s : (i,j) \in E_s\}, N_i \subseteq V_s$, причём $|N_i| = d_i$.

Требуется построить отображение $\Psi(\cdot): G \to [0,1]^M$, предсказывающее вероятности принадлежности молекулы к каждому из M классов. Качество оценивается метрикой ROC-AUC (площадь под ROC-кривой), усреднённой по всем M тестам токсичности.

Математические основания. Классическое уравнение реакции-адвекции-диффузии в непрерывном пространстве $\Omega \subseteq R^n$ для скалярной величины $u = u(r,t)$ имеет вид

$$\frac{\partial u}{\partial t} = \underbrace{D\nabla^2 u}_{\text{диффузия}} - \underbrace{v \cdot \nabla u}_{\text{адвекция}} + \underbrace{R(u)}_{\text{реакция}}, \qquad (3)$$

где $D > 0$ – коэффициент диффузии, $\nabla^2 u = \sum_{k=1}^{n} \partial^2 u / \partial {r_k}^2$ – лапласиан, v – поле скоростей адвекции, ∇u – градиент, $R(u)$ – реакционная функция. Первый член описывает рассеивание – изотропное распространение величины u из областей высокой концентрации в области низкой. Второй – направленный перенос средой со скоростью v: в отличие от диффузии, адвекция асимметрична и может переносить информацию без усреднения. Третий – локальную трансформацию: порождение или поглощение вещества.

Уравнение (3) является основой множества моделей в естественных науках: реакции Белоусова–Жаботинского в химической кинетике, уравнения Фишера–КПП в популяционной динамике, системы Тьюринга в морфогенезе, уравнений обработки изображений. Каждый из трёх членов описывает фундаментально различный физический процесс, и их комбинация позволяет моделировать сложные динамические системы.

Частные случаи. При $R(u) = 0$ уравнение (3) вырождается в уравнение конвекции-диффузии, лежащее в основе CDE-GRAND. При дополнительном условии $v = 0$ получаем чисто диффузионное уравнение, реализуемое архитектурой GRAND. Таким образом, уравнение реакции-адвекции-диффузии является наиболее общей формой, обобщающей все существующие ODE-подходы на графах.

Дискретизация на граф. Перенос уравнения (3) на граф $G = (V, E)$ с числом вершин $N = |V|$ требует замены дифференциальных операторов их графовыми аналогами. Для матрицы представлений $X(t) \in R^{N\times d_h}$, строки которой – векторы признаков вершин

$x_i(t) \in R^{d_h}$, лапласиан заменяется нормализованным графовым лапласианом, а градиент – разностями признаков вдоль рёбер.

Диффузия на графе в матричной форме допускает следующую формулировку:

$$\frac{\partial X(t)}{\partial t} = \big(A\big(X(t)\big) - I\big)X(t), \qquad (4)$$

где $A \in R^{N\times N}$ – обучаемая матрица внимания, $I \in R^{N\times N}$ – единичная матрица порядка N. Матрица внимания определяет силу связей между вершинами через функцию сходства $a(\cdot,\cdot)$: $R^{d_h} \times R^{d_h} \to R$. Элементы матрицы (коэффициенты внимания) $a_{ij} = a\left(x_i(t), x_j(t)\right)$ допускают различные аналитические определения. Например, в оригинальной статье про модель GRAND [8] используется механизм внимания GAT, подробно описанный в [3]. Для отдельной вершины $i \in V$ уравнение принимает вид:

$$\frac{\partial x_i(t)}{\partial t} = \sum_j \quad a_{ij}\left(x_j(t) - x_i(t)\right), \qquad (5)$$

Адвекция (конвекция) на графе моделирует направленный перенос информации. В непрерывном случае (3) адвекция описывается членом $-v \cdot \nabla u$, где v – поле скоростей. На графе аналогом градиента ∇u в вершине $i \in V$ вдоль ребра $(i,j) \in E$ служит разность представлений вершин $x_j(t) - x_i(t)$. Поле скоростей параметризуется обучаемой функцией $V_{ij}(t) \in R^{d_h}$, зависящей от этой разности. Поэлементное произведение (произведение Адамара) $V_{ij}(t) \odot x_j(t)$обеспечивает покоординатную модуляцию: каждая компонента вектора $x_j(t)$ масштабируется соответствующей компонентой скорости.

Реакционный член на графе является наиболее простым в плане дискретизации, поскольку он локален: $R\big(x_i(t)\big)$ зависит только от представления вершины $i \in V$ и не требует агрегирования информации от соседей. Это делает реакцию вычислительно эффективной и параллелизуемой.

Описание модели GRADE. Предлагаемая архитектура развивает существующие ODE-подходы, расширяя уравнение конвекции-диффузии до полного уравнения реакции-адвекции-диффузии на графе. Архитектура состоит из четырёх последовательных компонент: кодировщика атомов, блока интегрирования, графового пулинга и выходного классификатора.

Входной кодировщик. Динамика дифференциального уравнения требует, чтобы размерность состояния $x_i(t) \in R^{d_h}$ оставалась постоянной на всём интервале интегрирования $t \in [0,T]$. Поэтому исходные признаки вершин, имеющие произвольную размерность и природу, должны быть отображены в скрытое пространство фиксированной размерности d_h.

В оригинальной статье про модель CDE-GRAND [9] входной кодировщик представляет собой многослойный перцептрон (MLP), который отображает исходные признаки вершин графа $x_i \in R^d$ размерности d в скрытое пространство размерности d_h: $x_i(0) = MLP_{enc}(x_i) \in R^{d_h}$. Кодировщик состоит из двух линейных слоёв с функцией активации ReLU и регуляризацией Dropout между ними. В задаче классификации молекулярных графов ситуация принципиально иная: признаки атомов являются категориальными. Каждый атом описывается $K = 9$ дискретными характеристиками: атомный номер ($C_1 = 128$ категорий), хиральность ($C_2 = 16$), степень ($C_3 = 16$), формальный заряд ($C_4 = 16$), число водородов ($C_5 = 16$), число радикальных электронов ($C_6 = 8$), гибридизация ($C_7 = 8$), ароматичность ($C_8 = 8$), принадлежность к кольцу ($C_9 = 8$). Применение $MLP_{enc}(\cdot)$ к категориальным данным неэффективно, поскольку one-hot

кодирование порождает разреженные вектора высокой размерности ($d_h = \sum_k \quad C_k = 224$), большинство компонент которых равно нулю.

В задаче классификации графов молекул $MLP_{enc}(\cdot)$ заменяется *кодировщиком атомов* на основе таблиц вложений (embeddings). Для каждого типа признака $k \in \{1, \dots, K\}$ определяется обучаемая таблица вложений $Emb_k(\cdot): \{0, \dots, C_k - 1\} \to R^{d_h}$, представляющая собой матрицу $E_k \in R^{C_k \times d_h}$, строка которой с индексом $f_i^{(k)}$ извлекается как $Emb_k\left(f_i^{(k)}\right) = [E_k]_{f_i^{(k)}}$. Начальное представление вершины $i \in V$:

$$x_i^{in} \coloneqq \sum_{k=1}^{K} \quad Emb_k\left(f_i^{(k)}\right) = \sum_{k=1}^{K} \quad [E_k]_{f_i^{(k)}} \in R^{d_h}. \qquad (6)$$

Суммирование вместо конкатенации сохраняет размерность d_h. Каждая таблица представлений обучается совместно с остальными параметрами модели, что позволяет автоматически выучить информативные представления для каждой категории каждого признака.

Начальное условие ОДУ. Результат кодирования формирует матрицу начального условия $X^{in} = \{x_i^{in}\}_{i=1}^{N} \in R^{N \times d_h}$, строка i которой – вектор x_i^{in}. Дифференциальное уравнение интегрируется как задача Коши с начальным условием:

$$X(0) = X^{in}, \qquad X(0) = \big(x_i(0)\big)_{i=1}^{N} \in R^{N \times d_h}. \qquad (7)$$

Блок интегрирования. Дифференциальное уравнение, лежащее в основе модели GRADE для отдельной вершины $i \in V$:

$$\frac{\partial x_i(t)}{\partial t} = \alpha_i(t) D_i(t)_{\underbrace{\quad}_{\text{диффузия}}} + \beta\big(1 - \alpha_i(t)\big) C_i(t)_{\underbrace{\quad}_{\text{адвекция}}} + g_i(t) \odot R_i(t)_{\underbrace{\quad}_{\text{реакция}}}, \quad (8)$$

где тройка $\big(D_i(t), C_i(t), R_i(t)\big)$ отвечает соответственно за вклад диффузии, адвекции (конвекции) и реакции. Ключевым структурным отличием адвекционно-диффузионной части от CDE-GRAND является наличие эволюционирующих во времени весов у этих слагаемых, причем адаптивный коэффициент $\alpha_i(t) \in [0,1]$ и его дополнение $1 - \alpha_i(t)$ образуют *выпуклую комбинацию* диффузионного и адвекционного членов. В CDE-GRAND оба эти процесса всегда активны с полным весом; в GRADE же при $\alpha_i(t) \to 0$ диффузия подавляется, а адвекция усиливается, и наоборот. Это позволяет модели полностью переключаться между режимами в зависимости от локальной структуры. Реакционный член является принципиально новым. Ниже описаны все компоненты.

1) Диффузия. В GRADE диффузионный член реализован через обучаемый механизм масштабированного скалярного внимания [13]. Для каждой вершины $i \in V$ вычисляются вектор запроса и вектор ключа:

$$q_i(t) = W_Q x_i(t) \in R^{d_h}, \qquad k_i(t) = W_K x_i(t) \in R^{d_h}, \qquad (9)$$

где $W_Q, W_K \in R^{d_h \times d_h}$ – обучаемые матрицы проекций. Коэффициенты внимания вычисляются по рёбрам графа как выход функции $softmax(z_k) = \frac{exp\,(z_k)}{\sum_k \quad exp(z_k)}$:

$$a_{ij}(t) = \frac{exp\ exp\left(\dfrac{q_i(t)^{\intercal} k_j(t)}{\sqrt{d_h}}\right)}{\sum_{l \in N_i} \quad exp\ exp\left(\dfrac{q_i(t)^{\intercal} k_l(t)}{\sqrt{d_h}}\right)} \in R. \qquad (10)$$

Масштабирование через $\left(\sqrt{d_h}\right)^{-1}$ предотвращает попадание скалярных произведений в область насыщения softmax, стабилизируя градиенты. Свойство нормированности $\sum_j \ a_{ij}(t) = 1$ обеспечивает, что диффузионный член является взвешенным усреднением, не приводящим к неограниченному росту нормы. Матрица внимания $A \in R^{N\times N}$ определяется элементами $[A(t)]_{ij} = a_{ij}(t)$ при $(i,j) \in E$ и $[A(t)]_{ij} = 0$ иначе. Диффузионный вклад в матричной и векторной форме:

$$D(t) = (D_i(t))_{i=1}^N = (A(t) - I)X(t) \in R^{N\times d_h},$$

$$D_i(t) = \sum_j \ a_{ij}\left(x_j(t) - x_i(t)\right) \in R^{d_h}. \qquad (11)$$

Обучаемость коэффициентов позволяет модели адаптировать диффузию к признаковому содержанию: внимание может усилить влияние информативных соседей (например, атомов с электроноакцепторными группами) и ослабить нерелевантных.

2) Адвекция. Адвекционный член наследуется из CDE-GRAND. Единственным изменением является замена функции активации с ReLU на гиперболический тангенс. Вектор скорости потока от вершины j к вершине i:

$$V_{ij}(t) = tanh\left(W_V\left(x_j(t) - x_i(t)\right)\right) \in R^{d_h}, \qquad (12)$$

где $W_V \in R^{d_h\times d_h}$ – обучаемая матрица весов, а поэлементное применение гиперболического тангенса $tanh(\cdot)$: $R^{d_h} \to R^{d_h}$ обеспечивает ограниченность компонент скорости: $\left|\left[V_{ij}(t)\right]_k\right| \le 1,\ k = \underline{1, d_h}$. Соответственно, предотвращается бесконтрольный рост нормы вектора скорости, приводящий к нестабильному поведению модели. Матрица адвекции $C(t) = (C_i(t))_{i=1}^N \in R^{N\times d_h}$, строки которой отвечают за адвекционный вклад для вершины $i \in V$ и определяются как

$$C_i(t) = \sum_{j\in N_i} \ V_{ij}(t) \odot x_j(t) \in R^{d_h}. \qquad (13)$$

3) Адаптивный коэффициент. Адаптивный коэффициент $\alpha_i(t)$ модулирует адвекцию и диффузию на основе локальной оценки гомогенности ребер:

$$h_i^{local}(t) = \frac{1}{|N_i|}\sum_{j\in N_i} \ exp\ exp\ \left(-\|w \odot \left(x_j(t) - x_i(t)\right)\|^2\right) \in [0,1],$$

$$\alpha_i(t) = 1 - h_i^{local}(t) \in [0,1], \qquad (14)$$

где $w \in R^{d_h}$ – обучаемая вектор весов. В качестве нормы $\|\cdot\|$ используется l_2-норма, что преобразует экспоненту в гауссово ядро (RBF-ядро). Поэлементное произведение под знаком нормы обеспечивает свойство анизотропности ядер – свойстве, при котором признаковые координаты дифференцируются по значимости в процессе обучения. При инициализации каждая компонента вектора w полагается равной $w_k = \sqrt{\frac{\gamma_0}{d_h}}, k = \underline{1, d_h}$, $\gamma_0 = 0{,}5$ – эмпирически полученный результат. Далее в процессе обучения компоненты w_k принимают различные значения, что позволяет модели автоматически определять, какие измерения пространства представлений наиболее информативны для оценки гомогенности.

Физический смысл выпуклой комбинации: при $h_i^{local}(t) \approx 1$ (гомогенное окружение) $\alpha_i(t) \approx 0$ – вес диффузии стремится к нулю, вес адвекции к единице – доминирует адвекция. При $h_i^{local}(t) \approx 0$ (гетерогенное окружение) $\alpha_i(t) \approx 0$ – доминирует диффузия, адвекция подавлена. Пара коэффициентов $\left(\alpha_i(t), 1 - \alpha_i(t)\right)$, составляющая выпуклую комбинацию, гарантирует, что суммарный вклад диффузии и адвекции

ограничен в любой момент времени $t \in [0, T]$, что способствует устойчивости интегрирования. В отличие от CDE-GRAND, где адвекция и диффузия всегда имеют фиксированный вес 1, GRADE позволяет модели плавно перераспределять ресурсы между обоими процессами.

4) Реакция. Ключевым нововведением GRADE является реакционный член, вносящий локальную нелинейную динамику непосредственно в дифференциальное уравнение. Диффузия и адвекция – процессы межвершинного взаимодействия: представление вершины i изменяется только в зависимости от соседей N_i. Реакционный член позволяет вершине трансформировать своё представление *независимо* от окружения.

Реакционная функция – двухслойный перцептрон с гладкой нелинейностью:

$$R(x) = W_{R_2} GELU\left(W_{R_1} x + b_{R_1}\right) + b_{R_2} \in R^{d_h}, \qquad (15)$$

где $W_{R_1}, W_{R_2} \in R^{d_h \times d_h}$, $b_{R_1}, b_{R_2} \in R^{d_h}$ – обучаемые параметры. Поэлементная функция активации $GELU(z) = z \cdot \Phi(z)$ [14], где $\Phi(\cdot): R \to R$ – стандартная нормальная функция распределения, обеспечивает гладкость, критичную для корректности дифференцирования правой части уравнения (8).

Для контроля вклада реакции для каждой вершины $i \in V$ вводится вектор гейтирования:

$$g_i(t) = \sigma(W_G x_i(t) + b_G) \in [0, 1]^{d_h}, \qquad (16)$$

где $W_G \in R^{d_h \times d_h}$, $b_G \in R^{d_h}$ – обучаемые параметры, в качестве функции активации используется $\sigma(z) = \frac{1}{(1+e^{-z})}$ – поэлементная логистическая функция. Реакционный вклад:

$$R_i(t) = g_i(t) \odot R\left(x_i(t)\right) \in R^{d_h}. \qquad (17)$$

Гейтирование $g_i(t) \in [0, 1]^{d_h}$ выполняет две функции. Во-первых, оно предотвращает неконтролируемый рост нелинейного члена при длительном интегрировании ($S = \frac{T}{\tau}$ итераций, τ – шаг интегрирования), ограничивая каждую компоненту вектора реакции: $|[R_i(t)]_k| \le \|R\left(x_i(t)\right)\|_\infty$, $k = \underline{1, d_h}$. Во-вторых, оно позволяет модели поэлементно выбирать, какие компоненты признаков подвергаются реакции, а какие пропускаются. Векторы-строки $R_i(t)$ составляют гейтированную матрицу реакционных вкладов для каждой вершины графа $R(t) = \left(R_i(t)\right)_{i=1}^{N} \in R^{N \times d_h}$.

В контексте молекулярных графов реакционный член моделирует локальные электронные эффекты: атом может изменять своё представление независимо от соседей при наличии неподелённых электронных пар, формального заряда или резонансных структур. Такие эффекты невозможно выразить через диффузию (агрегирование соседей) и адвекцию (направленный перенос от соседей), поскольку оба механизма являются межвершинными.

5) Итоговое уравнение в матричной форме. Введём диагональную матрицу адаптивных коэффициентов $\Lambda(t) = diag(\alpha_1(t), \dots, \alpha_N(t)) \in R^{N \times N}$. Объединяя три члена, получаем матричное дифференциальное уравнение для GRADE:

$$\frac{\partial X(t)}{\partial t} = \Lambda(t) D(t) + \beta\left(I - \Lambda(t)\right) C(t) + R(t). \qquad (18)$$

6) Численное интегрирование и стабилизация. Решение уравнения (19) вычисляется методом Рунге–Кутты 4-го порядка (RK4) с фиксированным шагом τ на $t \in [0, T]$. Метод RK4 обеспечивает точность порядка $O(\tau^4)$, что существенно превосходит метод Эйлера первого порядка точности и позволяет использовать более крупный шаг τ при сохранении точности. Ценой является четырёхкратное вычисление правой части на каждом шаге. После каждой итерации интегрирования применяется нормализация слоя (Layer Normalization [15]). Для вектора $x \in R^{d_h}$:

$$x \leftarrow LayerNorm(x) = \frac{x-\mu}{\sqrt{\sigma^2+\varepsilon}} \odot \gamma + \delta, \qquad \mu = \frac{1}{d_h}\sum_{k=1}^{d_h} x_p,$$

$$\sigma^2 = \frac{1}{d_h}\sum_{k=1}^{d_h} \left(x_p - \mu\right)^2, \qquad (19)$$

где $\gamma, \delta \in R^{d_h}$ – обучаемые параметры масштаба и сдвига, $\varepsilon = 10^{-5}$. Нормализация предотвращает экспоненциальный рост или затухание признаков при длительном интегрировании. Дополнительно применяется покоординатное ограничение на представления вершин $[x_i(t)]_k \leftarrow max(-H, min(H, [x_i(t)]_k)),\ H = 10,\ k = \underline{1, d_h}$ для предотвращения числовых переполнений.

Графовый пулинг. Для перехода от вершинных представлений к представлению графа (молекулы) используется усреднение:

$$x_G = \frac{1}{N}\sum_{i=1}^{N} x_i(T) \in R^{d_h}. \qquad (20)$$

Выходной классификатор и функция потерь. Представление молекулы преобразуется в вектор логитов двухслойным перцептроном:

$$\hat{y}_G = W_{dec,2} SiLU\left(W_{dec,1} x_G + b_{dec,1}\right) + b_{dec,2} \in R^M, \qquad (21)$$

где $W_{dec,1} \in R^{d_h \times d_h}$, $W_{dec,2} \in R^{M \times d_h}$, $b_{dec,1} \in R^{d_h}$, $b_{dec,2} \in R^M$– обучаемые параметры. В качестве функции активации выступает $SiLU(z) = z \cdot \sigma(z) = \frac{z}{(1+e^{-z})}$ – сглаженная дифференцируемая версия $ReLU(z) = (0, z)$. Между слоями применяется Dropout ($p = 0{,}5$) для регуляризации. Функция потерь – бинарная кросс-энтропия с логитами по валидным меткам. Оптимизация выполняется алгоритмом Adam [16] с ограничением нормы градиента.

Экспериментальные данные. Валидация проведена на датасете ogbg-moltox21 [17] – стандартном бенчмарке из коллекции Open Graph Benchmark (OGB) для предсказания свойств молекулярных графов. Датасет содержит 7831 молекулу, каждая из которых представлена графом, где вершины – атомы, рёбра – химические связи. Графы этих молекул составляют описанное ранее множество $G = \{G_s\}_{s=1}^{S}$ мощности $|G| = 7831$ – то есть, $S = 7831$ уникальных молекул.

Задачей является одновременная бинарная классификация по $M = 12$ ассеям токсичности. Ассеи охватывают два класса биологических мишеней: (1) рецепторы ядерных гормонов – андрогенный рецептор (AR), эстрогенный рецептор (ER), ароматаза, рецептор прегнана X (PXR), рецептор ретиноевой кислоты (RAR) и другие; (b) пути стрессового ответа – путь p53, путь элемента стрессового ответа (HSE), митохондриальный мембранный потенциал (MMP), антиоксидантный ответ (ARE), путь активации каспазы (ATAD5). Каждая молекула может иметь отсутствующие метки для отдельных ассеев, которые маскируются при вычислении функции потерь.

Датасет ogbg-moltox21 особенно подходит для оценки ODE-моделей по нескольким причинам. Во-первых, многозадачная природа (12 ассеев) позволяет оценить обобщающую способность модели: одно представление молекулы должно быть информативным для различных биологических мишеней. Во-вторых, размер датасета (7831 молекула) достаточен для обучения моделей с умеренным числом параметров, но недостаточно велик для подавления преимущества более выразительных архитектур с регуляризацией. В-третьих, молекулярные графы обладают признаковой гетерогенностью: соседние атомы, как правило, имеют разные типы и свойства, что ограничивает применимость многих классических подходов.

Используется стандартное разбиение OGB: 6264/783/784 (обучение/валидация/тест), обеспечивающее воспроизводимость сравнений. Характеристики датасета приведены в таблице 1.

Таблица 1. Характеристики датасета ogbg-moltox21

Характеристика	Значение
Число молекул S	7831
Число тестов (ассеев) M	12
Среднее число атомов $\underline{N}$	18,6
Среднее число связей $\underline{\lvert E \rvert}$	19,3
Число типов атомных признаков K	9
Метрика	ROC-AUC
Разбиение	6264 / 783 / 784

Постановка эксперимента. Сравниваются четыре модели: GCN, GRAND, CDE-GRAND, GRADE. Общие гиперпараметры моделей перечислены в таблице 2.

Таблица 2. Гиперпараметры экспериментов

Параметр	Значение
Скрытая размерность d_h	128
Время интегрирования T	2,0
Шаг интегрирования τ	0,2
Число шагов $S = T/\tau$	10
Инициализация γ_0	0,5
Число эпох	100
Скорость обучения η	0,001
Размер батча	64
Dropout	0,5
Ранняя остановка (эпох)	15
Число запусков	3

Каждый эксперимент повторяется 3 раза с различными случайными инициализациями. Модель с наилучшей валидационной метрикой ROC-AUC сохраняется. Приводятся среднее и стандартное отклонение ROC-AUC на тестовом множестве. Все эксперименты проводились в онлайн-среде Google Colab с использованием языка Python 3 и аппаратного ускорителя – графического процессора NVIDIA Tesla T4. NVIDIA Tesla T4 – это энергоэффективный (75 Вт) ускоритель на архитектуре Turing, предназначенный для ИИ-инференса, видеообработки и графики в ЦОД. Карта оснащена 16 ГБ GDDR6 с высокой пропускной способностью, 2560 ядрами CUDA и 320 тензорными ядрами, обеспечивая высокую производительность в вычислениях FP16, INT8 и INT4 при низком потреблении энергии.

Результаты и анализ. Результаты сравнительного эксперимента представлены в таблице 3.

Таблица 3. ROC-AUC на ogbg-moltox21 (среднее ± стд. откл.)

Модель	ROC-AUC
GCN	0,734 ± 0,005
GRAND	0,725 ± 0,001
CDE-GRAND	0,728 ± 0,002
GRADE	**0,745 ± 0,005**

GRADE (0,745) превосходит все базовые модели, демонстрируя наибольший прирост (2 п.п.) в сравнении с GRAND. Этот прирост обеспечивается совокупностью четырёх архитектурных решений. Во-первых, выпуклая комбинация диффузии и конвекции с адаптивными весами позволяет модели для каждой вершины выбирать оптимальный режим: в гомогенном окружении (атомы с похожими представлениями) увеличивается вклад конвекции, а в гетерогенном – диффузии. Во-вторых, обучаемые коэффициенты внимания в диффузионном члене заменяют фиксированную спектральную нормализацию, позволяя модели усиливать влияние информативных соседей – например, атомов с электроноакцепторными или электронодонорными группами. В-третьих, гейтированный реакционный член вводит принципиально новый механизм – локальную нелинейную трансформацию, независимую от окружения. В контексте молекулярных графов он моделирует внутриатомные эффекты: формальный заряд, неподелённые электронные пары, резонансные структуры – всё то, что определяет реакционную способность атома, но не выражается через диффузию или конвекцию, являющихся по своей природе межвершинными процессами. В-четвёртых, обучаемое анизотропное ядро автоматически определяет, какие компоненты признаков наиболее значимы для оценки локальной гомогенности, устраняя необходимость ручного подбора масштаба.

Заключение. В настоящей работе рассмотрена задача предсказания токсичности молекулярных соединений – практически значимая задача для фармацевтической разработки и высокопроизводительного виртуального скрининга. Предложена архитектура GRADE, представляющая динамику признаков атомов молекулы как решение полного уравнения реакции-адвекции-диффузии на графе. Модель обобщает существующие ODE-подходы на графах, последовательно расширяя чисто диффузионную формулировку адвекцией, адаптивным балансом и реакцией. Экспериментальная валидация на стандартном бенчмарке ogbg-moltox21, содержащем 7831 молекулу и 12 ассеев токсичности, подтверждает эффективность подхода. GRADE достигает наивысшего значения ROC-AUC среди всех рассмотренных моделей, превосходя как дискретную архитектуру GCN, так и непрерывные ODE-модели GRAND и CDE-GRAND. Анализ результатов показывает, что на компактных молекулярных графах чисто диффузионные и диффузионно-конвекционные модели уступают дискретной архитектуре, и только полная модель реакции-адвекции-диффузии способна раскрыть потенциал непрерывного подхода.

Направления дальнейших исследований включают расширение модели рёберными признаками, отражающими тип химической связи и стереохимию, масштабирование на более крупные бенчмарки молекулярных свойств, разработку адаптивного шага интегрирования для повышения точности при сохранении вычислительной эффективности, а также теоретический анализ влияния реакционного члена на устойчивость представлений при длительном интегрировании.

Список литературы

[1] Wu Z. et al. MoleculeNet: a benchmark for molecular machine learning // Chemical Science. – 2018. – Vol. 9, № 2. – PP. 513–530. – DOI: 10.1039/C7SC02664A.

[2] Kipf T.N., Welling M. Semi-supervised classification with graph convolutional networks // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1609.02907.

[3] Veličković P., Cucurull G., Casanova A. et al. Graph attention networks // arXiv.org. – 2018. – DOI: 10.48550/arXiv.1710.10903.

[4] Gilmer J., Schoenholz S.S., Riley P.F. et al. Neural message passing for quantum chemistry // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1704.01212.

[5] Zhu J., Yan Y., Zhao L. et al. Beyond homophily in graph neural networks: Current limitations and effective designs // arXiv.org. – 2020. – DOI: 10.48550/arXiv.2006.11468.

[6] Li Q., Han Z., Wu X.-M. Deeper insights into graph convolutional networks for semi-supervised learning // arXiv.org. – 2018. – DOI: 10.48550/arXiv.1801.076 06.

[7] Chen R.T.Q., Rubanova Y., Bettencourt J., Duvenaud D.K. Neural ordinary differential equations // arXiv.org. – 2018. – DOI: 10.48550/arXiv.1806.07366.

[8] Chamberlain B.P., Rowbottom J., Gorinova M.I. et al. GRAND: Graph neural diffusion // arXiv.org. – 2021. – DOI: 10.48550/arXiv.2106.10934.

[9] Zhao K., Kang Q., Song Y.et al. Graph neural convection-diffusion with heterophily // arXiv.org.– 2023. – DOI: 10.48550/arXiv.2305.16780.

[10] Alvarez L., Guichard F., Lions P.-L., Morel J.-M. Axioms and fundamental equations of image processing // Archive for Rational Mechanics and Analysis. – 1993. – Vol. 123, № 3. – P. 199–257. – DOI: 10.1007/BF00375127.

[11] Wang T., Dou Z., Bao C., Shi Z. Diffusion mechanism in residual neural network: Theory and applications // arXiv.org. – 2024. – DOI: 10.48550/arXiv.2105.03155.

[12] Hu W., Fey M., Zitnik M. et al. Open graph benchmark: Datasets for machine learning on graphs // arXiv.org. – 2020. – DOI: 10.48550/arXiv.2005.00687.

[13] Vaswani A., Shazeer N., Parmar N. et al. Attention is all you need // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1706.03762.

[14] Hendrycks D., Gimpel K. Gaussian error linear units (GELUs) // arXiv.org. – 2016. – DOI: 10.48550/arXiv.1606.08415.

[15] Ba J.L., Kiros J.R., Hinton G.E. Layer normalization // arXiv.org. – 2016. – DOI: 10.48550/arXiv.1607.06450.

[16] Kingma D.P., Ba J. Adam: A method for stochastic optimization // arXiv.org. – 2017. – DOI: 10.48550/arXiv.1412.6980.

[17] Hu W., Fey M., Zitnik M. et al. Open graph benchmark: Datasets for machine learning on graphs// arXiv.org. – 2020. – DOI: 10.48550/arXiv.2005.00687.

Авторский вклад

Довгий Александр Сергеевич – разработка математического аппарата модели, разработка архитектуры GRADE, реализация модели и проведение вычислительных экспериментов, анализ результатов.

Василевский Константин Викторович – постановка задачи исследования, разработка математического аппарата модели, описание методологии экспериментальной валидации, разработка направлений для дальнейших исследований.

GRAPH NEURAL NETWORK BASED ON AN ADAPTIVE REACTION-ADVECTION-DIFFUSION EQUATION FOR MOLECULAR GRAPH CLASSIFICATION

A.S. Douhi
Undergraduate student, Faculty of Applied Mathematics and Computer Science, Belarusian State University

K.V. Vasilevsky
Associate Professor, Department of Computer Technologies and Systems, Belarusian State University, PhD in Physics and Mathematics

Annotation. The GRADE (Graph Reaction-Advection-Diffusion Equation) architecture is proposed – a graph neural network that models the continuous dynamics of node representations in a graph as the solution of the full reaction-advection-diffusion equation. The architecture extends CDE-GRAND with four key innovations: (1) a convex combination of both advection and diffusion instead of fixed diffusion with modulated convection; (2) a learnable attention coefficient in the diffusion term; (3) a gated reaction term; (4) a learnable anisotropic RBF kernel with a learnable weight vector. Experimental validation on the multitask binary classification problem for molecular graphs ogbg-moltox21 (12 toxicity assays, 7,831 molecules) demonstrates a ROC-AUC of 0.745 for GRADE, which surpasses GCN (0.734), GRAND (0.725), and CDE-GRAND (0.725).

Keywords: graph neural networks, reaction-advection-diffusion equation, molecular graphs, convex combination, learnable attention.

УДК 004.8:616.12-073.97:004.94

РАЗРАБОТКА И ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА МОДЕЛИ УЧЁТА ФИЗИОЛОГИЧЕСКОГО СОСТОЯНИЯ ПОЛЬЗОВАТЕЛЯ НА ОСНОВЕ ЭКГ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ BIG DATA И ПРОДВИНУТОЙ АНАЛИТИКИ В HCI-СИСТЕМАХ



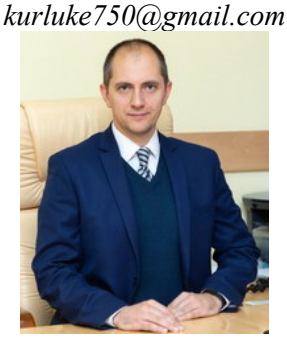
Н.А. Ларченко
Студент факультета компьютерного проектирования кафедры электронной техники и технологий БГУИР
nikitadeve@gmail.com



Е.А. Курлюк
Студент факультета компьютерного проектирования кафедры электронной техники и технологий БГУИР
kurluke750@gmail.com



А.Ч. Ганиев
Студент компьютерных систем и сетей кафедры информатики БГУИР
ganievabubakr.64@gmail.com



М.В. Давыдов
Первый проректор, доцент, кандидат технических наук
davydov-mv@bsuir.by

Н.А. Ларченко
Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки медицинских данных, построением нейросетевых систем анализа физиологических сигналов, разработкой масштабируемых информационно-компьютерных систем и сервисов.

Е.А. Курлюк
Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки медицинских данных, созданием программных модулей извлечения и структурирования информации из медицинских документов, разработкой масштабируемых информационно-компьютерных систем и сервисов.

А.Ч. Ганиев
Студент Белорусского государственного университета информатики и радиоэлектроники. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки данных, построением и исследованием нейросетевых систем, разработкой программного обеспечения и масштабируемых информационно-компьютерных систем и сервисов.

М.В. Давыдов

Первый проректор Белорусского государственного университета информатики и радиоэлектроники, доцент, кандидат технических наук. Область научных интересов связана с разработкой интеллектуальных методов и алгоритмов обработки данных, цифровой обработкой сигналов, электростимуляцией и транскраниальной магнитной стимуляцией.

Аннотация. В работе рассматриваются возможности использования сигналов ЭКГ в сочетании с методами Big Data и продвинутой аналитики для учета физиологического состояния пользователя в системах HCI. Рассматриваются особенности ЭКГ как источника данных о состоянии пользователя, включая возможность получения характеристик сердечного ритма и его вариабельности. Показано, что обработка ЭКГ в условиях HCI-систем связана с необходимостью анализа больших объемов временных данных. Предложен подход к построению адаптивной HCI-системы, в которой результаты анализа ЭКГ интегрируются в логику взаимодействия с пользователем.

Ключевые слова: электрокардиография, HCI, физиологическое состояние пользователя, вариабельность сердечного ритма, Big Data, продвинутая аналитика, машинное обучение, анализ временных рядов, адаптивные интерфейсы, биосигналы.

Введение. Эффективность выполнения профессиональных задач в организациях и на предприятиях в значительной степени связана с физиологическим и психоэмоциональным состоянием работников. При этом существующие решения, направленные на мониторинг состояния человека, как правило, ограничены по времени наблюдения, требуют специализированного оборудования либо не интегрированы в повседневные процессы взаимодействия пользователя с цифровыми системами. Аналогичная проблема наблюдается и в повседневной деятельности, где учет физиологического состояния пользователя в большинстве случаев не осуществляется [1].

Одним из доступных источников информации о состоянии организма являются физиологические сигналы, в частности электрокардиограмма (ЭКГ), позволяющая получать данные о функционировании сердечно-сосудистой системы. Анализ параметров ЭКГ, таких как частота сердечных сокращений и вариабельность сердечного ритма, может использоваться для оценки функционального состояния пользователя, однако получаемые результаты носят вероятностный характер и зависят от контекста наблюдения [2].

В последние годы технологии Big Data и методы продвинутой аналитики получили широкое распространение в различных областях, включая медицину, промышленность и информационные технологии [3]. Их применение позволяет обрабатывать значительные объемы разнородных временных данных, учитывать индивидуальные особенности пользователей и строить модели, ориентированные на анализ физиологических сигналов в динамике. В частности, обработка потоков ЭКГ с использованием методов машинного обучения и статистического анализа открывает возможность получения оценок состояния пользователя в режиме, близком к реальному времени.

В связи с этим актуальной является задача разработки подхода к построению систем человек–компьютер, в которых учет физиологического состояния пользователя, полученного на основе анализа ЭКГ и методов Big Data, может использоваться в качестве дополнительного контекста взаимодействия. Целью данной работы является исследование возможностей применения методов Big Data и продвинутой аналитики для учета физиологического состояния пользователя на основе ЭКГ в HCI-системах, а также формирование общего подхода к интеграции таких данных в процессы взаимодействия человек–компьютер.

Анализ предметной области. Современные HCI-системы ориентированы на повышение эффективности взаимодействия пользователя с программными и аппаратными средствами за счёт оптимизации интерфейсов и логики работы приложений. При этом в большинстве существующих решений взаимодействие строится на основе явных действий пользователя, как ввод данных и выполнение команд без учёта его текущего физиологического состояния. Отсутствие информации о внутреннем состоянии пользователя может приводить к снижению эффективности взаимодействия, особенно в условиях повышенной когнитивной нагрузки, утомления или стрессовых факторов [4]. В таких ситуациях возрастает риск ошибок,

увеличивается времени выполнения задачи, возрастает вероятность некорректного использования системы. В то же время в ряде современных решений уже предпринимаются попытки учитывать физиологические данные пользователя. Например, платформа Affective Health от компании Biotricity использует данные ЭКГ и вариабельности сердечного ритма для визуализации уровня стресса пользователя, а гарнитуры NeuroSky применяются для оценки внимания и адаптации интерфейсов обучающих и игровых приложений.

В связи с этим в последние годы развивается направление адаптивных HCI-систем, в которых учитываются дополнительные контекстные данные о пользователе. К таким данным относятся поведенческие характеристики, параметры активности, а также физиологические сигналы, позволяющие косвенно оценивать состояние пользователя [5]. Практическая реализация подобных подходов наблюдается в системах, интегрирующих биосигналы в пользовательские сценарии, включая исследовательские и коммерческие решения, ориентированные на адаптацию интерфейсов и поддержку пользователя в зависимости от его состояния.

Физиологические сигналы человека в целом рассматриваются как один из источников информации о его функциональном состоянии. К такого рода сигналам относят электрокардиограммы, электроэнцефалограммы, электромиограммы, показатели кожно-гальванической реакции и дыхания. Их применение в задачах человек-компьютерного взаимодействия связано с возможностью получения объективных данных без необходимости явного участия пользователя. В частности, в современных устройствах и системах используются комбинированные подходы, включающие несколько типов сигналов, например ЭКГ и ЭЭГ, что реализовано в ряде решений для оценки когнитивного и эмоционального состояния пользователя.

ЭКГ занимает особое место среди физиологических сигналов благодаря доступности измерения и информативности. На основе анализа ЭКГ можно получить такие характеристики, как частоту сердечных сокращений и вариабельность сердечного ритма, отражающие работу автономной нервной системы [6]. Эти параметры дают возможность косвенной оценки функционального состояния пользователя, включая признаки утомления, напряжения или повышенной нагрузки. При этом важным является тот факт, что интерпретация физиологических сигналов носит вероятностный характер и зависит от множества факторов, включая индивидуальные особенности пользователя, условия измерения и внешние воздействия, из чего следует, что использование ЭКГ в HCI-системах требует применение методов обработки сигналов и аналитических моделей, способных учитывать вариативность данных. Практическое применение анализа ЭКГ демонстрируют такие решения, как KardiaMobile (AliveCor), ориентированное на массового пользователя и позволяющее получать ЭКГ и автоматические заключения через мобильное приложение, а также платформа Cardiomatics, предназначенная для анализа длительных записей и интеграции в медицинские системы.

Анализ физиологических сигналов, в том числе ЭКГ, характеризуется необходимостью обработки больших объемов данных, представленных в виде временных рядов. В рамках систем человек–компьютер это может включать непрерывный сбор данных от большого числа пользователей, их хранение, обработку и последующую интерпретацию. Технологии Big Data позволяют решать задачи, связанные с обработкой разнородных и быстро изменяющихся данных, обеспечивая масштабируемость и устойчивость систем. Ключевыми характеристиками таких данных являются объем, скорость поступления и разнообразие форматов. Использование данных подходов позволяет организовать обработку потоков ЭКГ в режиме, близком к реальному времени, а также учитывать межпользовательскую вариативность. Такие требования к обработке данных реализуются в современных облачных и веб-платформах анализа ЭКГ, включая системы, поддерживающие обработку многоканальных записей и интеграцию с медицинскими информационными системами.

Методы продвинутой аналитики, включая машинное обучение и статистические методы анализа, применяются для извлечения информативных признаков из сигналов и построения моделей оценки состояния пользователя. Такие модели позволяют выявлять закономерности в данных, которые не могут быть получены с использованием простых пороговых методов, и адаптироваться к особенностям конкретных пользователей. Примером применения подобных методов являются системы биометрической аутентификации по ЭКГ, такие как технологии B-Secur (HeartKey) и устройства Nymi Band, а также исследовательские подходы, использующие нейронные сети для идентификации пользователя по сигналу сердечного цикла.

Постановка задачи. На основе анализа предметной области формулируется задача разработки HCI-системы для учета физиологического состояния пользователя на основе электрокардиограмм. Целью является получение, обработка и интерпретация ЭКГ-сигнала для оценки текущего функционального состояния пользователя с последующим использованием этой информации в адаптивных интерфейсах.

Входными данными являются ЭКГ-сигналы, регистрируемые с помощью носимых или стационарных устройств. Сигнал представляет собой временной ряд, отражающий электрическую активность сердца и содержащий информацию о сердечном ритме и его вариабельности. На его основе вычисляются производные характеристики, такие как частота сердечных сокращений и показатели вариабельности сердечного ритма, используемые в качестве признаков для анализа состояния пользователя [6].

Выходом системы является оценка физиологического состояния пользователя, представленная в виде категориальных или количественных показателей (например, уровень стресса или утомления). Полученная информация используется для адаптации интерфейса и логики работы системы.

Решение задачи включает этапы предварительной обработки ЭКГ-сигнала (фильтрация и детекция R-пиков), извлечения информативных признаков (временных, частотных и нелинейных характеристик HRV), а также применения методов машинного обучения для построения модели оценки состояния пользователя [7].

Задача характеризуется рядом ограничений. ЭКГ-сигнал подвержен шумам и артефактам, а физиологические показатели демонстрируют высокую межиндивидуальную вариативность. Дополнительно требуется обеспечение обработки данных в режиме, близком к реальному времени, что накладывает ограничения на вычислительную сложность алгоритмов.

Методы анализа ЭКГ и извлечения признаков. Электрокардиографический сигнал представляет собой временной ряд, отражающий электрическую активность сердца, и содержит ряд информативных характеристик, используемых для оценки физиологического состояния пользователя. К основным параметрам относятся частота сердечных сокращений (ЧСС), интервалы между характерными точками сигнала (в частности, RR-интервалы), а также показатели вариабельности сердечного ритма (HRV).

Частота сердечных сокращений определяется как количество сердечных циклов в единицу времени и является базовым индикатором состояния сердечно-сосудистой системы. Более информативным показателем является вариабельность сердечного ритма, представляющая собой изменения интервалов между последовательными сокращениями сердца. HRV отражает активность автономной нервной системы и баланс между её симпатическим и парасимпатическим отделами.

Для анализа HRV используются различные группы показателей. К временным характеристикам относятся, например, стандартное отклонение интервалов NN (SDNN) и среднеквадратичное отклонение последовательных различий (RMSSD). Частотные характеристики включают мощности в низкочастотном (LF) и высокочастотном (HF) диапазонах, а также их соотношение (LF/HF), применяемое для оценки симпато-вагального баланса. Кроме того, применяются нелинейные методы анализа, позволяющие учитывать сложную динамику сердечного ритма.

Процесс извлечения признаков из ЭКГ-сигнала включает несколько последовательных этапов обработки временного ряда. На первом этапе выполняется предварительная обработка сигнала, направленная на устранение шумов и артефактов. Обычно применяются полосовые фильтры для удаления низкочастотного дрейфа и высокочастотных помех, возникающих вследствие движения пользователя или внешних воздействий.

Далее осуществляется детекция характерных точек ЭКГ, прежде всего R-пиков, соответствующих максимальным значениям комплекса QRS. На основе последовательности R-пиков вычисляются RR-интервалы, являющиеся основой для дальнейшего анализа вариабельности сердечного ритма.

Следующий этап включает формирование признакового пространства. Из RR-интервалов извлекаются временные, частотные и нелинейные признаки. В ряде современных работ дополнительно используются методы анализа временных рядов и автоматического извлечения признаков, включая преобразования сигналов в спектральную или временно-частотную область, а также применение алгоритмов машинного обучения для выявления скрытых закономерностей в данных.

В условиях работы с большими объёмами данных и непрерывными потоками сигналов особое значение приобретают методы масштабируемой обработки и анализа, позволяющие учитывать динамический характер данных и межпользовательскую вариативность. Это включает использование потоковой обработки, агрегации данных и адаптивных моделей, способных обновляться по мере поступления новой информации.

Интерпретация физиологических сигналов, в том числе ЭКГ, носит вероятностный характер и не позволяет однозначно определить состояние пользователя. Одни и те же изменения в показателях HRV могут быть обусловлены различными факторами, такими как физическая активность, эмоциональное состояние, индивидуальные особенности или внешние условия.

Дополнительной проблемой является высокая межиндивидуальная вариативность физиологических показателей. Модели, обученные на данных одной группы пользователей, могут демонстрировать снижение точности при применении к другим пользователям или в иных условиях эксплуатации [1]. Это ограничивает возможность построения универсальных решений и требует учета контекста и адаптации моделей.

Также существенное влияние оказывают шумы и артефакты, возникающие при регистрации сигнала, особенно при использовании носимых устройств. Ограниченная длина временных окон при анализе в реальном времени может приводить к снижению стабильности оценок, особенно для показателей HRV, требующих достаточного объема данных для корректного вычисления.

Применение методов Big Data и продвинутой аналитики. Интерпретация физиологических сигналов, в том числе ЭКГ, носит вероятностный характер и не позволяет однозначно определить состояние пользователя. Изменения показателей HRV могут быть обусловлены различными факторами, включая физическую активность, эмоциональное состояние и индивидуальные особенности. Дополнительную сложность представляет высокая межиндивидуальная вариативность, из-за которой модели, обученные на ограниченной выборке, могут демонстрировать снижение качества при применении к новым пользователям [1].

Для анализа ЭКГ-сигналов применяются методы машинного обучения и статистического анализа. К классическим подходам относятся алгоритмы классификации и регрессии, такие как метод опорных векторов, случайный лес, k-ближайших соседей и логистическая регрессия, использующие признаки вариабельности сердечного ритма [5]. Современные подходы основаны на применении нейросетевых моделей, включая сверточные и рекуррентные сети, способные автоматически извлекать признаки из временных рядов и учитывать их динамическую структуру [5].

В ряде работ используются мультимодальные подходы, объединяющие ЭКГ с другими

физиологическими сигналами, что позволяет учитывать дополнительные аспекты состояния пользователя. Выбор модели определяется требованиями к точности, интерпретируемости и вычислительной сложности, особенно в условиях работы в режиме, близком к реальному времени.

Для повышения устойчивости моделей применяются методы персонализации, включая адаптацию под конкретного пользователя, трансферное обучение и дообучение на индивидуальных данных. Использование технологий Big Data позволяет агрегировать данные от большого числа пользователей, выявлять устойчивые закономерности и повышать обобщающую способность моделей.

Масштабируемость обработки обеспечивается за счёт распределённых вычислений и хранения данных, что особенно важно для систем, работающих с потоками ЭКГ и обслуживающих большое количество пользователей одновременно.

Подход к учету физиологического состояния в HCI-системах. Подход к учету физиологического состояния пользователя в системах человек–компьютерного взаимодействия основывается на многоуровневой архитектуре, обеспечивающей сбор, обработку и интерпретацию физиологических данных с последующим использованием результатов в логике работы системы. Визуализация архитектуры представлена на рисунке 1.

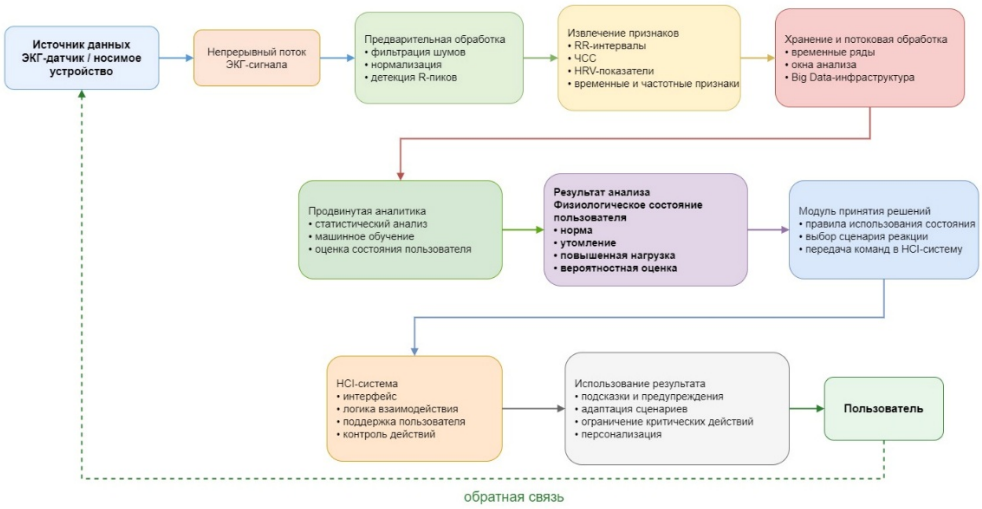


Рисунок 1. Архитектура учёта физиологического состояния пользователя в HCI-системе

На первом уровне осуществляется сбор данных с использованием сенсорных устройств, регистрирующих ЭКГ-сигнал. Источниками могут выступать носимые устройства, медицинские датчики или встроенные сенсорные модули. Получаемый сигнал представляет собой непрерывный временной ряд, требующий предварительной обработки.

Следующий уровень включает обработку сигнала, в рамках которой выполняется фильтрация шумов, нормализация и выделение ключевых характеристик, таких как RR-интервалы и показатели вариабельности сердечного ритма. На данном этапе формируется признаковое представление данных.

На уровне аналитики применяются методы машинного обучения, позволяющие оценивать текущее состояние пользователя. Результат может быть представлен в виде вероятностной оценки или дискретной категории.

Завершающий уровень связан с использованием полученной информации в HCI-системе. Оценка состояния пользователя передается в модуль управления взаимодействием и используется для адаптации интерфейса, изменения параметров системы или поддержки

принятия решений. Таким образом, реализуется последовательный конвейер: сбор данных → обработка → анализ → интерпретация → применение.

Оценка физиологического состояния может использоваться для адаптации взаимодействия с пользователем. При выявлении признаков повышенной нагрузки система может снижать сложность интерфейса, уменьшать объем информации или предоставлять дополнительные подсказки. В условиях, требующих высокой концентрации, возможно ограничение выполнения критически важных действий или введение дополнительных подтверждений.

Также состояние пользователя может применяться для динамической настройки параметров интерфейса и сценариев работы системы, что позволяет повысить эффективность взаимодействия и снизить когнитивную нагрузку. Дополнительно возможна долгосрочная адаптация на основе накопленных данных.

Интеграция модуля анализа физиологического состояния может быть реализована локально, в облаке или в гибридной архитектуре. Локальный подход обеспечивает минимальные задержки и повышенную конфиденциальность, но ограничен вычислительными ресурсами. Облачный подход позволяет использовать более сложные модели и масштабируемую инфраструктуру, однако требует передачи данных и может увеличивать задержки. Гибридная архитектура сочетает локальную предварительную обработку и облачную аналитику.

Выбор конкретного варианта интеграции определяется требованиями к времени отклика системы, уровню защиты данных и доступным вычислительным ресурсам. В практических HCI-системах часто применяется комбинированный подход, обеспечивающий баланс между производительностью и функциональностью.

Практическая часть. Для проверки применимости предложенного подхода была выполнена экспериментальная реализация программного конвейера обработки электрокардиографических сигналов и оценки физиологического состояния пользователя. В качестве источника данных использовался набор WESAD, содержащий синхронно записанные физиологические сигналы и разметку состояний испытуемых. Выбор данного набора обусловлен тем, что он включает электрокардиограммы, зарегистрированные в контролируемых условиях, а также аннотированные фрагменты, соответствующие различным функциональным состояниям пользователя. В рамках эксперимента была выбрана бинарная постановка задачи, в которой состояние пользователя интерпретировалось как «стресс» или «отсутствие стресса». К классу отсутствия стресса были отнесены состояния покоя, развлечения и медитации, тогда как к классу стресса относились соответствующие участки стрессового воздействия. Такой вариант постановки позволяет рассматривать задачу как базовую для последующего применения в адаптивных HCI-системах [8].

Входными данными являлись сигналы ЭКГ, зарегистрированные нагрудным устройством RespiBAN с частотой дискретизации 700 Гц. Для проведения эксперимента использовались записи пятнадцати испытуемых, для которых были доступны полные данные и корректная разметка состояний. С целью исключения некорректного завышения качества классификации разделение данных на обучающую, валидационную и тестовую выборки выполнялось не по отдельным временным окнам, а по пользователям. Такой способ разбиения является принципиально важным, поскольку позволяет оценивать способность модели переноситься на новых пользователей, данные которых не встречались на этапе обучения. Соответственно, задача рассматривалась в более реалистичной для практического использования постановке, когда система должна работать не только на уже известных, но и на новых пользователях.

На этапе предварительной обработки выполнялась полосовая фильтрация сигнала ЭКГ в диапазоне 0,5-40 Гц, что позволяло уменьшить влияние низкочастотного дрейфа и высокочастотных помех. После фильтрации выполнялась нормализация сигнала, обеспечивающая сопоставимость амплитудных характеристик между различными записями.

Далее запись разбивалась на временные окна длительностью 30 секунд с перекрытием 50 %. Использование оконного представления позволяет перейти от непрерывного сигнала к последовательности локальных фрагментов, для каждого из которых может быть построена отдельная оценка состояния пользователя. При формировании выборки учитывались только те окна, для которых разметка была достаточно однородной, что позволило уменьшить влияние переходных участков между состояниями и повысить надежность итоговых меток. Пример интерфейса просмотра реальных фрагментов ЭКГ и результатов классификации по временным окнам представлен на рисунке 2.

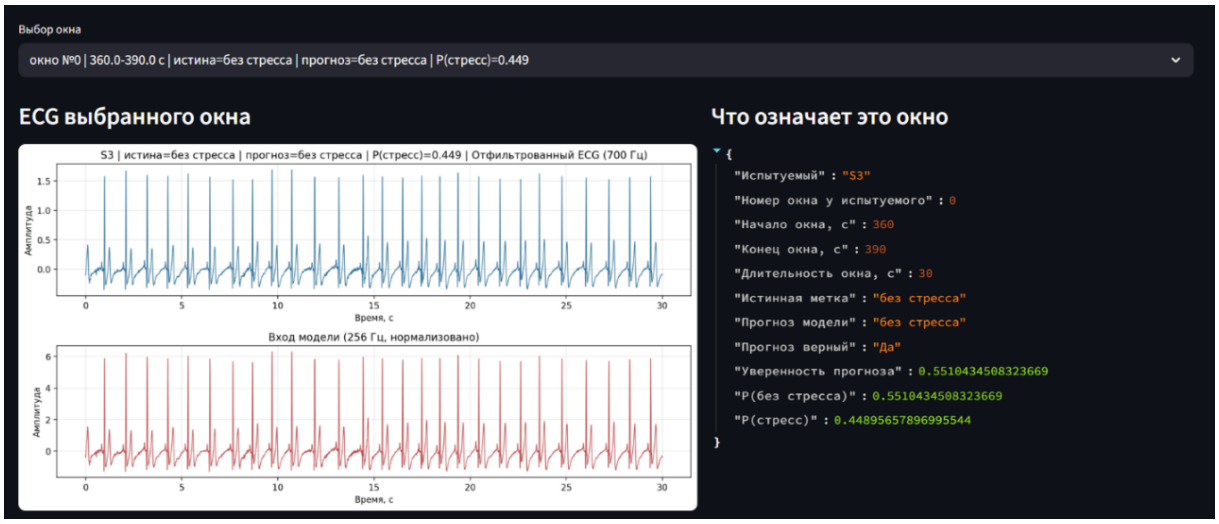


Рисунок 2. Пример интерфейса просмотра реальных фрагментов ЭКГ и результатов классификации по временным окнам

Для построения базовой модели использовался подход, основанный на анализе вариабельности сердечного ритма. На каждом окне выполнялась детекция R-пиков, после чего вычислялись RR-интервалы и производные признаки, характеризующие динамику сердечного ритма. В качестве признаков использовались средний RR-интервал, средняя частота сердечных сокращений, SDNN, RMSSD, pNN50, а также ряд вспомогательных характеристик, отражающих статистику последовательности интервалов. На основе полученного признакового описания были обучены классические модели машинного обучения, в частности логистическая регрессия и случайный лес. Такой подход представляет практический интерес, поскольку использует компактное и интерпретируемое представление сигнала, а вычислительная сложность построения признаков и классификации остается относительно невысокой.

Параллельно был реализован нейросетевой подход, в котором в качестве входных данных использовались не извлеченные признаки, а непосредственно фрагменты исходного ЭКГ-сигнала. Для этого каждый временной сегмент дополнительно приводился к частоте 256 Гц и подавался на вход одномерной сверточной нейронной сети. Использование такой архитектуры позволяет автоматизировать процесс извлечения информативных признаков и учитывать локальную структуру временного ряда. При этом нейросетевой подход предъявляет более высокие требования к объему обучающих данных и устойчивости модели к межпользовательской вариативности, что особенно важно при решении рассматриваемой задачи. Динамика изменения функции потерь и качества нейросетевой модели в процессе обучения представлена на рисунке 3.

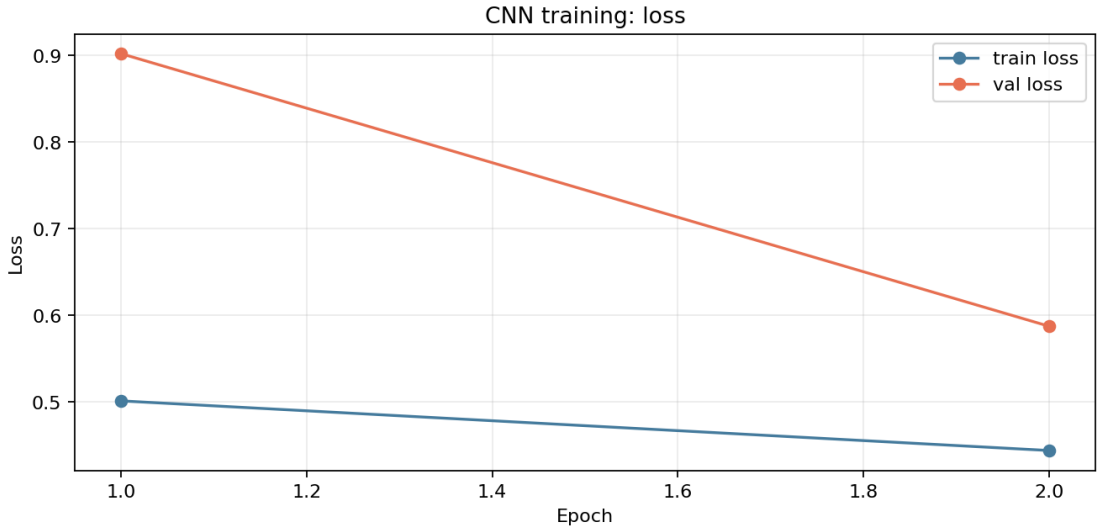


Рисунок 3. Динамика изменения функции потерь и качества нейросетевой модели в процессе обучения

Оценка качества моделей выполнялась с использованием метрик accuracy, precision, recall и F1-score. В качестве основной метрики сравнения использовался показатель macro F1, поскольку он позволяет более корректно оценивать качество классификации в условиях неоднородного распределения классов. По результатам эксперимента на валидационной выборке наилучшее значение macro F1 показала модель логистической регрессии, для которой данный показатель составил 0,827. На тестовой выборке эта модель обеспечила accuracy 0,706 и macro F1 0,622. Модель случайного леса продемонстрировала более высокие показатели на тестовой выборке, а именно accuracy 0,788 и macro F1 0,672, однако уступила логистической регрессии по качеству на этапе валидации. Нейросетевая модель 1D CNN показала accuracy 0,632 и macro F1 0,550, что свидетельствует о меньшей устойчивости при переносе на новых пользователей по сравнению с моделями, использующими HRV-признаки. Матрица ошибок модели классификации физиологического состояния пользователя на тестовой выборке представлена на рисунке 4.

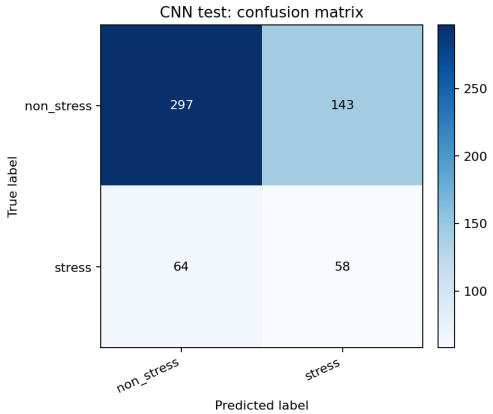


Рисунок 4. Матрица ошибок модели классификации физиологического состояния пользователя на тестовой выборке

Дополнительный анализ результатов показал, что наибольшую сложность для всех рассмотренных подходов представляет распознавание состояния стресса. Класс отсутствия стресса определяется более устойчиво, тогда как для класса стресса наблюдается снижение

точности и полноты. Это может быть связано с тем, что физиологические проявления стрессового состояния существенно зависят от индивидуальных особенностей пользователя, условий регистрации сигнала и характера стрессового воздействия. Кроме того, на итоговые показатели заметно влияет межпользовательская вариативность: качество классификации существенно различается для разных испытуемых, что подтверждает ограниченность универсальных моделей и необходимость дальнейшего развития персонализированных подходов. Сравнение качества baseline-модели на основе HRV-признаков и нейросетевой модели 1D CNN приведено на рисунке 5.

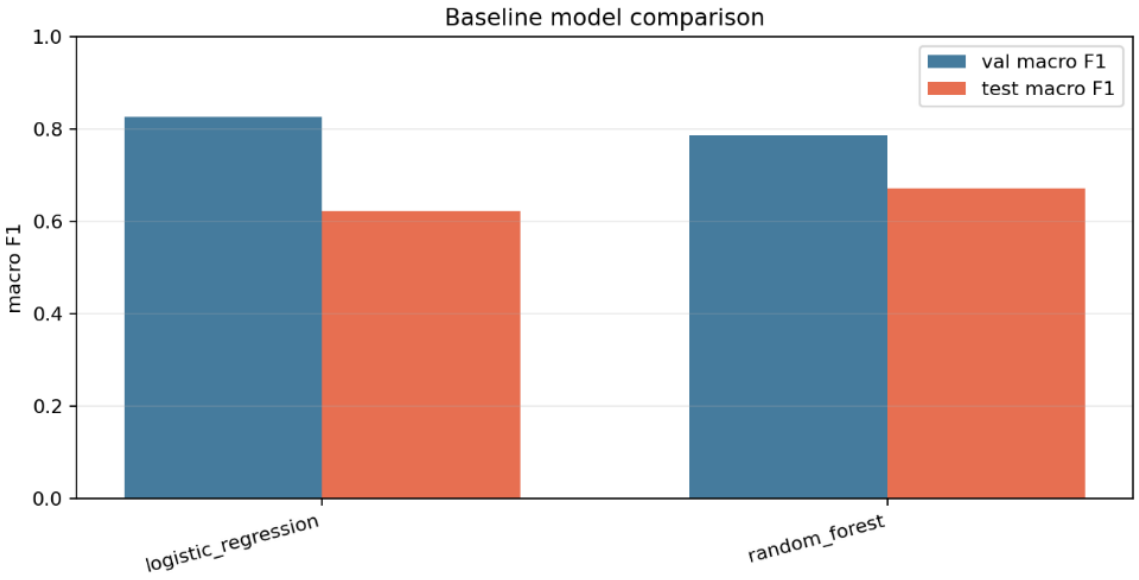


Рисунок 5. Сравнение качества baseline-модели на основе HRV-признаков и нейросетевой модели 1D CNN

В целом проведенный эксперимент показал, что электрокардиографический сигнал действительно может рассматриваться как информативный источник данных для оценки функционального состояния пользователя в задачах человеко-компьютерного взаимодействия. Наиболее практичным в рамках выполненной реализации оказался подход на основе признаков вариабельности сердечного ритма, поскольку он обеспечивает более высокую интерпретируемость результатов и более устойчивое качество при переносе на новых пользователей. Полученные результаты подтверждают перспективность применения методов анализа ЭКГ в адаптивных HCI-системах, где оценка состояния пользователя может использоваться для изменения параметров интерфейса, снижения когнитивной нагрузки и повышения надежности взаимодействия.

Заключение. В работе рассмотрены особенности применения физиологических сигналов, в частности электрокардиографии, в задачах человек–компьютерного взаимодействия. Проведён анализ предметной области, показавший ограниченность традиционных HCI-систем, основанных преимущественно на явных действиях пользователя и не учитывающих его текущее функциональное состояние. Обоснована актуальность использования физиологических данных как дополнительного источника информации для повышения эффективности взаимодействия.

Рассмотрены основные характеристики ЭКГ-сигнала и производные показатели, включая вариабельность сердечного ритма, позволяющие косвенно оценивать состояние пользователя. Отмечено, что данные показатели отражают работу автономной нервной системы и могут использоваться для выявления признаков нагрузки, утомления и стрессовых состояний. При этом интерпретация физиологических сигналов зависит от

множества факторов и требует применения методов обработки данных и аналитических моделей.

Показано, что анализ ЭКГ в условиях современных HCI-систем связан с необходимостью обработки больших объёмов данных, представленных в виде временных рядов. Рассмотрены возможности применения технологий Big Data и методов продвинутой аналитики, включая машинное обучение, для извлечения информативных признаков и построения моделей оценки состояния пользователя. Отмечено, что использование таких подходов позволяет учитывать динамический характер данных и межпользовательскую вариативность. Предложен подход к учёту физиологического состояния пользователя, включающий архитектуру системы, состоящую из этапов сбора данных, их обработки, анализа и интеграции результатов в логику HCI-системы. Рассмотрены возможные направления использования полученной информации, включая адаптацию интерфейса, поддержку пользователя и контроль действий. Отдельное внимание уделено вопросам интеграции разработанного подхода в программные системы с учётом требований к производительности и масштабируемости. Таким образом, использование ЭКГ и методов анализа данных представляет собой перспективное направление для развития адаптивных HCI-систем, ориентированных на учет текущего состояния пользователя. Дальнейшие исследования могут быть направлены на разработку более устойчивых моделей, способных учитывать индивидуальные особенности пользователей, а также на расширение перечня используемых физиологических сигналов и улучшение методов их интерпретации.

Список литературы

[1] Sano A., Picard R.W. Stress recognition using wearable sensors and mobile phones. IEEE Transactions on Affective Computing. 2013. DOI: 10.1109/ACII.2013.117.

[2] Shaffer F., Ginsberg J.P. An Overview of Heart Rate Variability Metrics and Norms. Frontiers in Public Health. 2017;5:258. DOI: 10.3389/fpubh.2017.00258.

[3] Sagiroglu S., Sinanc D. Big Data: A Review. IEEE International Conference on Collaboration Technologies and Systems. 2013. DOI: 10.1109/CTS.2013.6567202.

[4] Prajod P., Mahesh B., André E. Stressor Type Matters! Exploring Factors Influencing Cross-Dataset Generalizability of Physiological Stress Detection. In: Proceedings of the 26th International Conference on Multimodal Interaction (ICMI). New York: Association for Computing Machinery; 2024:508–517. DOI: 10.1145/3678957.3685738.

[5] Haque Y., Zawad R.S., Rony C.S.A., et al. State-of-the-Art of Stress Prediction from Heart Rate Variability Using Artificial Intelligence. Cognitive Computation. 2024;16:455–481. DOI: 10.1007/s12559-023-10200-0.

[6] Wang L., Hao J., Zhou T.H. ECG Multi-Emotion Recognition Based on Heart Rate Variability Signal Features Mining. Sensors. 2023;23:8636. DOI: 10.3390/s23208636.

[7] Ahmad Z., Rabbani S., Zafar M.R., Ishaque S., Krishnan S., Khan N. Multilevel Stress Assessment From ECG in a Virtual Reality Environment Using Multimodal Fusion. IEEE Sensors Journal. 2023;23(23):29559–29570. DOI: 10.1109/JSEN.2023.3323290/.

[8] ECG-HCI [Электронный ресурс] : репозиторий исходного кода. – Режим доступа: https://github.com/ImaAkito/ECG-HCI. – Дата доступа: 29.03.2026.

Авторский вклад

Ларченко Никита Александрович – остановка задачи исследования, разработка общей архитектуры решения, формирование методологии обработки и анализа ЭКГ-сигналов в контексте HCI-систем, реализация практической части исследования, включая сбор и предобработку данных, разработку и обучение моделей машинного обучения, проведение экспериментальных исследований и анализ полученных результатов.

Курлюк Евгений Александрович – частие в постановке задачи исследования, анализ существующих подходов к учёту физиологического состояния пользователя в HCI-системах, обзор и систематизация методов обработки физиологических сигналов, формирование теоретической базы исследования и участие в интерпретации полученных результатов.

Ганиев Абубакр Чалишевич – участие в постановке задачи исследования, анализ и сравнение современных методов обработки временных рядов и физиологических сигналов, формирование направлений дальнейших исследований, а также участие в подготовке теоретической части и обсуждении результатов.

Давыдов Максим Викторович научное руководство исследованием, формирование направления работы и постановки задач, методическое сопровождение разработки, участие в валидации полученных результатов, критический анализ содержания работы и редактирование научного текста.

DEVELOPMENT AND EXPERIMENTAL EVALUATION OF A MODEL FOR MONITORING THE USER'S PHYSIOLOGICAL STATE BASED ON ECG DATA USING BIG DATA AND ADVANCED ANALYTICS IN HCI SYSTEMS

N.A. Larchenko
Student of the Faculty of Computer-Aided Design of the Department of Electronic Engineering and Technologies BSUIR

E.A. Kurlyuk
Student of the Faculty of Computer-Aided Design of the Department of Electronic Engineering and Technologies BSUIR

A.Ch. Ganiev
Student of the Faculty of Computer Systems and Networks of the Department of Computer Science BSUIR

M.V. Davydov
First Vice Rector, Associate Professor, Candidate of Technical Sciences

Abstract. This paper explores the potential of using ECG signals in conjunction with big data methods and advanced analytics to monitor a user's physiological state in HCI systems. It examines the characteristics of ECG as a source of data on the user's condition, including the ability to derive parameters of heart rate and heart rate variability. It is shown that ECG processing in HCI systems involves the need to analyze large volumes of time-series data. An approach is proposed for building an adaptive HCI system in which the results of ECG analysis are integrated into the logic of user interaction.

Keywords: electrocardiography, HCI, user physiological state, heart rate variability, Big Data, advanced analytics, machine learning, time series analysis, adaptive interfaces, biosignals.

УДК 004.89:57.087.1

МУЛЬТИМОДАЛЬНАЯ СИСТЕМА БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ НА ОСНОВЕ КОМБИНИРОВАНИЯ ФИЗИОЛОГИЧЕСКИХ И ПОВЕДЕНЧЕСКИХ ПРИЗНАКОВ



Е.М. Косарева
Ассистент кафедры проектирования информационно-компьютерных систем БГУИР, магистр
kksrvvv@gmail.com



Д.В. Лихачевский
Доцент кафедры проектирования информационно-компьютерных систем БГУИР, кандидат технических наук, доцент
likhachevskyd@bsuir.by



Д.О. Кожевников
Магистрант кафедры проектирования информационно-компьютерных систем БГУИР
deniskoj03@gmail.com

Е.М. Косарева

Окончила Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с автоматизированными системами распознавания, системами биометрической идентификации.

Д.В. Лихаческий

Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных интересов связана с исследованием проблем радиочастотной идентификации объектов, организацией учебного и научно-исследовательского процессов в техническом университете.

Д.О. Кожевников

Окончил Белорусский государственный университет информатики и радиоэлектроники. Область научных ин-тересов связана с интеллектуальной обработки данных, схемотехникой радиоэлектронных устройств.

Аннотация. В статье предложен подход к построению мультимодальной системы биометрической идентификации, основанный на комбинировании физиологических и поведенческих характеристик (лицо, походка и пульс). Обоснован выбор модальностей. Разработана архитектура системы и предложена топология.

Ключевые слова: биометрическая идентификация, слияние данных, мультимодальная идентификация, распознавание лица, анализ походки, фотоплетизмография

Введение. Традиционные методы аутентификации, основанные на знаниях (пароли, PIN-коды) или владении (карты доступа, токены), обладают рядом существенных недостатков, включая возможность передачи третьим лицам и утери. В этой связи биометрические методы, основанные на комбинации уникальных физиологических и поведенческих характеристик человека, рассматриваются как более надежная альтернатива. Таким образом, речь идет о мультимодальной идентификации. Согласно [1], понятия «мультимодальный» и «мультибиометрический» указывают на применение более одного типа биометрической характеристики, одного датчика, одного экземпляра и/или алгоритма в той или иной комбинации для принятия определенного решения в отношении биометрической идентификации или верификации. Такой подход повышает точность идентификации и устойчивость к атакам, что особенно актуально в случае использования системы в критически важных отраслях.

Группы модальностей, применяемых в системах биометрической идентификации. Модальности, применяемые в системах биометрической идентификации, можно разделить в зависимости от природы используемых признаков на физиологические и поведенческие.

К физиологическим модальностям можно отнести характеристики, обусловленные анатомическими особенностями человека, например:

– отпечатки пальцев;

– радужную оболочку глаза;

– сетчатку глаза;

– геометрию лица и т.д.

Отпечатки пальцев характеризуются папиллярным рисунком кожного покрова пальцев, формируемым системой чередующихся гребней и впадин. Структура отпечатка включает глобальные признаки (общий тип рисунка – дуги, петли, завитки) и локальные особенности – минуции (окончания и разветвления линий), а также поры и текстурные характеристики поверхности кожи.

Радужная оболочка глаза представляет собой пигментированную мышечную структуру, окружающую зрачок. Она характеризуется сложной текстурой, включающей крипты, складки, радиальные и концентрические структуры, пигментные пятна и сосудистые элементы.

Сетчатка глаза определяется уникальным сосудистым рисунком глазного дна. Основными элементами являются артерии и вены, формирующие разветвленную топологическую структуру с характерными точками бифуркации и пересечения. Пространственное распределение сосудов и их взаимное расположение формируют индивидуальный признак, обладающий высокой устойчивостью.

Геометрия лица характеризуется совокупностью антропометрических и текстурных признаков. К структурным компонентам относятся взаимное расположение ключевых точек (глаза, нос, рот, подбородок), пропорции лица, контуры и форма отдельных областей. Дополнительно учитываются текстурные особенности кожного покрова и трехмерная структура лица.

Поведенческие модальности основаны на анализе динамических характеристик действий пользователя. К ним можно отнести:

– голос;

– динамику подписи;

– клавиатурный почерк;

– походку и т.д.

Данные модальности, как правило, менее устойчивы во времени и более подвержены влиянию внешних факторов, однако могут эффективно использоваться в качестве дополнительного уровня идентификации.

Голос как поведенческая модальность определяется совокупностью физиологических и динамических характеристик речевого аппарата. Он включает особенности голосовых связок, резонансных полостей, а также индивидуальные характеристики произношения, интонации и темпа речи.

Динамика подписи представляет собой поведенческую характеристику, включающую пространственно-временную траекторию движения пишущего инструмента. Модальность определяется формой подписи, скоростью выполнения подписи, ритмом и вариациями движений.

Клавиатурный почерк характеризуется временными параметрами взаимодействия пользователя с клавиатурой. Основными признаками являются интервалы между нажатиями клавиш, длительность удержания клавиш и ритмические особенности набора текста.

Походка определяется динамикой движения человека при перемещении в пространстве. Она включает пространственно-временные параметры шага, координацию движений конечностей, амплитуду и частоту движений, а также индивидуальные особенности кинематики тела.

В условиях постоянного усложнения методов несанкционированного доступа, а также высокой вариативности поведения злоумышленников и воздействия внешних факторов, традиционные подходы, основанные на анализе одной модальности, перестают обеспечивать необходимый уровень точности идентификации. Альтернативным подходом является идентификация на основе комбинации физиологических и поведенческих модальностей. Комбинирование модальностей разных классов позволяет компенсировать ограничения каждой отдельной модальности: физиологические признаки обеспечивают высокую стабильность, тогда как включение поведенческих признаков обеспечивает дополнительный уровень различимости.

Мультимодальное слияние данных. Мультимодальные биометрические системы проектируются по двум схемам – последовательно и параллельно. В первом случае выходные результаты анализа одной биометрической модальности используются для сужения области поиска потенциальных пользователей и передачи данной информации на вход следующему блоку. При параллельном построении системы биометрическая информация разной природы анализируется единовременно. При этом существует три возможных уровня объединений:

– на уровне признаков;

– в процессе принятия итогового решения;

– на этапе сравнения [2].

Рассмотрим процесс объединения модальностей на уровне признаков. В общем виде он начинается с формирования множества входных биометрических сигналов (формула 1):

$$X = \{X_1, X_2, ..., X_n\}, \quad (1)$$

где X_i – данные, полученные от i-й биометрической модальности.

Для каждой модальности выполняется преобразование исходных данных в биометрический шаблон (формула 2):

$$T_i = \psi(\phi(X_i)), \tag{2}$$

где $\phi(\cdot)$ включает этапы предобработки и выделения признаков;

$\psi(\cdot)$ – формирование шаблона.

Сформированные биометрические шаблоны объединяются путем слияния (формула 3).

$$T = f(T_1, T_2, ..., T_n), \tag{3}$$

где $f(\cdot)$ – функция агрегирования шаблонов различных модальностей.

Полученное интегральное представление используется классификатором для определения принадлежности пользователя к одному из классов. На основе результата классификации принимается решение о допуске.

Обобщенная схема объединения модальностей на уровне признаков приведена на рисунке 1.

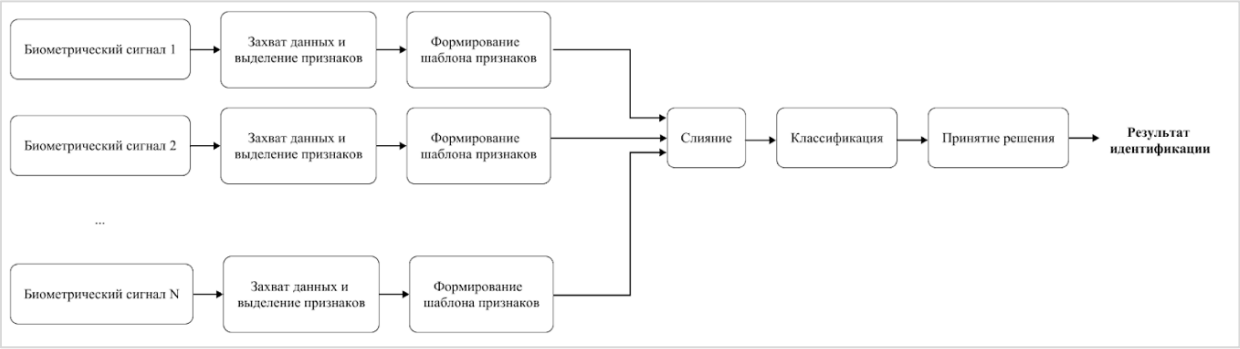


Рисунок 1. Обобщенная схема объединения модальностей на уровне признаков

Процесс объединения модальностей на уровне принятия решения аналогичен рассмотренному ранее процессу объединения модальностей на уровне сформированных признаков вплоть до этапа формирования шаблона. Различия в подходах возникают на этапе классификации. В данном случае каждая модальность рассматривается независимо (формула 4):

$$T_i \rightarrow D_i, \tag{4}$$

где D_i – результат классификации для *i*-й модальности.

Далее результаты классификации каждой модальности объединяются в единое решение по формуле 5:

$$D = g(D_1, D_2, ..., D_n), \tag{5}$$

где $g(\cdot)$ – функция агрегирования решений.

Обобщенная схема объединения модальностей на уровне уровне принятия решения приведена на рисунке 2.

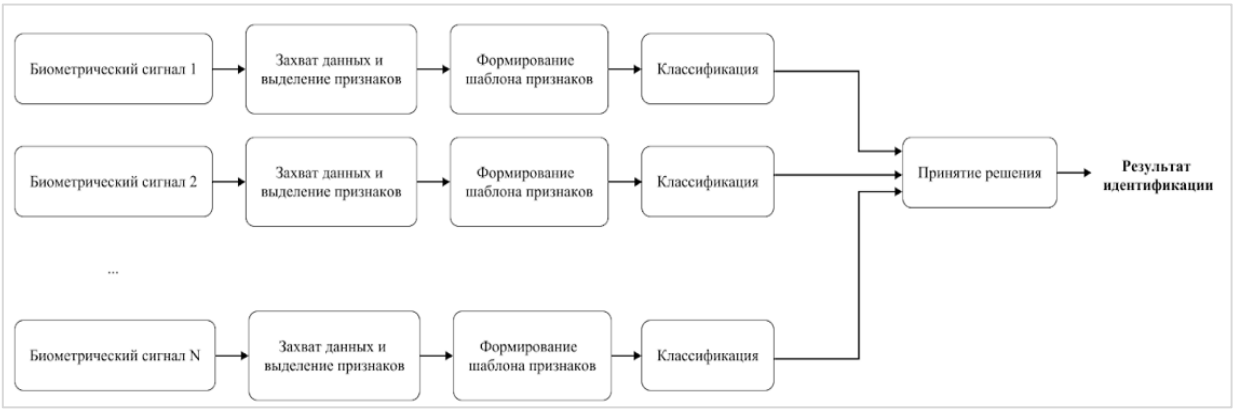


Рисунок 2. Обобщенная схема объединения модальностей на уровне принятия решения

Вариант слияния на этапе сравнения в данной работе не рассматривается, так как он относится только к задачам верификации пользователя [2].

Для задачи слияния физиологических и поведенческих признаков, таких как лицо и походка был выбран алгоритм объединения модальностей на уровне принятия решения. Это обуславливается разнородностью используемых биометрических характеристик: выбранные модальности имеют различную природу, размерность и чувствительность к внешним и психофизиологическим факторам.

Независимая обработка каждой модальности с последующим объединением результатов классификации позволяет повысить устойчивость системы к шумам и вариативности данных, а также позволяет учитывать надежность каждой модальности при формировании итогового решения.

Обоснование выбора модальностей. В контексте проектируемой системы мультимодальной биометрической идентификации в качестве основных модальностей были выбраны лицо, походка и пульс.

Лицо является часто используемой и хорошо изученной биометрической модальностью. Идентификация по лицу обеспечивает высокую точность при наличии достаточного освещения, не требует физического контакта субъекта с оборудованием и допускает захват данных на расстоянии [3].

Преимуществом является возможность использования стандартных камер видеонаблюдения, что существенно снижает стоимость инфраструктуры. Однако при этом данная модальность подвержена влиянию внешних факторов – изменения освещения, ракурса, наличия маски или головного убора – что обусловливает необходимость ее дополнения другими модальностями [3].

Включение походки обусловлено ее способностью обеспечивать идентификацию на расстоянии и без необходимости явного взаимодействия с системой. Походка формируется под влиянием индивидуальных биомеханических характеристик человека, включая длину шага, ритм и координацию движений, что делает ее трудновоспроизводимой для злоумышленников. При этом походка менее чувствительна к качеству изображения, что компенсирует ограничения первой модальности в неблагоприятных условиях съемки [4].

Использование пульса в качестве дополнительной модальности обусловлено тем, что параметры сердечного ритма, включая вариабельность интервалов между ударами, определяются внутренними биологическими процессами и практически не поддаются сознательному контролю или имитации. Это делает пульс надежным фактором для подтверждения «живости» субъекта (liveness detection) и снижает риск атак с использованием статических изображений или записей [5].

Система мультимодальной биометрической идентификации. В настоящем исследовании предлагается архитектура системы биометрической идентификации, состоящая из шести модулей: модуля захвата данных, модуля предобработки данных, модуля извлечения признаков, модуля биометрической идентификации, модуля анализа психофизиологического состояния и модуля принятия решения.

Модуль захвата данных обеспечивает первичный сбор биометрических сигналов и представлен набором сенсоров: сенсор А представляет собой видеокамеру для фиксации кинематики тела субъекта; сенсор Б, также видеокамера, захватывает изображение лица; сенсор В осуществляет регистрацию пульсового сигнала методом фотоплетизмографии с извлечением показателей частоты сердечных сокращений (ЧСС) и вариабельности сердечного ритма (ВСР).

Модуль предобработки данных выполняет очистку и нормализацию сигналов, поступающих от каждого сенсора.

Модуль извлечения признаков преобразует предобработанные данные в числовые векторы признаков, пригодные для последующего анализа.

На выходе модуля формируются три признаковых вектора, которые поступают параллельно в модуль биометрической идентификации и в модуль анализа психофизиологического состояния.

Модуль биометрической идентификации решает задачу идентификации личности субъекта. На этапе слияния признаков производится конкатенация векторов походки и лица с применением совместного эмбеддинга FaceNet и GaitNet.

Полученный агрегированный вектор сопоставляется с профилями субъектов, хранящимися в базе данных системы.

Результатом работы модуля является идентификатор личности с соответствующим скором уверенности.

Модуль анализа психофизиологического состояния функционирует параллельно с модулем идентификации и решает задачу оценки текущего состояния субъекта.

Модуль принятия решения интегрирует полученные результаты и формирует итоговый профиль субъекта, включающий идентификатор личности, категорию состояния и метку времени события. Сформированный профиль передаётся во внешнюю систему контроля и управления доступом (СКУД) через интеграционный интерфейс API.

Структурная схема системы приведена на рисунке 3.

Топология системы. С точки зрения топологии инфраструктуры системы основу системы составляют две боковые видеокамеры (сенсоры А.1 и А.2) размещенные по обе стороны траектории движения человека и формирующие бинокулярную схему наблюдения. Их основная функция – захват кинематических параметров походки с минимизацией эффекта самоокклюзии и обеспечением устойчивого анализа движения. Фронтальная видеокамера (сенсор Б.1) ориентирована на лицо субъекта и предназначена для детекции и анализа микровыражений.

Контактный пульсометр (сенсор В.1) интегрирован в зону остановки пользователя и обеспечивает регистрацию физиологических параметров в момент взаимодействия.

Все сенсоры синхронизированы по времени и передают данные в единый вычислительный модуль, где осуществляется их слияние и комплексный анализ.

Такая топология обеспечивает одновременный сбор поведенческих и физиологических признаков (рисунок 4).

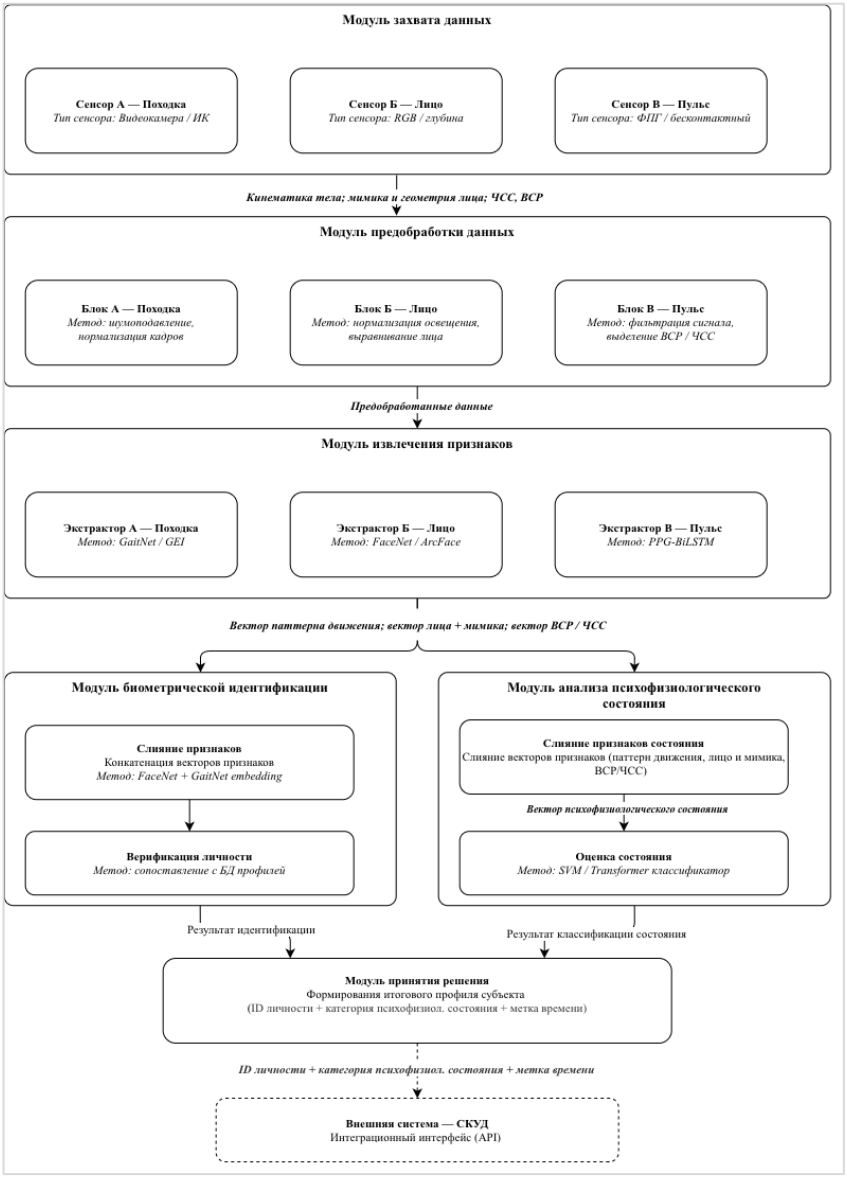


Рисунок 3. Структурная схема системы

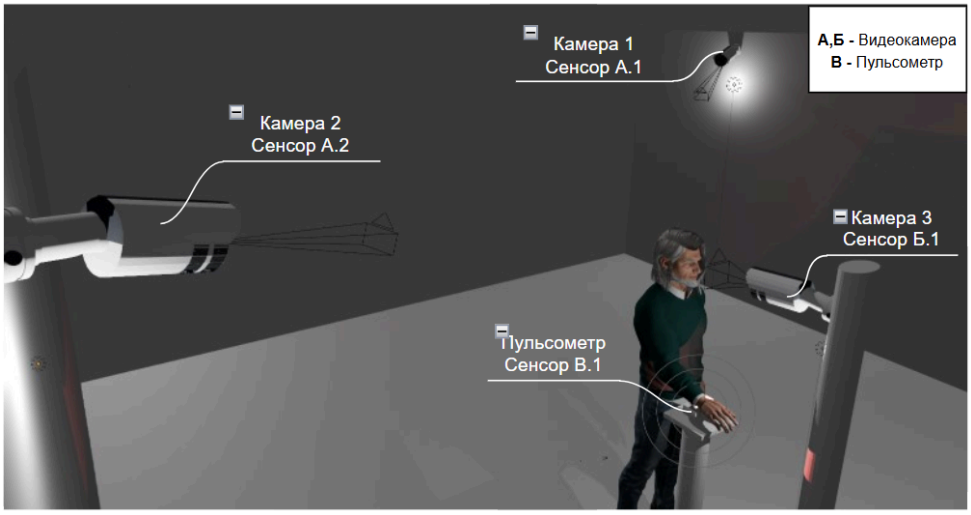


Рисунок 4. Топология системы

Заключение. В работе предложен подход к построению мультимодальной системы биометрической идентификации на основе объединения физиологических и поведенческих признаков.

Показано, что использование модальностей лица, походки и пульса, а также их слияние на уровне принятия решений.

Разработанная архитектура и топология могут быть применены в системах контроля доступа с повышенными требованиями к безопасности.

Список литературы

[1] ГОСТ Р 54411–2018. Информационные технологии. БИОМЕТРИЯ. Мультимодальные и другие мультибиометрические технологии / ФГУП «СТАНДАРТИНФОРМ». – Москва, 2019. – 22 с.

[2] Стефаниди, А. Ф. Исследование мультимодальных алгоритмов биометрической идентификации на основе методов цифровой обработки речевых сигналов и изображений : дис. ... канд. техн. наук : 2.2.13/ А. Ф. Стефаниди; Ярославский государственный университет им. П.Г. Демидова. – Ярославль, 2022. – 130 с.

[3] Gupta A. Advancements and Challenges in Face Recognition Technology // International Journal of Computer Trends and Technology (IJCTT). – 2024. – Vol. 72, No. 11. – Pp. 92–104. – DOI: 10.14445/22312803/IJCTT-V72I11P110.

[4] Munusamy V., Senthilkumar S. Emerging Trends in Gait Recognition Based on Deep Learning: A Survey // PeerJ Computer Science. – 2024. – Vol. 10. – Art. e2158. – DOI: 10.7717/peerj-cs.2158.

[5] Aziz M.A., et al. PPG-Based Biometric Authentication: A Review on Architectures, Datasets, Attacks and Security Challenges // Sensors / PMC. – 2025.

Авторский вклад

Косарева Екатерина Максимовна – подготовка обзора модальностей, используемых в системах биометрической идентификации; проведение сравнительного анализа подходов к мультимодальному слиянию данных; разработка структурной схемы системы.

Лихаческий Дмитрий Викторович – постановка задачи исследования, обоснование выбора модальностей.

Кожевников Денис Олегович – разработка топологии системы.

MULTIMODAL BIOMETRIC IDENTIFICATION SYSTEM BASED ON A COMBINATION OF PHYSIOLOGICAL AND BEHAVIORAL CHARACTERISTICS

E.M. Kosareva
Assistant of the Department of Information Computer Systems Design BSUIR, Master of Tech. sciences

D.V. Likhachevsky
Associate Professor, Department of Information Computer Systems Design BSUIR, PhD of Technical sciences, Associate Professor

D.O. Kojevnikov
Master student of the Department of Information Computer Systems Design BSUIR

Abstract. This article proposes an approach to constructing a multimodal biometric identification system based on combining physiological and behavioral characteristics (face, gait, and heart rate). The choice of modalities is substantiated. The system architecture is developed, and a topology is proposed.

Keywords: biometric identification, data fusion, multimodal identification, face recognition, gait analysis, photoplethysmography.

Научное издание

BIG DATA И АНАЛИЗ ВЫСОКОГО УРОВНЯ

BIG DATA AND ADVANCED ANALYTICS

СБОРНИК НАУЧНЫХ СТАТЕЙ
XII МЕЖДУНАРОДНОЙ НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ

(Республика Беларусь, Минск, 23 апреля 2026 года)

В двух частях

Часть 2

В авторской редакции
Ответственный за выпуск *С.К. Дик*
Компьютерная верстка *А.В. Воробей, Е.М. Косарева*

Подписано в печать 08.04.2026. Формат 60×84 1/8. Бумага офсетная. Гарнитура «Таймс».
Отпечатано на ризографе. Усл. печ. л. 41,6. Уч.-изд. л. 35,8. Тираж 50 экз. Заказ 49.

Издатель и полиграфическое исполнение: учреждение образования
«Белорусский государственный университет информатики и радиоэлектроники».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий №1/238 от 24.03.2014,
№2/113 от 07.04.2014, №3/615 от 07.04.2014.
Ул. П. Бровки, 6, 220013, г. Минск

ISBN 978-985-543-878-7

