

ISSN 2413 - 0133
Scientific journal

ИНФОРМАЦИОННЫЕ И МАТЕМАТИЧЕСКИЕ ТЕХНОЛОГИИ В НАУКЕ И УПРАВЛЕНИИ

№3(43)/2026

ИНФОРМАЦИОННЫЕ И МАТЕМАТИЧЕСКИЕ ТЕХНОЛОГИИ В НАУКЕ И УПРАВЛЕНИИ

Научный журнал
№ 3 (43)



EDITORIAL BOARD**РЕДАКЦИОННАЯ КОЛЛЕГИЯ ЖУРНАЛА**

Ablameyko S.V.	Абламейко С.В., академик НАН Беларуси, Минск, БГУ
Arshinskiy L.V.	Аршинский Л.В., д.т.н., Иркутск, ИрГУПС
Berestneva O.G.	Берестнева О.Г., д.т.н., Томск, ТПУ
Boukhanovsky A.V.	Бухановский А.В., д.т.н., Санкт-Петербург, НИУ ИТМО
Bychkov I.V.	Бычков И.В., академик РАН, Иркутск, ИДСТУ СО РАН
Woern H.	Вёрен Х., Германия, Karlsruhe Institute of Technology (KIT)
Wolfengagen V.E.	Вольфенгаген В.Э., д.т.н., Москва, МИФИ
Gornov A.Y.	Горнов А.Ю., д.т.н., Иркутск, ИДСТУ СО РАН
Gribova V.V.	Грибова В.В., чл.-корр. РАН, д.т.н., Владивосток, ИАПУ ДВО РАН
Groumpos P.	Грумπος Π., Греция, University of Patras
Zorina T.G.	Зорина Т.Г., д.т.н., Республика Беларусь, Институт энергетики НАН Беларуси
Kazakov A.L.	Казаков А.Л., д.ф.-м.н., профессор РАН, Иркутск, ИДСТУ СО РАН
Kalimoldaev M.N.	Калимолдаев М.Н., академик НАН РК, Республика Казахстан, ИИВТ
Karpenko A.P.	Карпенко А.П., д.ф.-м.н., Москва, МГТУ им. Баумана
Komendantova N.P.	Комендантова Н.П., PhD, Австрия, Лаксенбург, ПАСА
Kureichik V.V.	Курейчик В.В., д.т.н., профессор, Таганрог, ЮФУ
Lis R.	Лис Р., Польша, Wroclaw University of Science and Technology
Marchenko M.A.	Марченко М.А. д.ф.-м.н., профессор РАН, Новосибирск, ИВМиМГ СО РАН
Massel L.V.	Массель Л.В., д.т.н., Иркутск, ИСЭМ СО РАН
Moskvichev V.V.	Москвичев В.В., д.т.н., Красноярск, ФИЦ ИВТ
Ovtcharova J.	Овчарова Ж., Германия, Karlsruhe Institute of Technology (KIT)
Popov G.T.	Попов Г.Т., Болгария, г. София, Технический университет
Sidorov D.N.	Сидоров Д.Н., д.ф.-м.н., профессор РАН, Иркутск, ИСЭМ СО РАН
Smirnov S.V.	Смирнов С.В., д.т.н., Самара, ИПУСС РАН
Stennikov V.A.	Стенников В.А., академик РАН, Иркутск, ИСЭМ СО РАН
Stylios C.	Стилос Х., Греция, Technological Educational Institute of Epirus
Taratukhin V.V.	Таратухин В.В., Германия, ERCIS, University of Muenster
Khamisov O.V.	Хамисов О.В., д.ф.-м.н., Иркутск, ИСЭМ СО РАН
Hodashinsky I.A.	Ходашинский И.А., д.т.н., Томск, ТУСУР
Chubarov L.B.	Чубаров Л.Б., д.ф.-м.н., Новосибирск, ФИЦ ИВТ

EXECUTIVE EDITORIAL**ИСПОЛНИТЕЛЬНАЯ РЕДАКЦИЯ ЖУРНАЛА**

Chief Editor	Massel L.V.	Главный редактор	Массель Л.В.	д.т.н., Иркутск, ИСЭМ СО РАН
Editor	Tuktartova P.A.	Редактор	Туктарова П.А.	к.э.н., Иркутск, ИСЭМ СО РАН
Editor	Kuzmin V.R.	Редактор	Кузьмин В.Р.	к.т.н., Иркутск, ИСЭМ СО РАН
Editor	Massel A.G.	Редактор	Массель А.Г.	к.т.н., Иркутск, ИСЭМ СО РАН
Tech. Editor	Popova M.D.	Техн. редактор	Попова М.Д.	Иркутск, ИСЭМ СО РАН
Designer	Pesterev D.V.	Дизайнер	Пестерев Д.В.	Иркутск, ИСЭМ СО РАН

Учредитель: Федеральное государственное бюджетное учреждение науки

Институт систем энергетики им. Л.А. Мелентьева Сибирского отделения Российской академии наук
(ИСЭМ СО РАН)

Адрес учредителя, издателя и редакции

664033 г. Иркутск, ул. Лермонтова, 130

Тел: (3952) 42-47-00 Факс: (3952) 42-67-96

Сайт журнала - <https://www.imt-journal.ru/>

Журнал включен в Российский индекс научного цитирования (РИНЦ). Номер контракта 202-04/2016.

Журнал включен в перечень рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученых степеней кандидата и доктора наук (по состоянию на 29.03.2022) по научным специальностям 1.2.2; 2.3.1; 2.3.3; 2.3.5; 2.4.5.

Журнал зарегистрирован в Роскомнадзоре. Регистрационный номер ПИ № ФС 77 – 73539 от 31.08.2018.

Отпечатано в типографии «Дубль Принт»

СС ВУ

Адрес типографии: 664046, г. Иркутск, ул. Волжская, 14

© Издательство ИСЭМ СО РАН

*Подписка на журнал доступна со второго полугодия 2021 г.

Цена свободная (12+)

Дата выхода 03.09.2026 г. Тираж 50 экз.

Методологические аспекты информационных и математических технологий

- Бочков А.В.** Риск как имманентное свойство эргатических систем в эпоху цифровых двойников: от «цифрового Левиафана» к парадигме симбиотического управления 5

Методы, технологии и приложения искусственного интеллекта

- Морозов М.И.** Нейросетевой стекинг для прогнозирования убытков в страховании 22
- Потапов А.Д., Лукьянов Н.Д.** Адаптивная гибридная нейросетевая архитектура на основе внимания для идентификации дикторов в корпоративных совещаниях 31
- Углев В.А., Смирнов Г.А.** Интеллектуальная автоматизированная обучающая система AESU и лежащие в её основе концепции, механизмы и методы 42

Математическое моделирование и его применение в научных исследованиях

- Савотченко С.Е., Афанасьева Н.О.** Моделирование свойств критического значения показателя преломления при переходах пика интенсивности между нелинейной и градиентной оптическими средами 56
- Туренова И.А., Моисеева С.П., Шкленник М.А.** Перспективы использования систем массового обслуживания для моделирования работы маркетплейсов 70
- Иваньо Я.М., Энхбат Ариунжаргал.** Оптимизация численности сельскохозяйственных животных с учетом деградации пастбищ в Монголии 83

Математические, информационные и интеллектуальные технологии в энергетике

- Жиравок А.Н., Зуев А.В.** Метод идентификации дефектов в технических системах, описываемых линейными моделями 93
- Панкратьев П.С.** Разработка гибридного метода оптимального распределения электрической нагрузки с учетом трудноформализуемых факторов 103
- Щукин Н.И.** Разработка цифровой модели цифрового двойника ветровой электростанции с использованием онтологического моделирования 119

Цифровая экономика и управление

- Ерженин Р.В.** Построение целевой функции учетной информационной системы публично-правового образования 134
- Бастанов А.Э.** Интеллектуальный анализ спроса и оптимизация распределения оборотного капитала селлера на электронных маркетплейсах: методологическая основа, алгоритм и дизайн экспериментов 144
- Карамушко Д.М., Чернов А.В.** Оптимизация расписания авиакомпании с учётом моделирования трансферных пассажиропотоков 154

Программные системы и комплексы

- Фелагерева Е.Ю., Рудоман Н.Р., Новиков А.Д.** Разработка гибридного модуля сопровождения воздушных объектов на основе GRU и фильтра Калмана 165
- Гун Шэншо.** Миварная система обработки информации и принятия решений для группового управления гетерогенными складскими роботами 178
- Исмагулов М.Е.** Иерархическая модель структуры документа для решения задачи генерации конспектов лекций 197

Methodological Aspects of Information and Mathematical Technologies

- Bochkov A.V.** Risk as an immanent property of ergatic systems in the era of digital twins: from the "digital Leviathan" to the paradigm of symbiotic management 5

Methods, technologies and applications of artificial intelligence

- Morozov M.I.** Neural network stacking for loss forecasting in Insurance 22
- Potapov A.D., Lukyanov N.D.** Adaptive hybrid neural network architecture with attention for speaker identification in corporate meetings 31
- Ughev V.A., Smirnov G.A.** AESU Intelligent Tutoring Systems and its underlying concepts, mechanisms and methods 42

Mathematical Modeling and Its Application in Scientific Research

- Savotchenko S.E., Afanasyeva N.O.** Modeling the properties of the critical refractive index during peak intensity transitions between nonlinear and gradient optical media 56
- Turenova I.A., Moiseeva S.P., Shklennik M.A.** Prospects of using queuing systems for modeling the operation of marketplaces 70
- Ivanyo Ya.M., Ariunjargal Enkhbat.** Optimization of the number of farms livestock taking into account the degradation of pastelands in Mongolia 83

Mathematical, Information, and Intelligent Technologies in Energy

- Zhirabok A.N., Zuev A.V.** Method of fault estimation in technical systems described by linear models 93
- Pankratev P.S.** Development of a hybrid method for optimal electric load distribution considering hard-to-formalize factors 103
- Shchukin N.I.** Digital model of a digital twin of a wind power plant development using ontological modeling 119

Digital Economy and Management

- Erzhenin R.V.** Construction of the target function of the accounting information system of a public-law entity 134
- Bastanov A.E.** Intelligent demand analysis and working capital allocation optimization for sellers on electronic marketplaces: methodological foundation, algorithm and design of experiments 144
- Karamushko D.M., Chernov A.V.** Airline schedule optimization with transfers modelling 154

Software Systems and Complexes

- Felagereva E.Yu., Rudoman N.R., Novikov A.D.** Development of a hybrid module for tracking air objects based on GRU and the Kalman filter 165
- Gong Shengshuo.** A Mivar information processing and decision-making system for group control of heterogeneous warehouse robots 178
- Ismagulov M.E.** A Hierarchical model of document structure for the task of generating lecture summaries 197

Методологические аспекты информационных и математических технологий

УДК 005.334:004.94

DOI:10.25729/ESI.2026.43.3.001

Риск как имманентное свойство эргатических систем в эпоху цифровых двойников: от «цифрового Левиафана» к парадигме симбиотического управления

Бочков Александр Владимирович

АО «НИИАС», Россия, Москва, a.bochkov@gmail.com

Аннотация. Современный переход к «управлению, основанному на данных», опирающемуся на создание полномасштабных цифровых двойников сложных социально-экономических систем, сталкивается с фундаментальным вызовом. В статье доказывается, что оптимизационные задачи, решаемые для таких цифровых моделей, являются некорректными по отношению к оригиналу, поскольку ключевые объекты управления – эргатические системы – принципиально неформализуемы из-за наличия человеческого компонента. Риск в этой парадигме перестает быть внешней угрозой, становясь имманентным свойством самой системы управления, порождаемым неизбежным и нарастающим расхождением траекторий виртуальной модели и реального эргатического процесса. Стратегии, оптимальные для «цифрового Левиафана» (тотальной симуляции), ведут к дестабилизации оригинала, эрозии человеческого компонента и выходу системы за границы управляемости. Автор обосновывает необходимость смены парадигмы: от поиска точечного оптимума для неполной модели к синтезу робастных и адаптивных стратегий, явно минимизирующих функционал риска. Вводится и противопоставляется два концепта: «цифровой Левиафан» (искушение управлять симуляцией) и «формализм риска» (дисциплина постоянного сравнения карты и территории). Ключевым элементом новой парадигмы «симбиотического управления» является формальное включение человеческого оператора, как корректирующего контура в итеративный диалог между моделью и реальностью. Целью управления становится не абсолютный оптимум в виртуальном пространстве, а динамическая адаптивная эффективность в реальном мире, достигаемая через постоянное управление имманентным риском расхождения траекторий. Предложенная архитектура симбиотического управления, основанная на принципах байесовского обновления и минимакса, позволяет не просто минимизировать риск, но и активно управлять эпистемической неопределенностью, превращая человеческую интуицию в структурированный источник коррекции модели. В работе показано, что такой подход обеспечивает устойчивость сложных социотехнических систем в условиях принципиальной неполноты знания, предлагая путь от иллюзии абсолютного контроля к мудрости совместной навигации.

Ключевые слова: цифровой двойник, эргатическая система, имманентный риск, оптимизационная задача, некорректность модели, симбиотическое управление, адаптивная стратегия, человеческий фактор, эпистемическая неопределенность

Цитирование: Бочков А.В. Риск как имманентное свойство эргатических систем в эпоху цифровых двойников: от «цифрового Левиафана» к парадигме симбиотического управления / А.В. Бочков // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 5-21. – DOI:10.25729/ESI.2026.43.3.001.

Введение. Современное управление все чаще строится на использовании «цифровых двойников» – виртуальных моделей физических объектов или процессов, обновляемых в реальном времени. Их применение формально сводится к задаче оптимизации заданного функционала (эффективности, затрат) на траекториях модельного процесса. Однако ключевые объекты управления – предприятия, города, транспортные системы – относятся к классу эргатических систем, чьи цели (устойчивость, легитимность, адаптивность) шире операционной эффективности. Полное пространство их состояний, особенно из-за человеческого фактора, принципиально неформализуемо. Любая цифровая модель оперирует лишь упрощенным подпространством, а её целевой функционал отражает лишь часть истинных, часто противоречивых целей оригинала. Это порождает фундаментальный парадокс: оптимизация идеализированной абстракции не гарантирует улучшения реальности.

Следствием становится онтологизация риска: риск превращается из внешней угрозы во внутреннее свойство архитектуры управления, вытекающее из несоответствия модели и реальности [1]. Цель статьи – проанализировать природу этого имманентного риска, описать деструктивную динамику, порождаемую «цифровым Левиафаном», и обосновать контуры новой парадигмы: формализма риска и симбиотического управления, где риск становится ключевым управляемым функционалом, а человек – главным корректирующим контуром.

1. «Цифровой Левиафан»: Некорректность оптимизационной задачи и генезис имманентного риска. Современный этап технологического развития, характеризуемый взрывным ростом вычислительных мощностей и доступности данных, породил новую управленческую утопию – возможность построения полномасштабного цифрового двойника сложной социально-экономической или социотехнической системы. Этот двойник, будучи насыщенным историческими и текущими данными, должен был стать идеальным полигоном для экспериментов, точным инструментом прогнозирования и, что самое главное, основой для принятия оптимальных управленческих решений. Формальная постановка задачи в рамках этой парадигмы выглядит безупречно с точки зрения математической теории управления и исследования операций. Пусть реальная управляемая система существует в некотором фазовом пространстве состояний $\mathcal{X}_R$. Управляющее воздействие $u(t)$ выбирается из допустимого множества $\mathcal{U}$. Динамика системы описывается стохастическим дифференциальным уравнением или их системой:

$$dX_t = f(X_t, u_t, t)dt + \sigma(X_t, u_t, t)dW_t,$$

где $X_t \in \mathcal{X}_R$ – состояние системы в момент времени t , f – детерминированный дрейф, σ – волатильность, а W_t – винеровский процесс, моделирующий случайные возмущения. Цель управления формализуется через целевой функционал $J_R(u)$, который требуется максимизировать (например, суммарный доход) или минимизировать (например, совокупные издержки или риск). Часто он имеет вид математического ожидания от некоторого интегрального или терминального критерия:

$$J_R(u) = E \left[\int_{t_0}^T L(X_t, u_t, t)dt + \Phi(X_T) \right],$$

где L – функция текущих затрат/доходов, а Φ – терминальный выигрыш. Классическая задача стохастической оптимизации заключается в нахождении стратегии управления $u^*(t)$ (часто в форме обратной связи $u^*(X_t, t)$), которая доставляет оптимальное значение функционалу: $u^* = \arg \max_{u \in \mathcal{U}} J_R(u)$.

Казалось бы, при наличии достаточно точной модели функций f , σ и критерия J_R , современные методы динамического программирования, принципа максимума Понтрягина или обучения с подкреплением позволяют приблизиться к решению этой задачи.

Однако фундаментальная проблема, лежащая в основе всей этой конструкции, заключается в том, что для эргатических систем исходное предположение об изоморфизме $\mathcal{X}_M \cong \mathcal{X}_R$ и точном соответствии (f, σ, J) и (f_M, σ_M, J_M) принципиально невыполнимо. Эргатическая система, по определению, является целостностью, возникающей из неразрывного взаимодействия человеческого, технологического и средового компонентов. Человеческий компонент вносит в систему качество, которое можно назвать «рефлексивной неформализуемостью» [2]. Пространство состояний $\mathcal{X}_R$ такой системы включает в себя не только измеримые технические параметры (обороты, напряжение, объемы), но и ментальные состояния операторов, их интенции, уровень доверия к системе, неформальные договоренности внутри коллектива, культурные нормы, имплицитное знание [3], которое невозможно вербализовать и, тем более, представить в виде структурированных данных. Эти

аспекты не просто сложно измерить; они принадлежат к качественно иной онтологической категории, нежели данные, обрабатываемые компьютером. Следовательно, оператор Π не является изоморфизмом, а представляет собой жесткую редукцию, проецирующую многомерное, отчасти континуальное, отчасти дискретное, отчасти качественное пространство реальности на конечномерное, чисто количественное подпространство модели: $\mathcal{X}_M \subset \mathcal{X}_R$, причем $\dim \mathcal{X}_M \ll \dim \mathcal{X}_R$). Более того, это подпространство не является статичным: по мере изменения системы, ее целей и контекста, релевантное подпространство состояний может «плавать» и модель, построенная на исторических данных, эти изменения уловить не в состоянии.

Это приводит к первому и главному источнику имманентного риска: некорректности оптимизационной задачи в постановке Адамара [4]. Задача называется корректно поставленной, если ее решение существует, единственно и непрерывно зависит от исходных данных. В нашем случае «исходными данными» является сама модель, как отображение реальности. Некорректность возникает именно из-за разрыва между моделью и оригиналом.

Имманентный риск возникает из трех ключевых ограничений при моделировании эргатических систем:

Редукция целей: Оптимизация по узкому формальному критерию (прибыль, эффективность) ведет к деградации по другим жизненно важным, но неформализуемым целям (устойчивость, адаптивность).

Игнорирование рефлексивности: Модель предполагает статичное поведение людей, тогда как в реальности они наблюдают и адаптируются к управляющим воздействиям, меняя систему и обесценивая прогнозы.

Слепота к «черным лебедям»: Модели, обученные на исторических данных, не учитывают катастрофические события за пределами этой выборки. Оптимизированная под «спокойную» модель система становится хрупкой к непредсказуемым потрясениям.

Таким образом, генезис риска коренится в применении аппарата оптимизации к принципиально неформализуемой эргатической системе. «Цифровой Левиафан» – это автономная управляющая сущность, действующая на основе своей редуцированной модели. Парадокс в том, что чем совершеннее его модель и алгоритмы, тем увереннее он в своей «оптимальности» и тем глубже пропасть между симуляцией и реальностью.

Риск при этом не уменьшается, а, напротив, маскируется и накапливается. Он становится не внешним параметром, который можно оценить и застраховать, а внутренним свойством архитектуры управления, его скрытой пружинной, которая рано или поздно разожмет свою разрушительную силу, когда накопленное расхождение ρ_T превысит критический порог и система выйдет из области, где модель хоть как-то адекватна.

2. Деструктивная динамика «цифрового Левиафана»: механизмы накопления риска и потеря управляемости. Фундаментальная некорректность оптимизационной задачи для эргатической системы, рассмотренная в предыдущей главе, не является статическим, пассивным свойством. Она запускает активный, самоусиливающийся процесс деградации, в котором имманентный риск ρ_T не просто существует как константа, а монотонно нарастает во времени, приводя к качественным изменениям в состоянии системы.

Первичным и самым прямым следствием является **дестабилизация и эволюция в неописанные области пространства состояний**. Стратегия u_M^* вычислена, как аргумент максимума функционала $J_M(u)$ на траекториях модельного процесса X_t^M , подчиняющегося уравнению $dX_t^M = f_M(X_t^M, u_t, t)dt + \sigma_M dW_t$. Геометрически, в пространстве $\mathcal{X}_M$ эта стратегия «заставляет» модельную траекторию двигаться в область, где значение J_M максимально. Однако, из-за редукции $\Pi: \mathcal{X}_R \rightarrow \mathcal{X}_M$, эта область в модельном пространстве соответствует не

компактной, четко определенной области в реальном пространстве $\mathcal{X}_R$, а скорее конусу или многообразию высокой размерности. Точнее, прообраз $\Pi^{-1}(\mathcal{A}_M)$ некоторой «оптимальной» области $\mathcal{A}_M \subset \mathcal{X}_M$ в $\mathcal{X}_R$ может быть невыпуклым, несвязным и содержать состояния, качественно различные с точки зрения реальной динамики f_R . Применяя u_M^* , мы насильственно направляем реальный процесс X_t^R в этот прообраз, но поскольку реальная динамика f_R отличается от f_M , траектория X_t^R начинает отклоняться от проекции идеальной модельной траектории. Важно, что это отклонение не является случайным блужданием; оно имеет систематический характер, определяемый разностью операторов дрейфа: $\Delta f(X, u) = f_R(X, u) - f_M(\Pi(X), u)$. Эта разность, которую можно назвать **систематической ошибкой модели**, действует как дополнительное «силовое поле», сдвигающее реальную траекторию относительно модельной. Формально, если записать динамику ошибки $e_t = \widetilde{X}_t^R - X_t^{M,*}$, где $\widetilde{X}_t^R = \Pi(X_t^R)$, то, пренебрегая для простоты стохастическими членами, получим:

$$\dot{e}_t \approx \frac{\partial f_M}{\partial x} \big|_{X_t^{M,*}} e_t + \Delta f(X_t^R, u_M^*).$$

Первое слагаемое описывает линейную неустойчивость модельной траектории, второе – постоянно действующее возмущение из-за неадекватности модели. Поскольку стратегия u_M^* часто выводит систему на границу допустимой области в $\mathcal{X}_M$ (так как оптимум часто лежит на границе), линейная часть $\frac{\partial f_M}{\partial x}$ может иметь положительные собственные значения, делая ошибку e_t растущей даже при малом Δf . Но главное – это сам член Δf . По мере движения в область $\mathcal{A}_M$, где модель особенно плохо описывает реальность (например, в область высокой нагрузки на человеческий компонент, которую модель считает просто увеличением «производительности»), величина $|\Delta f|$ возрастает. Это приводит к тому, что реальная система X_t^R быстро покидает «коридор» $\Pi^{-1}(\mathcal{A}_M)$, в котором модель хотя бы приблизительно адекватна, и выходит в области пространства $\mathcal{X}_R$, вообще не представленные в модели. В этих областях предсказательная сила модели падает до нуля, так как для состояний X^R , для которых $\Pi(X^R)$ далеко от X^M , модель не может дать даже грубой оценки динамики. Управляющие воздействия u_M^* , рассчитываемые на основе текущего модельного состояния X_t^M , становятся не просто неоптимальными, а откровенно вредными, так как основываются на фундаментально неверной «карте местности».

Этот процесс можно описать, как динамику расходящихся траекторий с положительной обратной связью. Введем показатель степени расхождения $d(t) = E[|e_t|^2]$. Его производная по времени будет включать в себя:

1. Член, пропорциональный самой $d(t)$ (из-за возможной неустойчивости).
2. Член, пропорциональный ковариации ошибки и систематического сдвига $\text{Cov}(e_t, \Delta f)$.
3. Член, связанный с ростом нормы Δf .

Поскольку Δf сам зависит от состояния, а состояние все дальше уходит от области адекватности модели, возникает нелинейная зависимость: $|\Delta f| \approx \kappa(d(t))$, где κ – монотонно возрастающая функция. Тогда уравнение для $d(t)$ принимает вид:

$$\frac{d}{dt} d(t) \geq \alpha d(t) + \beta \kappa(d(t)) - \gamma,$$

где $\alpha, \beta > 0$, а γ – возможный демпфирующий член.

При достаточно больших $d(t)$ и возрастающей κ (например, $\kappa(d) \sim d^\delta, \delta > 0$), решение этого неравенства имеет взрывной характер: $d(t) \rightarrow \infty$ за конечное время T^* . Это математическое выражение катастрофического расхождения, когда разрыв между моделью и реальностью становится настолько велик, что система фактически выходит из-под контроля, основанного на этой модели. Момент T^* – точка потери управляемости.

Второй ключевой механизм деструктивной динамики – «эрозия» человеческого компонента и рост системной диссипации. В рамках модели M человеческий фактор часто представлен в предельно редуцированном виде: как источник задержек, шума (аддитивного или параметрического) или как детерминированный преобразователь входных сигналов в действия по фиксированному, хотя, возможно, и стохастическому, правилу. Оптимизация J_M естественным образом направлена на минимизацию негативного влияния этого «шума» и максимизацию «пропускной способности» человеческого звена.

На практике это выражается в:

- стандартизации и алгоритмизации действий оператора, лишаящих его свободы манёвра и возможности применять имплицитное знание¹;
- интенсификации труда для увеличения скорости обработки информации или принятия решений, что ведет к когнитивной перегрузке;
- сокращению «избыточных» с точки зрения модели функций, таких, как взаимный контроль, неформальное общение, создание смыслов, которые не имеют прямого измеримого вклада в J_M .

Формально, если представить человеческий компонент как подсистему с внутренним состоянием $h_t \in \mathcal{H}$ (включающим мотивацию, усталость, компетенцию, уровень доверия), а его вклад в общую динамику системы – через оператор $\psi(h_t, x_t, u_t)$, то в модели происходит грубая аппроксимация: $\psi_M = \psi_{id} + \xi_t$, где ψ_{id} – идеализированный, детерминированный отклик, а ξ_t – гауссов шум. Стратегия u_M^* , оптимизируя систему в целом, будет стремиться снизить зависимость от ψ (как от ненадежного элемента) и минимизировать дисперсию $E[|\xi_t|^2]$. В реальности это приводит к тому, что состояние h_t деградирует по определенному закону. Можно смоделировать эту деградацию как динамический процесс:

$$\dot{h}_t = Ah_t + Bu_t^{human} + \eta_t,$$

где u_t^{human} – управляющие воздействия, направленные на человека со стороны системы (задачи, инструкции, стимулы), вычисленные в соответствии со стратегией u_M^* .

Матрица A описывает внутреннюю динамику человеческого состояния (например, усталость накапливается, мотивация снижается без подкрепления). Оптимальная стратегия u_M^* порождает последовательность $\{u_t^{human}\}$, которая максимизирует краткосрочный вклад в J_M , но игнорирует долгосрочную динамику h_t . Фактически, u_t^{human} становится эксплуататорским, истощающим ресурс h_t . Например, если h_t включает запас «внимания» или «психической энергии», то u_t^{human} постоянно требует высокого уровня этого ресурса, не предусматривая периодов для его восстановления. Со временем это приводит к переходу состояния h_t в область низких значений (выгорание, цинизм, потеря квалификации). В этой области реальный оператор $\psi(h_t, \cdot)$ перестает быть даже приблизительно похожим на модельный ψ_M : вместо того, чтобы быть источником малого шума, он становится источником крупных, систематических ошибок или пассивно-сопротивляющимся элементом. Это напрямую увеличивает систематическую ошибку модели Δf , так как теперь f_R зависит от $\psi(h_t)$, а f_M – от ψ_M . Возникает петля положительной обратной связи: применение u_M^* ухудшает h_t , ухудшение h_t увеличивает Δf , увеличение Δf ускоряет расхождение траекторий $d(t)$, что, в свою очередь, может приводить к еще более агрессивным и деструктивным с точки зрения

¹ Имплицитное знание – неосознаваемое, скрытое знание или навык, который человек приобретает без явных инструкций и не может легко вербализовать, но применяет на практике. Оно формируется через «имплицитное научение», когда мозг усваивает закономерности, например, при езде на велосипеде, изучении языка или восприятии дизайна, и проявляется как интуитивное понимание, а не как набор фактов.

человека управляющим воздействиям со стороны модели, стремящейся «вернуть систему на курс».

Понятие **диссипации** здесь приобретает глубокий смысл. Если в термодинамике диссипация – это рассеяние энергии, превращение упорядоченной формы энергии в неупорядоченную (тепло), то в контексте эргатической системы, управляемой «цифровым Левиафаном», диссипация проявляется как рассеяние **системного потенциала**, который включает в себя не только материально-энергетические ресурсы, но и социальный капитал, доверие, адаптивный резерв, креативность человеческого компонента. Формально, можно ввести функционал системного потенциала $\Psi(t) = \Psi_S(X_t^R, h_t)$, отражающий способность системы к выполнению своих истинных, многоаспектных целей J_R . В нормальных условиях часть ресурсов тратится на поддержание этого потенциала (обучение, коммуникация, создание доверия, поддержание избыточности). Стратегия, оптимизирующая узкий критерий J_M , воспринимает эти затраты как неэффективные и стремится их минимизировать, перенаправляя ресурсы на непосредственное увеличение J_M . Это ведет к монотонному снижению потенциала: $d\Psi/dt < 0$. Скорость снижения может описываться уравнением:

$$\frac{d\Psi}{dt} = -\lambda_J \cdot (J_M(t) - J_M^{target}) - \lambda_\Psi \cdot \Psi(t) + v(t),$$

где первый член отражает «перекачку» ресурсов из потенциала в текущий модельный выход (чем агрессивнее оптимизация, тем больше отток), второй – естественную деградацию потенциала при отсутствии инвестиций, а $v(t)$ – случайные или внешние пополнения.

Решение такого уравнения при $\lambda_J > 0$ приводит к экспоненциальному затуханию $\Psi(t)$. Система, лишенная потенциала, теряет устойчивость (resilience): ее способность поглощать возмущения и восстанавливаться падает. Малое отклонение, которое раньше было бы компенсировано, теперь вызывает лавинообразный сбой. Она теряет адаптивность: для обучения и изменения уже нет свободных ресурсов и мотивации. Наконец, падает легитимность: участники системы перестают верить в ее справедливость и осмысленность, что ведет к росту транзакционных издержек, оппортунизму и, в пределе, к отказу от сотрудничества. Все эти потери напрямую увеличивают имманентный риск ρ_T , так как делают поведение системы все более непредсказуемым и все менее соответствующим простой, детерминированной логике модели M .

Третий механизм – формирование порочных циклов [6] и режимов с обострением [7].

Описанные выше процессы не независимы; они переплетаются, усиливая друг друга.

Кульминацией этой деструктивной динамики является катастрофическое расхождение и потеря управляемости. Точку, где система окончательно выходит из-под контроля, можно определить несколькими способами.

Первый – достижение расхождением $d(t)$ некоторого критического порога d_{crit} , после которого ни одно управляющее воздействие из допустимого множества $\mathcal{U}$, рассчитанное на основе модели M , не может вернуть систему в область, где модель хотя бы приблизительно адекватна. Формально, это момент T^* , когда множество достижимости реальной системы $\mathcal{R}_{\mathcal{R}}(T^*)$ под действием стратегий, основанных на M , полностью отделено от прообраза области адекватности модели $\Pi^{-1}(\mathcal{A}_{adm})$.

Второй способ – достижение потенциалом $\Psi(t)$ критически низкого уровня Ψ_{crit} , после которого система теряет структурную целостность: начинается массовый выход человеческих агентов, отказы технологических компонентов из-за отсутствия обслуживания, распад кооперативных связей. В этот момент система перестает быть управляемой в принципе, так как перестает существовать как целостный объект управления.

Третий способ – возникновение в системе новых, эмерджентных режимов, которые не только не описаны в модели, но и прямо противоречат ее базовым принципам. Например, вместо кооперации возникает жесткая конкуренция за истощающиеся ресурсы, вместо решения задач – их имитация, вместо инноваций – регресс к примитивным, но энергосберегающим практикам. Эти эмерджентные режимы обладают собственной, часто хаотической, динамикой, которую первоначальная модель M и связанная с ней стратегия u_M^* абсолютно не способны ни предсказать, ни парировать.

Таким образом, «цифровой Левиафан», начав с решения, казалось бы, рациональной задачи оптимизации, запускает лавинообразный процесс саморазрушения эргатической системы. Имманентный риск ρ_T , изначально присутствующий как следствие редукции, не остается постоянным, а становится функцией времени $\rho_T(t)$, монотонно растущей, часто нелинейно или даже гиперболически. Движущей силой этого роста являются не внешние шоки, а внутренние противоречия, заложенные в самой попытке подчинить живую, рефлексивную, многомерную реальность логике упрощенной симуляции. Система не просто отклоняется от плана; она активно трансформируется под воздействием этого «плана», и трансформируется в сторону большей хрупкости, меньшей адаптивности и, в конечном счете, к состоянию, где понятия «оптимальности» и «управления» теряют всякий смысл. Как показано в работе [5], анализ подобных деструктивных сценариев требует перехода от статических оценок безопасности к динамическим моделям, учитывающим эволюцию системного потенциала и скорости накопления риска. Выход из этой ловушки требует не уточнения модели в рамках старой парадигмы, а радикальной смены самой цели управления – от максимизации узкого функционала в симуляции к постоянному управлению и минимизации имманентного риска расхождения. Как это может быть реализовано в рамках новой парадигмы симбиотического управления, будет рассмотрено далее.

3. Формализм риска: от декомпозиции неопределенности к синтезу робастных стратегий. Кульминацией критического анализа «цифрового Левиафана» является не отрицание моделирования как такового, но осознание необходимости принципиально иной теоретической основы для управления эргатическими системами. Эта основа, обозначенная как формализм риска, предлагает совершить эпистемологический поворот: вместо того чтобы рассматривать риск как побочный продукт неточной модели, принять его имманентность в качестве центрального объекта управления. В рамках этого подхода цель смещается с поиска условного оптимума $u_M^* = \arg \max J_M(u)$ для неполной модели M к синтезу стратегий, минимизирующих функционал, явным образом измеряющий расхождение между моделью и реальностью. Однако для конструктивного синтеза таких стратегий необходима более тонкая, чем в предыдущих главах, декомпозиция самой природы неопределенности, порождающей риск.

Первый шаг формализма риска заключается в переходе от рассмотрения единой модели M к работе с ансамблем моделей или, что более фундаментально, с вероятностной мерой на пространстве моделей. Это признает тот факт, что наше знание об эргатической системе R принципиально неполно и может быть представлено лишь в виде распределения над возможными её описаниями. Пусть Θ – пространство параметров, определяющих класс допустимых моделей. Вместо единственной модели M_{θ_0} мы рассматриваем апостериорное распределение $p(\theta|\mathcal{D})$, где $\mathcal{D}$ – исторические и текущие данные. Тогда прогноз поведения системы и оценка последствий управления должны учитывать всю эту неопределенность. Целевой функционал, который в парадигме Левиафана был детерминированным $J_M(u)$, теперь сам становится случайной величиной, зависящей от θ : $J(u, \theta)$. Однако ключевое отличие формализма риска состоит в том, что мы отказываемся от простого взятия математического

ожидания $E_\theta[J(u, \theta)]$ в качестве нового критерия. Такой подход (стандартный в байесовской оптимизации) по-прежнему стремится к некоему «усредненному» оптимуму, который может оказаться столь же деструктивным, если распределение $p(\theta)$ широко или смещено.

Вместо этого формализм риска предлагает работать непосредственно с функционалом риска $\rho(u)$, который количественно измеряет потенциальный ущерб от расхождения между предсказаниями ансамбля моделей и реальностью.

Один из способов построения такого функционала – это определение риска как условной дивергенции между прогнозным и эмпирическим распределениями траекторий. Пусть P_θ^u – распределение траектории системы на горизонте T при управлении u , предсказанное моделью с параметрами θ . Пусть Q^u – истинное (неизвестное) распределение реальных траекторий. Мы никогда не знаем Q^u , но мы можем наблюдать эмпирические реализации $\{\xi^{(i)}\}_{i=1}^N$, порожденные реальной системой под управлением u . Тогда риск можно определить как верхнюю границу некоторой дивергенции D между прогнозными распределениями из нашего ансамбля и эмпирическим распределением, взятым по всем моделям, согласующимся с данными в пределах заданной доверительной области. Например, используя дивергенцию Реньи [8] порядка α :

$$\rho_\alpha(u) = \sup_{\theta \in \Theta_\varepsilon} D_\alpha(P_\theta^u | \widehat{Q}_N^u),$$

где $\Theta_\varepsilon = \{\theta: d(P_\theta^u, \widehat{Q}_N^u) \leq \varepsilon\}$ – множество параметров, чьи прогнозы не слишком противоречат эмпирическим данным в смысле некоторой метрики d , а $\widehat{Q}_N^u$ – эмпирическое распределение.

Такой функционал измеряет наихудшее возможное расхождение среди всех моделей, которые ещё не отвергнуты данными. Минимизация $\rho_\alpha(u)$ по u приводит к стратегии, которая старается сделать систему такой, чтобы даже самые пессимистичные из ещё допустимых моделей предсказывали поведение, близкое к наблюдаемому. Это есть не что иное, как формализация принципа минимакса применительно к эпистемической неопределенности.

Однако подобный подход, будучи теоретически строгим, зачастую вычислительно неосуществим. Практической альтернативой является введение модели рассогласования $\Delta(x, u)$, которая явно оценивает ошибку прогноза основной модели M в точке состояния x при управлении u . Эта модель может быть параметрической, $\Delta_\phi(x, u)$, и обучаться на исторических ошибках прогнозирования:

$$\phi^* = \arg \min_{\phi \in \Sigma_k} |x_{k+1} - (f_M(x_k, u_k) + \Delta_\phi(x_k, u_k))|^2.$$

В этом случае общая динамика системы представляется как:

$$x_{k+1} = f_M(x_k, u_k) + \Delta_{\phi^*}(x_k, u_k) + w_k,$$

где w_k – остаточный шум.

Риск тогда можно связать с величиной $|\Delta_{\phi^*}(x, u)|$ или, более точно, с ковариацией её оценки. Если модель рассогласования также выдает оценку неопределенности $\Sigma_\Delta(x, u)$, то функционал риска может быть определен, как ожидаемая норма ошибки с учетом этой неопределенности:

$$\rho(u) = E \left[\int_0^T |\Delta_{\phi^*}(x_t, u_t)|_{\Sigma_\Delta^{-1}(x_t, u_t)}^2 dt \right],$$

где норма взвешена обратной ковариацией, что придает больший вес областям, где модель рассогласования сама ненадежна. Задача управления тогда формулируется как:

$$u_\rho^* = \arg \min_{u \in \mathcal{U}} (\rho(u) + \lambda \cdot J_M(u)),$$

где $J_M(u)$ – традиционный функционал эффективности, а λ – параметр, регулирующий компромисс между минимизацией риска и достижением цели.

Это уже классическая задача стохастической оптимального управления, но с принципиально иным составным критерием.

Второй ключевой элемент формализма риска – явное разделение эпистемической и алеаторной неопределенности. Алеаторная неопределенность – это принципиально неустранимый шум, присущий самой системе (например, колебания спроса, погодные условия). Эпистемическая неопределенность – это незнание, проистекающее из неполноты модели, и она может быть сокращена путем сбора данных и обучения. Формализм риска фокусируется на управлении именно эпистемической неопределенностью, так как она является источником имманентного риска расхождения. В терминах байесовских глубоких нейронных сетей или гауссовских процессов это соответствует разделению предсказательной дисперсии на две компоненты.

Пусть для данной точки (x, u) модель предсказывает распределение следующего состояния $p(x'|x, u)$. Его общую дисперсию $\text{Var}[x']$ можно декомпозировать (в приближении) как:

$$\text{Var}[x'] \approx \underbrace{E_{\theta}[\text{Var}[x'|\theta]]}_{\text{Алеаторная}} + \underbrace{\text{Var}_{\theta}[E[x'|\theta]]}_{\text{Эпистемическая}},$$

где усреднение ведется по апостериорному распределению параметров $p(\theta|\mathcal{D})$.

Функционал риска тогда естественно связать с эпистемической компонентой, так как она отражает наше незнание правильной динамики. Стратегия, минимизирующая риск, будет стремиться выбирать такие управления u , которые уменьшают эпистемическую дисперсию, то есть вести систему в области, где наши модели более надежны, или, что эквивалентно, активно исследовать те области, где неопределенность велика, чтобы сократить её. Это приводит к фундаментальному пересмотру логики управления: не гнаться за сиюминутной эффективностью в области, где модель уверена (но, возможно, ошибается), а балансировать между эксплуатацией текущего знания и исследованием для его улучшения.

Наиболее перспективным направлением синтеза робастных стратегий в рамках формализма риска представляется стохастическое модельно-прогнозирующее управление (Stochastic Model Predictive Control, SMPC) с явным учетом ограничений на риск. В классической постановке на каждом шаге решается конечно-горизонтная задача оптимизации, первое управление применяется, и процесс повторяется. В SMPC учитывается стохастичность модели, и ограничения часто задаются в вероятностной форме. Формализм риска предлагает заменить вероятностные ограничения на ограничения по риску.

Сложность решения такой задачи заключается в том, что даже оценка риска ρ_k для заданной последовательности управлений является нетривиальной вычислительной проблемой, требующей методов Монте-Карло или аппроксимаций. Однако развитие методов глубокого обучения с подкреплением с учетом неопределенности [9] (например, на основе байесовских нейросетей или ансамблей) открывает пути для приближенного, но практически реализуемого синтеза таких стратегий.

Отдельного рассмотрения заслуживает проблема включения человеческого компонента в формализм риска. В парадигме «цифрового Левиафана» человек – источник шума. В формализме риска он должен стать активным источником информации для сокращения эпистемической неопределенности и корректирующим механизмом. Математически это можно представить как расширение пространства состояний системы x_t до (x_t, b_t) , где b_t – состояние доверия или убеждений оператора относительно адекватности модели. Динамика b_t описывает, как оператор обновляет свои убеждения, наблюдая расхождения. Стратегия управления теперь должна зависеть и от (b_t) : $(u_t = \pi(x_t, b_t))$. Более того, в контуре управления появляется второе управляющее воздействие – корректирующий сигнал оператора v_t ,

который может напрямую модифицировать управление, предлагаемое моделью: $u_t^{applied} = u_t^{model} + v_t$, или даже модифицировать саму модель: $\theta_{t+1} = \theta_t + g(v_t, data)$. Задача тогда становится игрой двух агентов (модели и человека) с общей, но не полностью совпадающей целью. Общая цель – минимизация долгосрочного риска ρ . Но у модели есть доступ только к J_M и оценке ρ через свои алгоритмы, а у человека – к качественным, контекстуальным знаниям и интуиции, которые не формализованы. Синтез совместной стратегии может быть описан в рамках теории совместной оптимизации или иерархического управления, где политика высокого уровня (человек) вмешивается в политику низкого уровня (модель) при превышении порога локального риска, оцениваемого человеком на основе своего b_t .

Это естественным образом приводит к **концепции симбиотического управления**, являющейся практической реализацией формализма риска. Его ядро – итеративный диалог, который можно формализовать, как цикл Байесовского обновления, но не только параметров модели, а всей структуры взаимодействия. Такой итеративный процесс принципиально сужает эпистемическую неопределенность, но не за счет создания всё более сложной модели, а за счет динамического уточнения границ её применимости и обогащения её контекстом от человеческого оператора. Риск ρ_t становится не объектом слепого минимизации, а управляющей переменной, значение которой поддерживается в допустимых пределах за счет диалога. Это перекликается с идеями адаптивного управления и управления по возмущениям, но на более высоком, мета-уровне, где возмущением является собственное несовершенство модели.

4. Архитектура симбиотического управления: итеративный диалог как механизм сжатия эпистемической неопределенности. Формализм риска, устанавливающий новую философию управления, требует для своей практической реализации конкретной системной архитектуры. Такой архитектурой, способной обеспечить динамический баланс между алгоритмической эффективностью и человеческой интуицией, является **симбиотическое управление**. Его сущность заключается не в простом соседстве человека и машины в контуре управления, а в организации между ними интенсивного, структурированного итеративного диалога, где каждый акт коммуникации приводит к сужению эпистемической неопределенности и уменьшению имманентного риска. Остановимся более подробно на построении математического аппарата и принципов функционирования такого диалога, рассмотрению его динамики, устойчивости и пределов эффективности.

Центральным объектом симбиотической системы является не статичная модель, а динамический эпистемический контур, состояние которого в момент времени t описывается кортежем

$$E_t = (\mathcal{M}_t, \mathcal{B}_t, \mathcal{C}_t, \rho_t),$$

где $\mathcal{M}_t$ – текущая агностическая модель системы, представленная, например, параметрическим семейством $\{f_{\theta_t}, \Sigma_{\theta_t}\}$, где $\theta_t \in \Theta$; $\mathcal{B}_t$ – модель убеждений человеческого оператора относительно адекватности $\mathcal{M}_t$ (это может быть распределение над некоторым пространством гипотез $\mathcal{H}$ или просто вектор доверия $b_t \in [0,1]^K$, где каждая компонента отражает уверенность в определенном аспекте модели (например, точность прогноза в разных режимах)); $\mathcal{C}_t$ – контекстуальный слой, содержащий неформализуемую или слабо формализуемую информацию, доступную оператору: социальную атмосферу, неписанные правила, слухи, интуитивные предчувствия, этические соображения (формально $\mathcal{C}_t$ можно рассматривать как точку в многообразии высокоразмерных качественных данных); $\rho_t: \mathcal{U} \rightarrow R^+$ – текущий оператор риска, оценивающий ожидаемое расхождение для любого допустимого управления u , с учетом $\mathcal{M}_t$, $\mathcal{B}_t$ и, возможно, $\mathcal{C}_t$.

Динамика всей эргатической системы теперь описывается не одним уравнением, а связанной системой уравнений для физического состояния $x_t \in \mathcal{X}_{\mathcal{R}}$ и эпистемического состояния E_t :

$$\begin{aligned} x_{t+1} &= F(x_t, u_t, w_t; \mathcal{M}_t) + \varepsilon_t, \\ E_{t+1} &= G(E_t, x_t, u_t, y_{t+1}, v_t; \Omega). \end{aligned}$$

Здесь F – реальная, неизвестная динамика, которая в нашей парадигме аппроксимируется моделью $\mathcal{M}_t$, но содержит неустранимую ошибку ε_t . Функция G – закон эпистемического обновления, описывающий, как взаимодействие с реальностью (наблюдение нового состояния x_{t+1} или измерений y_{t+1}) и коррекция оператора v_t изменяют наше знание и доверие. Ω – набор мета-параметров, определяющих «скорость обучения» и «степень доверия» в диалоге.

Сердцем симбиоза является процесс генерации и коррекции управляющего воздействия. Он реализуется как двухуровневая оптимизация. На нижнем уровне алгоритмический агент решает задачу, сформулированную в рамках формализма риска:

$$u_t^0 = \arg \min_{u \in \mathcal{U}} \{J_{\mathcal{M}_t}(u) + \lambda_t \cdot \rho_t(u)\},$$

где λ_t – адаптивный коэффициент, регулирующий баланс между эффективностью (первый член) и осторожностью (второй член).

Ключевое отличие – функционал риска ρ_t учитывает не только неопределённость модели $\mathcal{M}_t$, но и степени доверия к ней $\mathcal{B}_t$. Если доверие к прогнозам в какой-либо области низкое, ρ_t будет штрафовать управляющие воздействия, ведущие систему в эту область.

Полученное оптимальное управление u_t^0 вместе с аналитикой (прогноз, оценка неопределённости, уязвимости модели) представляется оператору в виде структурированного предложения $\mathcal{P}_t$. На верхнем уровне человеческий оператор, обладая состоянием $\mathcal{B}_t$ и контекстом $\mathcal{C}_t$, производит оценку предложения $\mathcal{P}_t$. Он выступает как байесовский процессор с собственной неформализуемой моделью системы. Сравнивая прогноз со своей мысленной симуляцией, оператор вычисляет субъективную оценку риска $\rho_t^{\widehat{human}}$ и выявляет контекстуальные факторы, снижающие надёжность модели. Его задача – не просто отклонить предложение, а сформулировать структурированную коррективу v_t из заданного пространства $\mathcal{V}$, минимизирующую коммуникационные потери. Это может быть:

- аддитивная или мультипликативная поправка к управлению;
- семантическая коррекция (изменение приоритета целей);
- контекстуальная аннотация (добавление в модель нового ограничения).

Оператор выбирает корректив v_t^* , минимизирующий его субъективную оценку общего риска:

$$v_t^* = \arg \min_{v \in \mathcal{V}} \rho_t^{\widehat{human}}(\Phi(u_t^0, v), \mathcal{C}_t),$$

где Φ – оператор применения коррективы к исходному управлению (например, $\Phi(u, v) = u + v$). Итоговое управление, применяемое к системе, есть $u_t = \Phi(u_t^0, v_t^*)$.

После применения u_t система переходит в состояние x_{t+1} . Ошибка прогноза служит сигналом для **эпистемического обновления** – обучения модели на основе реального исхода. Описанный цикл (предложение, коррекция, применение, обновление) представляет собой одну итерацию эпистемического диалога.

Эффективность симбиоза оценивается по скорости снижения среднего реализованного риска: $\overline{\rho}_T = \frac{1}{T} \sum_{t=0}^{T-1} \rho_t(u_t) \rightarrow \rho_{min}$ при $T \rightarrow \infty$, где ρ_{min} – неустранимый уровень неопределённости. Скорость сходимости зависит от обучаемости модели и качества человеческих корректив.

Математически процесс можно представить как двух-агентное обучение с подкреплением, где наградой служит отрицательный риск: $r_t = -\rho_t(u_t)$. Политика алгоритмического агента параметризована π_ϕ^A , политика человеческого агента – π_ξ^H , где ξ – его неформализуемый опыт. Оптимизация направлена на нахождение точки равновесия, возможно, седловой, в игре с общей, но неполностью наблюдаемой наградой. Сложность заключается в нестационарности: политика π_ξ^H меняется по мере изменения доверия B_t , а политика π_ϕ^A меняется по мере обучения модели.

Важнейшим вопросом является устойчивость симбиотического контура. Существует опасность возникновения патологических режимов. Для предотвращения этих режимов необходимы метаправила управления самим диалогом. Архитектурно симбиотическая система требует специальных интерфейсов, которые превращают неявное знание оператора в структурированные коррективы. Симбиотическое управление – это не статичная архитектура, а мета-архитектура для «выращивания» совместной познавательной системы. Её ядро – итеративный диалог – представляет собой алгоритм с внешним оракулом (человеком), где каждая итерация выполняет два действия: улучшает формальную модель за счет данных и формализует часть контекстуального знания за счет человеческих коррективов. Динамическим инвариантом этого процесса является имманентный риск ρ_t , который служит и целью, и мерой прогресса. В пределе, успешный симбиоз ведет не к созданию идеальной модели, а к установлению устойчивого режима ко-эволюции, где модель и оператор совместно адаптируются к меняющемуся миру, постоянно поддерживая риск в приемлемых границах. Это превращает управление из задачи оптимизации в задачу совместного познания и адаптации, что является единственно адекватным ответом на вызовы, порождаемые сложностью эргатических систем.

5. Практическая имплементация и вызовы симбиотического управления: от теории к инженерии социотехнических систем. Теоретический каркас симбиотического управления, построенный на принципах формализма риска и архитектуре итеративного диалога, должен быть переведён в плоскость практической реализации. Эта задача сопряжена с комплексом междисциплинарных вызовов, лежащих на стыке теории управления, машинного обучения, когнитивных наук, человеко-машинного взаимодействия (HCI) и организационного проектирования [10].

5.1. Математическое и вычислительное ядро: от байесовских ансамблей к обучающимся картам риска. Практическая реализация эпистемического контура требует выбора конкретных математических представлений для модели $\mathcal{M}_t$, оператора риска ρ_t и механизма обновления. Ключевой дилеммой является баланс между выразительностью (способностью улавливать сложные зависимости) и вычислительной эффективностью (возможностью работать в режиме реального времени).

Для представления агностической модели $\mathcal{M}_t$ наиболее перспективным представляется использование ансамблей моделей или байесовских глубоких сетей. В первом случае $\mathcal{M}_t = \{f_{\theta^{(i)}}\}_{i=1}^N$, где каждый $\theta^{(i)}$ обучается на бутстраппированной выборке данных или с разными начальными условиями. Прогноз и неопределенность выводятся из статистики по ансамблю: $\widehat{x}_{t+1} = \frac{1}{N} \sum_i f_{\theta^{(i)}}(x_t, u_t)$, а эпистемическая дисперсия оценивается как $\Sigma_t^{epist} = \text{Var}_i[f_{\theta^{(i)}}(x_t, u_t)]$. Такой подход, хотя и требует в N раз больше вычислений, прозрачен и хорошо калибруется. Байесовские нейросети, использующие приближенные методы вывода вроде вариационного вывода или MC-Dropout [11], позволяют получить оценку неопределенности в рамках одной сети, но могут страдать от недооценки эпистемической

неопределенности. Выбор между ними зависит от критичности времени отклика и доступных вычислительных ресурсов.

Оператор риска ρ_t должен агрегировать информацию о неопределенности по всей прогнозируемой траектории. Пусть для предложенного управления $u_{t:t+H-1}$ модель генерирует распределение траекторий $p(x_{t+1:t+H}|x_t, u_{t:t+H-1}, \mathcal{M}_t)$. Тогда интегральный риск можно определить как математическое ожидание некоторой функции потерь L_{risk} , зависящей от состояния и его неопределенности, вдоль траектории:

$$\rho_t(u_{t:t+H-1}) = E_{p(x_{t+1:t+H}|\dots)} \left[\sum_{k=1}^H \gamma^k L_{risk}(x_{t+k}, \Sigma_{t+k}^{epist}) \right].$$

Функция L_{risk} должна штрафовать как попадание в нежелательные области фазового пространства (например, близость к границам безопасной работы), так и высокую эпистемическую неопределенность. Один из вариантов – использовать стоимость, зависящую от дисперсии: $L_{risk}(x, \Sigma) = x^T Q x + \text{Tr}(S \Sigma)$, где матрица S определяет «цену незнания». Вторая компонента прямо стимулирует выбор управлений, ведущих в области с низкой эпистемической неопределенностью. Задача оптимального управления тогда становится задачей стохастического MPC с составной стоимостью, которая может решаться методами стохастической оптимизации (например, на основе сэмплирования траекторий – Model Predictive Path Integral, MPPI [12]) или аппроксимации динамического программирования с глубоким обучением [13].

Отдельная сложность – эффективное обновление модели $\mathcal{M}_t$ в реальном времени после получения новой пары данных (x_t, u_t, x_{t+1}) . Полный пересчёт байесовского апостериорного распределения непрактичен. Необходимы инкрементальные методы. Для ансамблей можно применять онлайн-обучение каждого члена ансамбля на потоке данных с небольшой скоростью обучения. Для байесовских сетей существуют методы онлайн-вариационного вывода, где вариационные параметры ϕ_t (например, средние и дисперсии весов) обновляются по градиенту логарифмического правдоподобия нового наблюдения с учётом априорного распределения, играющего роль регуляризатора, предотвращающего катастрофическую забывчивость:

$$\phi_{t+1} = \phi_t + \eta \nabla_{\phi} [\log p(x_{t+1}|x_t, u_t, \phi_t) - D_{KL}(q_{\phi_t}(\theta) || p(\theta))].$$

Человеческие коррективы v_t должны также интегрироваться в процесс обучения. Если корректив был семантическим (например, «учитывать фактор Z »), система должна попытаться найти латентную переменную z_t , такую, что добавление её в модель уменьшает ошибку прогноза. Это можно сформулировать как задачу обучения с частичным привлечением учителя: по данным $\{(x_t, u_t, x_{t+1})\}$ и меткам-подсказкам от оператора в моменты вмешательства, обучить модель $f_{\theta}(x, u, z)$ и энкодер $q_{\psi}(z|x, u, \text{hint})$, предсказывающий латентную переменную.

5.2. Моделирование и интеграция человеческого компонента: от когнитивных архитектур к структурированным интерфейсам. Самой сложной частью является формализация и интеграция человеческой составляющей. Оператор – не детерминированный алгоритм, его состояние $\mathcal{B}_t$ и контекст $\mathcal{C}_t$ динамичны и подвержены когнитивным искажениям. Главный инженерный вызов – проектирование интерфейса передачи контекстуального знания с минимальными потерями и искажениями. Традиционные интерфейсы (кнопки, графики) не годятся для выражения качественных нюансов. Необходимы интерфейсы нового поколения. Для предотвращения когнитивной перегрузки оператора система должна сама определять моменты, когда вмешательство критически необходимо. Это можно сделать, отслеживая «сигналы беспокойства»: быстрый рост

эпистемической дисперсии $\frac{d}{dt} \text{Tr}(\Sigma_t^{\text{epist}})$, приближение траектории к ранее неисследованным областям фазового пространства, или расхождение в прогнозах между разными членами ансамбля моделей. Только в такие моменты система должна запрашивать активное внимание оператора, в остальное время работая автономно.

5.3. Организационные и этические аспекты: распределение ответственности и обучение культуре симбиоза. Симбиотическая система не существует в вакууме – она встроена в организацию со своей иерархией, нормами и культурой. Ключевой вопрос: распределение ответственности за решения. Если управляющее воздействие u_t является результатом коррективы оператора v_t^* к рекомендации модели u_t^0 , то кто отвечает в случае аварии? Оператор, утвердивший решение? Разработчик модели, чья рекомендация была ошибочной? Создатель интерфейса, который не позволил оператору правильно выразить свои опасения? На практике необходимо юридическое и организационное закрепление принципа совместной ответственности с чётким протоколом анализа происшествий, который реконструирует эпистемическое состояние E_t на момент принятия решения.

Это требует ведения полного протокола диалога, включающего не только последовательность $(u_t^0, v_t^*, u_t, x_{t+1})$, но и метаданные: прогнозы с неопределённостью, визуализации, представленные оператору, его аннотации, оценку субъективного риска ρ_t^{human} . Такой протокол позволит впоследствии определить, была ли ошибка следствием «слепого пятна» модели, которую оператор не заметил, или же оператор проигнорировал явные предупреждения системы. Другим организационным вызовом является подготовка операторов. Они должны быть не просто пассивными наблюдателями, а активными «напарниками» искусственного интеллекта. Этические аспекты включают проблему смещения в симбиозе. Если оператор обладает неосознанными предубеждениями (например, систематически не доверяет рекомендациям в определённых ситуациях из-за стереотипов), его коррективы могут «обучить» модель этим же предубеждениям, закрепляя их в алгоритме. Необходимы механизмы обнаружения таких паттернов, например, статистический анализ корреляции между характеристиками ситуаций и частотой/направленностью человеческих вмешательств, с последующей калибровкой.

5.4. Области применения и границы парадигмы. Симбиотическое управление наиболее востребовано в областях с высокой степенью неопределённости, где формальные модели заведомо неполны, а цена ошибки велика. Однако существуют и границы применимости. Симбиотическое управление малоэффективно в двух крайних случаях.

Во-первых, в полностью формализуемых, детерминированных системах, где модель может быть исчерпывающе точной (например, управление роботом в контролируемой среде). Здесь человеческое вмешательство только замедлит работу.

Во-вторых, в системах, где человеческий компонент настолько доминирует, а его поведение настолько иррационально или непредсказуемо, что любая формализация бессмысленна (например, управление творческим коллективом в момент кризиса). Здесь модель не сможет построить даже приближённо адекватное представление.

Оптимальная область – это зона сложности, где существует устойчивая, но неполная регулярность, а человеческая интуиция служит мощным, но не абсолютным источником информации о латентных переменных.

5.5. Выводы: путь к инженерии доверия. Практическая реализация симбиотического управления – это не просто техническая задача создания алгоритмов и интерфейсов. Это комплексная задача инженерии доверия между человеком и машиной. Её успех зависит от создания устойчивого, самообучающегося контура. Ключевым метрическим показателем

успеха такой системы является не только итоговая эффективность J_R , но и динамика эпистемической неопределенности – скорость, с которой сокращается разрыв между моделью и реальностью, и уровень доверия в системе, измеряемый через стабильность и адекватность убеждений B_t . Таким образом, переход от парадигмы «цифрового Левиафана» к симбиотическому управлению представляет собой не просто смену алгоритмов, а глубокую трансформацию самой логики взаимодействия с технологией в сложном мире. Это путь от иллюзии абсолютного контроля к мудрости совместной навигации в условиях принципиальной неполноты знания, где риск перестаёт быть угрозой и становится общим компасом для человека и машины.

Заключение. Цифровизация управления – необратимый тренд, однако искушение построить «цифрового Левиафана», полагающегося на иллюзию абсолютного знания и тотальной оптимизации, является опасной ловушкой. Такая парадигма превращает риск из внешней угрозы во внутреннее, деструктивное свойство системы, неизбежно ведущее к дестабилизации, эрозии человеческого компонента и потере управляемости. Выход лежит в смене фундаментальной философии: мы должны отказаться от погони за призрачным оптимумом в виртуальном мире и честно признать принципиальную ограниченность наших моделей. Предлагаемый формализм риска предлагает сделать эту ограниченность предметом явного управления, а парадигма симбиотического управления выступает практическим механизмом его реализации, выстраивая непрерывный диалог между человеческой интуицией и машинным расчетом. Устойчивое будущее лежит через «приручение» цифровых двойников, превращая их из самодовлеющих симулякров в инструменты, служащие подлинным целям живой, адаптивной и человеко-центрированной эргатической системы. Это путь от иллюзии контроля к мудрости совместной навигации в условиях принципиальной неполноты знания, где риск становится не угрозой, а общим компасом.

Список источников

1. Бочков А.В. На пути к формализму риска. Размышления о риске и его природе / А.В. Бочков, В.В. Лесных. – Вологда: Инфра-Инженерия, 2024. – 296 с. – EDN: MXDJTV.
2. Гастев Ю.А. Гомоморфизмы и модели. Логико-алгебраические аспекты моделирования. – М.: Наука, 1975. – 152 с.
3. Пуговкина А.С. Имплицитные знания как аспект интуиции и практического интеллекта // Евразийский Союз Ученых, 2015. – № 3-7 (12). – URL: <https://cyberleninka.ru/article/n/implitsitnye-znaniya-kak-aspekt-intuitsii-i-prakticheskogo-intellekta> (дата обращения: 07.01.2026).
4. Журавлев В.Ф. Некорректные задачи механики // Вестник МГТУ им. Н.Э. Баумана. Серия «Приборостроение», 2017. – № 2 (113). – URL: <https://cyberleninka.ru/article/n/nekorrektnye-zadachi-mehaniki> (дата обращения: 07.01.2026). – DOI:10.18698/0236-3933-2017-2-77-85.
5. Бочков А.В. На пути к формализму риска. Безопасность и устойчивость функционирования умных экспансивных систем. – Вологда; М.: Инфра-Инженерия, 2025. – 348 с. – EDN: VQVVDB.
6. Belikov A. Age-related diseases as vicious cycles. Preprints, 2017, DOI: 10.20944/preprints201704.0102.v1.
7. Самарский А.А. Режимы с обострением в задачах для квазилинейных параболических уравнений / А.А. Самарский, В.А. Галактионов, С.П. Курдюмов, А.П. Михайлов. – М.: Наука, 1987. – 470 с.
8. Прелов В.В. f-дивергенция и склеивание вероятностных распределений // Проблемы передачи информации, 2021. – Т. 57. – № 1. – С. 64–80. – DOI:<https://doi.org/10.31857/S0555292321010034>.
9. Шереужев М.А. Модификация алгоритма глубокого обучения для распределения функций и задач между робототехническим комплексом и человеком в условиях неопределенности и переменности окружающей среды / М.А. Шереужев, У. Го, В.В. Серебряный // Известия КБНЦ РАН, 2024. – № 6. – URL: <https://cyberleninka.ru/article/n/modifikatsiya-algoritma-glubokogo-obucheniya-dlya-raspredeleniya-funktsiy-i-zadach-mezhdu-robototekhnicheskim-kompleksom-i> (дата обращения: 07.01.2026).
10. Кучкаров З.А. Методы концептуального анализа и синтеза в теоретическом исследовании и проектировании социально-экономических систем: учебное пособие: [курс лекций: для слушателей специальности «Концептуальное проектирование систем организационного управления»: в 2 т.]. – Т. 1: Методология концептуального анализа и синтеза; Методология концептуального проектирования систем организационного управления; Организационное консультирование. – М.: Концепт, 2008. – 262 с.

11. Son S., Seok J. Improving monte carlo dropout uncertainty estimation with stable output layers. *Neurocomputing*, 2025, vol. 661, p. 131927, DOI: 10.1016/j.neucom.2025.131927.
12. Kim T., Park G., Kwak K. et al. Smooth model predictive path integral control without smoothing. *IEEE Robotics and Automation Letters*, 2022, vol. 7, pp. 1-8, DOI: 10.1109/LRA.2022.3192800.
13. Панфилов И.А. Разработка математического и алгоритмического обеспечения интеллектуального проектирования аппаратно-программных комплексов обработки информации в распределенных высокопроизводительных системах / И.А. Панфилов, Н.А. Смирнов, В.А. Терсков // Сибирский аэрокосмический журнал, 2011. – № 5 (38). – URL: <https://cyberleninka.ru/article/n/razrabotka-matematicheskogo-i-algoritmicheskogo-obespecheniya-intellektualnogo-proektirovaniya-apparatno-programmnyh-kompleksov> (дата обращения: 07.01.2026).

Бочков Александр Владимирович. Доктор технических наук, ученый секретарь, АО «НИИАС» (Россия, Москва). Автор более 100 научных публикаций, в том числе 6-х монографий. Область научных интересов: системы жизнеобеспечения человека в экстремальных условиях окружающей среды, исследования нестационарных процессов, вопросы синтеза и анализа риска, безопасности критической инфраструктуры, системный анализ и исследование операций, экспертные методы. Scopus Author ID: 57188739597, ORCID: 0000-0003-0211-8320, ResearcherID: A-4101-2014, SPIN: 8277-6641, AuthorID: 660808 a.bochkov@gmail.com. 107078, Россия, Москва, Орликов переулок, д.5, стр.1.

UDC 005.334:004.94

DOI:10.25729/ESI.2026.43.3.001

Risk as an immanent property of ergatic systems in the era of digital twins: from the "digital Leviathan" to the paradigm of symbiotic management

Alexander V. Bochkov

JSC "NIAS", Moscow, Russia, a.bochkov@gmail.com

Abstract. The current transition to so-called "data-driven management," based on the creation of full-scale digital twins of complex socio-economic systems, faces a fundamental challenge. Managed systems are ergatic, meaning they include ultimately non-formalizable human, technological, and environmental components. This article demonstrates that optimization problems solved for such digital models are inherently ill-posed with respect to the original, as the human element introduces reflexive non-formalizability and implicit knowledge that cannot be captured by any formal structure. Risk in this paradigm ceases to be an external threat, becoming an immanent property of the management system itself, generated by the inevitable and growing discrepancy between the trajectories of the virtual model and the real ergatic process. Strategies optimal for the "Digital Leviathan" (total simulation) lead to the destabilization of the original, erosion of the human component, and the system exceeding the boundaries of controllability. The authors substantiate the need for a paradigm shift: from searching for a point optimum for an incomplete model to synthesizing robust and adaptive strategies that explicitly minimize the risk functional. A key element of the new "symbiotic management" paradigm is the formal inclusion of the human operator as a corrective loop in the iterative dialogue between the model and reality. The proposed architecture of symbiotic management, grounded in Bayesian updating and minimax principles, enables not merely risk minimization but active management of epistemic uncertainty, transforming human intuition into a structured source of model correction. This approach ensures the resilience of complex socio-technical systems under conditions of fundamental knowledge incompleteness, offering a path from the illusion of absolute control to the wisdom of collaborative navigation. The goal of management becomes not an absolute optimum in virtual space, but dynamic adaptive efficiency in the real world, achieved through constant management of the immanent risk of trajectory divergence. The article introduces and contrasts two concepts: the "Digital Leviathan" (the temptation to manage the simulation) and the "formalism of risk" (the discipline of constantly comparing the map and the territory).

Keywords: digital twin, ergatic system, immanent risk, optimization problem, model incorrectness, symbiotic management, adaptive strategy, human factor, epistemic uncertainty

References

1. Bochkov A.V., Lesnykh V.V. Na puti k formalizmu riska. Razmyshleniya o riske i ego prirode [On the path to formalism of risk. Reflections on risk and its nature]. Vologda, Infra-Inzheneriya Publ., 2024, 296 p. ISBN 978-5-9729-2146-1, EDN: MXDJTV.
2. Gastev Yu.A. Gomomorfizmy i modeli. Logiko-algebraicheskiye aspekty modelirovaniya [Homomorphisms and models. Logical-algebraic aspects of modeling]. Moscow, Nauka [Science] Publ., 1975, 152 p.
3. Pugovkina A.S. Implitsitnyye znaniya kak aspekt intuitsii i prakticheskogo intellekta [Implicit knowledge as an aspect of intuition and practical intelligence]. Evraziyskiy Soyuz Uchenykh [Eurasian Union of Scientists], 2015, no. 3-7 (12). Available at: <https://cyberleninka.ru/article/n/implitsitnye-znaniya-kak-aspekt-intuitsii-i-prakticheskogo-intellekta> (accessed: 01.07.2026).
4. Zhuravlev V.F. Nekorrektnyye zadachi mekhaniki [Ill-posed problems of mechanics]. Vestnik MGTU im. N.E. Baumana. Seriya "Priborostroyeniye" [Bulletin of Bauman MSTU. Series "Instrument Engineering"], 2017, no. 2 (113). Available at: <https://cyberleninka.ru/article/n/nekorrektnye-zadachi-mekhaniki> (accessed: 01.07.2026), DOI:10.18698/0236-3933-2017-2-77-85.
5. Bochkov A.V. Na puti k formalizmu riska. Bezopasnost' i ustoychivost' funktsionirovaniya umnykh ekspansivnykh sistem [On the path to formalism of risk. Safety and stability of smart expansive systems]. Vologda; Moscow, Infra-Inzheneriya Publ., 2025, 348 p. ISBN 978-5-9729-2373-1, EDN: VQVVDDB.
6. Belikov A. Age-related diseases as vicious cycles. Preprints, 2017, DOI: 10.20944/preprints201704.0102.v1.
7. Samarskii A.A., Galaktionov V.A., Kurdyumov S.P., Mikhailov A.P. Rezhimy s obostreniyem v zadachakh dlya kvazilineynykh parabolicheskikh uravneniy [Blow-up in problems for quasilinear parabolic equations]. Moscow, Nauka [Science] Publ., 1987, 470 p. (English translation: Samarskii A.A., Galaktionov V.A., Kurdyumov S.P., Mikhailov A.P. Blow-up in problems for quasilinear parabolic equations. Berlin; N.Y., Walter de Gruyter, 1995, 533 p.)
8. Prelov V.V. f-divergence and gluing of probability distributions. Problems of Information Transmission, 2021, vol. 57, no. 1, pp. 64–80, DOI: 10.31857/S0555292321010034.
9. Sherezhev M.A., Guo U., Serebrenny V.V. Modifikatsiya algoritma glubokogo obucheniya dlya raspredeleniya funktsiy i zadach mezhdru robototekhnicheskimi kompleksami i chelovekom v usloviyakh neopredelennosti i peremennosti okruzhayushchey sredy [Modification of a deep learning algorithm for distributing functions and tasks between a robotic complex and a human in conditions of uncertainty and variability of the environment]. Izvestiya KBNTs RAN [Proceedings of the Kabardino-Balkarian Scientific Center of RAS], 2024, no. 6. Available at: <https://cyberleninka.ru/article/n/modifikatsiya-algoritma-glubokogo-obucheniya-dlya-raspredeleniya-funktsiy-i-zadach-mezhdru-robototekhnicheskimi-kompleksami-i> (accessed: 01.07.2026).
10. Kuchkarov Z.A. Metody kontseptual'nogo analiza i sinteza v teoreticheskom issledovanii i proyektirovanii sotsial'no-ekonomicheskikh sistem: uchebnoye posobiye [Methods of conceptual analysis and synthesis in theoretical research and design of socio-economic systems: textbook]. Vol. 1. Moscow, Kontsept Publ., 2008, 262 p.
11. Son S., Seok J. Improving monte carlo dropout uncertainty estimation with stable output layers. Neurocomputing, 2025, vol. 661, p. 131927, DOI: 10.1016/j.neucom.2025.131927.
12. Kim T., Park G., Kwak K. et al. Smooth model predictive path integral control without smoothing. IEEE Robotics and Automation Letters, 2022, vol. 7, pp. 1-8, DOI: 10.1109/LRA.2022.3192800.
13. Panfilov I.A., Smirnov N.A., Terskov V.A. Razrabotka matematicheskogo i algoritmicheskogo obespecheniya intellektual'nogo proyektirovaniya apparatno-programmnykh kompleksov obrabotki informatsii v raspredelennykh vysokoproizvoditel'nykh sistemakh [Development of mathematical and algorithmic support for intelligent design of hardware-software complexes for information processing in distributed high-performance systems]. Sibirskiy aerokosmicheskiy zhurnal [Siberian Aerospace Journal], 2011, no. 5 (38). Available at: <https://cyberleninka.ru/article/n/razrabotka-matematicheskogo-i-algoritmicheskogo-obespecheniya-intellektualnogo-proektirovaniya-apparatno-programmnykh-kompleksov> (accessed: 01.07.2026).

Bochkov Alexander Vladimirovich. Doctor of Technical Sciences, Scientific Secretary, JSC "NIIAS" (Moscow, Russia). Author of over 100 scientific publications, including 5 monographs. Research interests: human life support systems in extreme environmental conditions, studies of non-stationary processes, risk synthesis and analysis, critical infrastructure safety, systems analysis and operations research, expert methods. Scopus AuthorID: 57188739597, ORCID: 0000-0003-0211-8320, ResearcherID: A-4101-2014, SPIN: 8277-6641, AuthorID: 660808 a.bochkov@gmail.com. 107078, Russia, Moscow, Orlikov Pereulok, 5/1.

Статья поступила в редакцию 28.01.2026; одобрена после рецензирования 18.02.2026; принята к публикации 11.08.2026.

The article was submitted 01/28/2026; approved after reviewing 02/18/2026; accepted for publication 08/11/2026.

Методы, технологии и приложения искусственного интеллекта

УДК 004.85:004.8:368.2

DOI:10.25729/ESI.2026.43.3.002

Нейросетевой стекинг для прогнозирования убытков в страховании

Морозов Михаил Ильич

Югорский Государственный Университет,

Россия, Ханты-Мансийск, mikhailmorozov99@bk.ru

Аннотация. В работе раскрываются основные ограничения древовидных систем принятия решений в страховании – чувствительность к шуму и выбросам, склонность к переобучению, ограниченная способность к обобщению. Предлагаются способы их преодоления за счет методов машинного обучения, нейронных сетей и методов стекинг-ансамблирования. Ставится гипотеза о улучшении качества прогнозирования убытков, где прогнозы от нескольких базовых моделей (XGBoost, CatBoost, LogisticRegression, RandomForest, TabTransformer), в совокупности с исходными признаками набора данных, объединяются в более сильный ансамбль моделей, где строится метамодель на основе полносвязной нейронной сети с дополнительным входом исходным признакам. На данных страхования автомобилей в России за 2023-2025 годы, проводится исследование, где описывается порядок эксперимента с обучением разнородных моделей машинного обучения и нейронных сетей. Описывается алгоритм предобработки набора данных с акцентом на специфичное кодирование категорий (европейская классификация автомобилей по сегменту люксовости), делается акцент на несбалансированности целевого класса – менее 3% убытков во всем объеме данных. На основе обученных моделей строятся модели ансамблей, используя пять методов стекинг-ансамблирования: Simple Average, LogisticRegression-Constraint, Adaptive Regression by Mixing, ARM-Tweedie, Multilayer Perceptron (MLP). Проводится качественное сравнение базовых моделей с ансамблями моделей на основе классификационных (ROC-AUC, PR-AUC) и вероятностных метрик (Logloss, BrierScore). По результатам исследования лучшей базовой моделью является TabTransformer (BrierScore = 0.037, LogLoss = 0.185, ROC-AUC = 0.635, PR-AUC = 0.054). Ансамбль на основе полносвязной нейронной сети (MLP) превзошел ее по всем ключевым метрикам (BrierScore = 0.026, LogLoss = 0.131, ROC-AUC = 0.649, PR-AUC = 0.055). Полученный результат исследования позволяет подтвердить поставленную гипотезу: ансамбль на основе полносвязной нейронной сети позволил получить прирост по основным метрикам оценки. Полученный результат доказывает методологическую ценность подхода, даже в условиях ограниченного сигнала данных нейросетевой ансамбль в совокупности с исходными признаками обеспечивает более эффективное объединение прогнозов по сравнению с методами предиктивной аналитики и методами стекинг-ансамблирования.

Ключевые слова: деревья решений, нейронные сети, машинное обучение, автомобильное страхование, убыточность, стекинг ансамблирование

Цитирование: Морозов М.И. Нейросетевой стекинг для прогнозирования убытков в страховании / М.И. Морозов // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 22-30. – DOI:10.25729/ESI.2026.43.3.002.

Введение. В страховании одной из важных задач является задача предсказания вероятности наступления страхового случая для оценки риска и формирования страховой премии [1, 2]. Делается это с целью обеспечить экономическую стабильность страховой организации и формирования бюджета на выплату страховых случаев [3,4].

В оценке рисков в страховании используют как актуарные оценки данных клиентов, так и системы принятия решений на основе древовидной структуры. Подобные системы чаще всего используют вербальные правила с условиями «если...то...», что позволяет настраивать необходимую сегментацию по портфелю страхования. Ключевой особенностью использования деревьев решений в страховании является интерпретируемость и прозрачность, страховая организация может объяснить, почему было принято то или иное решение. Однако такие системы строятся и поддерживаются пользователями вручную, что неизбежно приводит к многим проблемам деревьев решений: чувствительность к шуму и выбросам, склонность к переобучению, ограниченная способность к обобщению.

Не менее популярным методом оценки рисков в страховых организациях являются построение моделей машинного [5] и нейросетевого обучения [6,7], что позволяет выявлять нелинейные зависимости в данных и чаще всего именно комбинация деревьев решений и моделей машинного обучения дает более точную оценку страхового риска по портфелю страхования. В рамках данной работы автор формулирует гипотезу, что комбинация нескольких индивидуальных моделей машинного и нейросетевого обучения позволит построить более сильный ансамбль на основе полносвязной нейронной сети для точной оценки страхового риска в задаче предсказания наступления страховых случаев. Для подтверждения гипотезы сравниваются несколько методов классического стекинга с простой полносвязной нейронной сетью, целью которой является объединение прогнозов индивидуальных моделей с дополнительным входом исходных данных для более точного прогнозирования вероятности наступления страхового случая.

Данное исследование является расширением предыдущей работы [8]. Дополнительно добавлен новый метод трансформерной нейросети TabTransformer в ансамбль, а также изменена предобработка данных для более качественного обучения моделей и сравнения полученных результатов

2. Модели и методы исследования. В рамках текущей работы обучаются пять моделей, которые лягут в основу ансамбля: XGBoost [9], CatBoost [10], LogisticRegression, RandomForest [11], TabTransformer [12]. В качестве методов ансамблирования используются: Simple Average (SA), LogisticRegression-Constraint (LR-C), Adaptive Regression by Mixing (ARM), Adaptive Regression by Mixing Tweedie [13], Multilayer Perceptron (MLP).

2.1. Описание набора данных. Набор данных содержит информацию о страховании автомобилей в РФ на период с 2023 по 2025 год. Он содержит данные полисов о страховании клиентов. Набор данных является несбалансированным, менее 3% убытков из всего датасета, содержит более 50 уникальных признаков. Набор данных основан на промышленных логах высоконагруженной системы, предоставленных партнерской организацией. В соответствии с соглашением о конфиденциальности, полный набор данных не подлежит публичному раскрытию.

2.2. Предварительная обработка данных. На основе данных проводится извлечение исходных признаков, где на основе гипотез строятся новые признаки набора данных. С помощью корреляции проверяются новые построенные признаки на основе вклада в целевую переменную. Специфические категориальные признаки кодируются особым способом для сохранения связи между признаками датасета:

1. Особым случаем кодирования является признак марки и модели транспортного средства, т.к. размер категорий марки и моделей может насчитывать более нескольких тысяч, что приводит к проблеме кодирования категорий. В зависимости от стоимости автомобиля и его комплектации может меняться и стоимость страхования, и, чтобы сохранить такую взаимосвязь, используется европейская классификация автомобилей, где на основе стоимости, марки и модели автомобиля определяется сегмент люксовости автомобиля [14].
2. На основе информации о водителях строятся как индивидуальные, так и географические признаки, позволяющие учесть различные взаимосвязи в данных.

При извлечении исходных признаков строится матрица корреляции, где рассматриваются взаимосвязи между целевым и исходными признаками. Удаляются признаки с высоким уровнем корреляции, выше значения 0.9, так как такие признаки не привносят в данные новый сигнал для прогнозирования целевой переменной.

Далее происходит кодирование набора данных. Даты кодируются на основе синусного и косинусного расстояния по месяцам и дням, для учета цикличности календарного года. После

проведения основной предобработки данных происходит разделение данные на 3 целевых выборки:

- Train – данные между 2024 и 2025 годом
- Valid – данные за последующие 3 месяца 2025 г.
- Test – данные старше 2025 г.

Последующим этапом заполняются пропуски, для бинарных признаков происходит заполнение нулевыми значениями, для числовых – медианное заполнение и масштабирование на основе MinMax.

В категориальных данных на уровне предобработки заполняются пропуски значениями 'Unknown' и кодируются они в последнюю очередь, т.к. различные методы построения моделей обрабатывают категориальные признаки по-разному.

Для всех методов, представленных в разделе 2.3, кроме метода TabTransformer, используется кодирование категорий с помощью метода OneHotEncoder. Для метода TabTransformer категории подаются в сыром виде, где сам метод строит эмбединги для кодирования категорий

На предварительном этапе обработки вычисляются отдельные веса для целевого признака на основе частоты встречаемого класса во всей выборке данных, что позволяет учесть редкость предсказываемого класса за счет использования полученных весов в функции потерь. Веса для классов в функции потерь позволяют вводить дополнительный штраф при обучении моделей, что позволяет обратить внимание обучаемой модели на редкий класс. Полный алгоритм предобработки данных можно видеть на рис. 1.

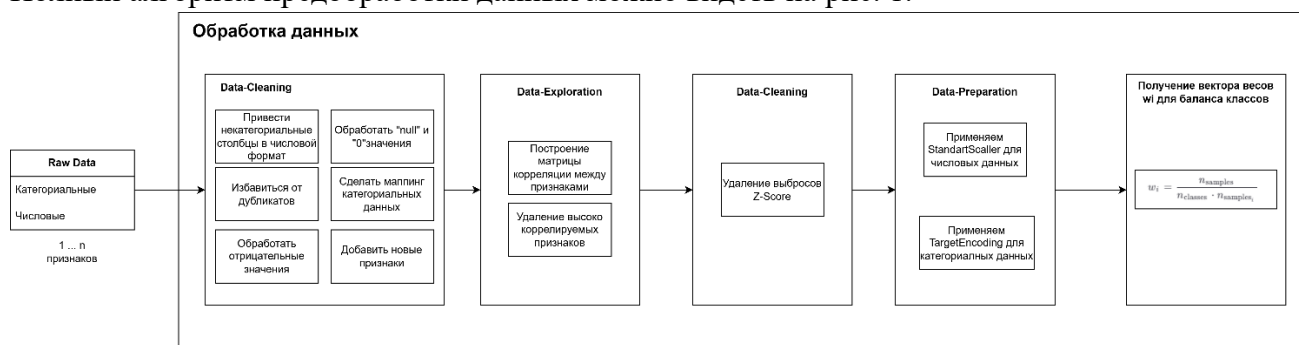


Рис. 1. Алгоритм предобработки данных

2.3. Методы предиктивной аналитики. В рамках текущей работы используются пять методов предиктивной аналитики, которые определяют основу ансамбля моделей. Выбор этих методов обуславливается их популярностью в задачах бинарной классификации и разнородностью самих методов в способах обучения моделей, а также популярностью использования этих методов в задачах предсказания убытков в страховой сфере [15,16]:

- **XGBoost** – устойчивый алгоритм градиентного бустинга, который позволяет эффективно обрабатывать пропуски и обеспечивать высокую точность прогнозирования в шумных данных, что является критичной и важной особенностью для страховых данных с естественной вариативностью.
- **CatBoost** – более продвинутая версия градиентного бустинга для более удобной работы с категориальными признаками. Метод использует способ кодирования категорий TargetEncoding и симметричные деревья, что снижает переобучение и повышает стабильность прогнозов на данных с большим количеством категорий.
- **LogisticRegression** – стандартная линейная модель, оценивающая вероятность наступления через логистическую функцию. С регуляризацией L1/L2 позволяет обеспечить устойчивость к переобучению, выделить ключевые факторы риска.

- **RandomForest** – ансамбль из несвязанных деревьев решений, основанный на методах бэггинга и случайного отбора признаков. Обладает устойчивостью к выбросам и шуму, позволяет обеспечить оценку важности признаков и служит базовым методом деревьев решений для сравнения с более сложными методами.
- **TabTransformer** – трансформерная архитектура обучения моделей, адаптированная для работы с табличными данными. Особенностью данного метода является кодирование категориальных признаков в виде эмбедингов, наличие механизма самовнимания, что позволяет выявлять сложные зависимости между исходными признаками в данных. Указанные особенности позволяют улавливать более сложные взаимодействия в данных, что полезно при анализе редких событий.

2.4 Объединение прогнозов моделей. Для объединения прогнозов моделей, которые будут строиться на основе методов предиктивной аналитики, применяется метод стекинга. Стекинг позволяет объединить прогнозы нескольких базовых моделей с помощью финальной метамодели, которая, в свою очередь, позволит усилить общую предсказательную способность ансамбля за счет компенсации индивидуальных особенностей каждой из моделей. Указанные ниже методы широко встречаются в исследованиях [13] для построения метамodelей стекинга:

- **Simple Average (SA)** – метод простого среднего, полученные вероятности от базовых моделей складываются и делятся на количество моделей. Метод не учитывает особенности каждой из моделей, предполагается, что мы можем доверять каждой из моделей одинаково, когда на практике степень доверия может зависеть от различных факторов в данных.
- **LogisticRegression-Constraint (LR-C)** – метод позволяет строить взвешенную комбинацию моделей, что позволяет исключить доминирование отдельных моделей в ансамбле.
- **Adaptive Regression by Mixing (ARM)** – адаптивный смешиватель, где вес каждой базовой модели динамически корректируется в зависимости от ее точности.
- **Adaptive Regression by Mixing Tweedie** – модификация метода ARM с использованием распределения Tweedie в функции потерь, что позволяет учитывать специфику страховых данных с распределением убытков на хвосте, что позволяет повысить качество оценки как вероятности.
- **Multilayer Perceptron (MLP)** – простая полносвязная нейронная сеть, используемая как мета-классификатор. На вход принимает базовые прогнозы моделей и дополнительный поток исходных признаков, что, на основе гипотезы, должно позволять выявлять нелинейные зависимости между ошибками разных базовых моделей и исходными данными. В текущем исследовании используется нейронная сеть, состоящая из 3-х слоев (128, 64, 32) с регуляризацией (dropout, L2) для предотвращения переобучения ансамбля.

2.5 Методы оценки моделей. В качестве методов оценки используется ряд ключевых метрик, которые позволяют оценить эффективность моделей как на основе калибровки вероятностей, так и способности выявлять редкие события с учетом несбалансированности данных:

- **BrierScore** – стандартная среднеквадратическая ошибка между прогнозируемой вероятностью и фактическим результатом в данных. Оценивает калибровку модели, где низкое значение говорит о соответствии прогнозируемых вероятностей реальным частотам наступления страховых случаев, что является критичным для расчета тарифа страхования на основе страховой премии. Теоретическим минимумом для данной задачи при 3% целевого класса является значение = 0.03.

- **LogLoss** – штрафует модель за неуверенные прогнозы, чем дальше прогноз от истинного класса, тем выше штраф. Метрика очень чувствительна к качеству оценки редких событий. Теоретическим минимумом для данной задачи при 3% целевого класса является значение, равное 0.13.
- **ROC-AUC** – оценивает способность модели ранжировать клиентов по риску независимо от порога бинаризации (threshold). Используется, как вспомогательная метрика для сравнения дискриминативной силы моделей, но является менее информативной при сильном дисбалансе классов.
- **PR-AUC** – предпочтительная метрика для задач предсказания случаев на несбалансированных данных, фокусируется на качестве прогнозов для редкого класса, отражает баланс между выявленными рисками и ложными срабатываниями. Высокое значение указывает на способность модели концентрировать внимание на рискованных клиентах.

3. Экспериментальная установка. В качестве инструментов для проведения эксперимента использовалось программное обеспечение JupyterHub – для обучения моделей, MLFlow – для логирования результатов эксперимента. Порядок эксперимента состоит из нескольких ключевых шагов (Рис. 2).

После разделения данных на тренировочную, тестовую и валидационную выборки выполняется обучение моделей на полной тренировочной выборке. На основе методов предиктивной аналитики с помощью кросс валидации формируется датасет, который участвует в построении и обучении ансамблей. Датасет состоит из пяти колонок, где каждая колонка – это название метода предиктивной аналитики, каждая строка – предсказанная вероятность по экземпляру данных. На основе полученного датасета строится матрица корреляции обученных моделей, для выявления сильно коррелированных моделей выполняется построение ансамблей с помощью методов объединения прогнозов моделей.

В качестве финального этапа с помощью методов оценки моделей выполняется сравнение каждого из методов предиктивной аналитики с методами объединения прогнозов моделей. Для честной оценки и сравнения моделей обученных на основе методов предиктивной аналитики и полученных ансамблей финальная калибровка вероятностей на основе изотонической регрессии не проводится.

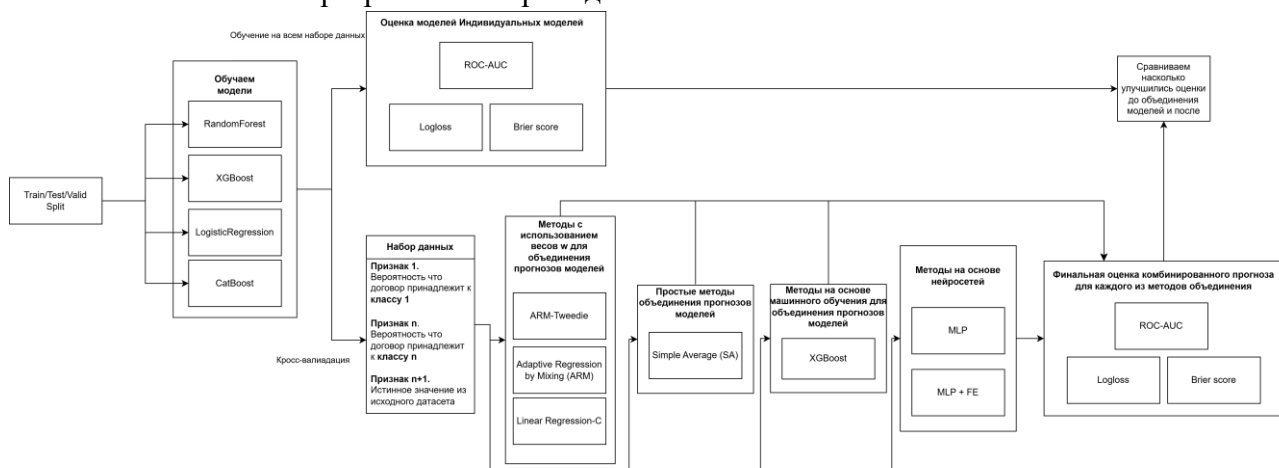


Рис. 2. Порядок проведения эксперимента

4. Интерпретация результатов. На основе предварительной обработки данных п. 2.2 и порядку эксперимента, описанному в п. 3, проводится обучение моделей. Ниже представлены результаты обучения на основе методов предиктивной аналитики (Таблица 1), матрица

корреляции обученных моделей (Рис. 3), результаты ансамблирования моделей (Таблица 2). Полученные метрики представлены на основе оценки моделей с помощью тестовой выборки.

Таблица 1. Результаты обучения моделей на основе методов предиктивной аналитики

Модель/Метрика	XGBoost	CatBoost	LR	RF	TabTransformer
BrierScore	0.170	0.134	0.252	0.215	0.037
LogLoss	0.521	0.435	0.701	0.624	0.185
ROC-AUC	0.617	0.621	0.643	0.660	0.635
PR-AUC	0.053	0.055	0.052	0.053	0.054

Таблица 2. Результаты обучения ансамблей

Модель/Метрика	SA	LR-C	ARM	ARM-T	MLP
BrierScore	0.114	0.208	0.232	0.232	0.026
LogLoss	0.401	0.608	0.658	0.657	0.131
ROC-AUC	0.648	0.644	0.647	0.649	0.649
PR-AUC	0.049	0.047	0.0506	0.051	0.055

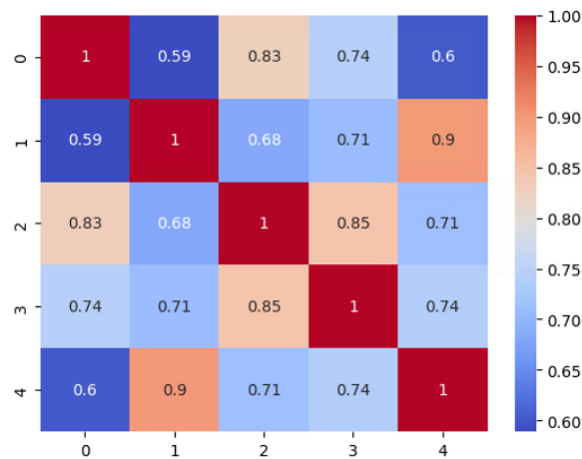


Рис. 3. Матрица корреляции моделей

Среди моделей (Табл. 1), обученных на основе методов предиктивной аналитики (п. 2.3), наилучшими показателями по метрикам BrierScore, LogLoss обладает метод TabTransformer, это означает, что данная модель точнее других предсказывает вероятности. По метрикам оценки способности модели различать классы, по метрике ROC-AUC лучшей моделью является RandomForest, по метрике PR-AUC – модель CatBoost. По комбинации вероятностных и дискриминационных метрик среди моделей предиктивной аналитики лидером является TabTransformer.

Среди моделей (Табл. 2), обученных на основе методов стекинг-ансамблирования (п. 2.4), по всем ключевым метрикам наблюдается превосходство полносвязной нейронной сети MLP, что позволяет подтвердить гипотезу о комбинации нескольких индивидуальных моделей как на основе машинного обучения, так и нейросетевого обучения с целью построить более сильный ансамбль для точной оценки страхового риска в задаче предсказания наступления страховых случаев. Однако низкий показатель PR-AUC во всех базовых моделях и ансамблях говорит об одном: в данных недостаточно сигнала для предсказания целевого события. Так как исходный набор данных содержит только информацию о договоре страхования, без обогащения данных из внешних источников, то такой информации недостаточно для предсказания целевого события на качественно значимом уровне, что и подтверждает метрика PR-AUC.

Полученные результаты не претендуют на практическую применимость в страховании без обогащения данных, однако доказывают методологическую ценность подхода: даже в

условиях ограниченного сигнала данных нейросетевой ансамбль в совокупность с исходными признаками данных обеспечивает более эффективное объединение прогнозов по сравнению с методами предиктивной аналитики и методами стекинг-ансамблирования.

Заключение. На основе полученных результатов исследования можно сделать следующие выводы:

- По результатам сравнения методов предиктивной аналитики лучшими являются: TabTransformer по метрикам BrierScore=0.037 и LogLoss=0.185, RandomForest по метрике ROC-AUC=0.66, CatBoost по метрике PR-AUC=0.055. По совокупности всех метрик TabTransformer лидирует, благодаря вероятностным метрикам и показателям PR-AUC, ROC-AUC, близким к моделям RandomForest и CatBoost.
- Среди методов стекинг-ансамблирования, лучшим является Multilayer Perceptron (MLP) с вероятностными метриками BrierScore=0.026, LogLoss=0.131, которые близки к теоретическому минимуму, дискриминационными метриками ROC-AUC=0.649, PR-AUC=0.055. Этот метод превосходит другие методы предиктивной аналитики и ансамблирования по общей совокупности и комбинации всех метрик.
- Ансамбль на основе полносвязной нейронной сети работает, поставленная гипотеза подтверждается. Исходные признаки датасета в совокупности с предсказаниями моделей предиктивной аналитики позволяют обучить более сильный ансамбль и добиться лучших результатов по метрикам.

Дальнейшим этапом исследования будет являться калибровка полученных моделей и ансамблей с последующей оценкой эффекта моделей по бизнес-метрикам на различных сегментах данных для проверки устойчивости поставленной гипотезы. Дополнительно планируется проверить гипотезу на данных с более сильным целевым сигналом, для осуществления практической применимости обученного ансамбля.

Благодарности. Автор выражает благодарность научному руководителю, заслуженному деятелю науки РФ, д-ру техн. наук, проф. Мельникову Андрею Витальевичу за профессиональное руководство, помощь и активное участие в развитии научного исследования.

Список источников

1. Ахвледиани Ю.Т. Актуальные направления развития страхового рынка в современных условиях / Ю.Т. Ахвледиани. – М.: МГУ, 2023. – С. 26–30.
2. Ritho B.M., Simiyu E., Omagwa J. The impact of loss ratio on the financial stability of insurance firms in Kenya. *Journal of Finance and Accounting*, 2023, vol. 7, no. 4, pp. 22–41.
3. Бронская Т.А. Искусственный интеллект в страховой сфере как инструмент повышения конкурентоспособности / Т.А. Бронская. – Минск: БГУ, 2023. – С. 389–391.
4. Кириченко А.О. Прогнозирование страховых рисков с использованием искусственного интеллекта / А.О. Кириченко, А.Л. Золкин, И.А. Поскряков // Прикладные экономические исследования, 2024. – № 3. – С. 196–203.
5. So B. Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM. arXiv:2307.07771, 2024, DOI: 10.48550/arXiv.2307.07771.
6. Singh U. et al. Comparative analysis of transformers for modeling tabular data: a case study using industry scale dataset. arXiv:2311.14335, 2023, DOI: 10.48550/arXiv.2311.14335.
7. Морозов М.И. Предсказание наступления страхового случая с помощью трансформерных нейросетей / М.И. Морозов // Системная инженерия и информационные технологии, 2025. – Т. 7. – № 2(21). – С. 96–102.
8. Морозов М.И. Прогнозирование страховых случаев на основе стекинг-ансамблирования / М.И. Морозов // Системная инженерия и информационные технологии, 2026. – Т. 8. – № 1(25). – С. 93–100.
9. Chen T., Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, DOI: 10.1145/2939672.2939785.

10. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features. arXiv:1706.09516, 2019, DOI: 10.48550/arXiv.1706.09516.
11. Breiman L. Random forests. Machine Learning, 2001, vol. 45, no. 1, pp. 5–32, DOI: 10.1023/A:1010933404324.
12. Huang X. et al. TabTransformer: tabular data modeling using contextual embeddings. arXiv:2012.06678, 2020, DOI: 10.48550/arXiv.2012.06678.
13. Chenglong Ye, Zhang L., Han M. et al. Combining predictions of auto insurance claims. Econometrics, 2022, vol. 10, iss. 2, 19, DOI: 10.3390/econometrics10020019.
14. Какие классы автомобилей существуют, и как их определить. – Major Auto. – URL: <https://www.major-auto.ru/blog/132/> (дата обращения: 29.03.2026).
15. Baran S., Rola P. Prediction of motor insurance claims occurrence as an imbalanced machine learning problem. arXiv, 2022, DOI:10.48550/arXiv.2204.06109.
16. Mohamed Hanafy, Ruixing Ming. Machine learning approaches for auto insurance big data. Risks, 2021, vol. 9(2), no. 42, DOI: 10.3390/risks9020042.

Морозов Михаил Ильич. Аспирант спец. «Системный анализ, управление и обработка информации», Югорский Государственный Университет, SPIN: 7231-0420, ORCID: 0009-0004-1217-9439, mikhailmorozov99@bk.ru, Ханты-Мансийский автономный округ - Югра, г. Ханты-Мансийск, ул. Чехова, 1

UDC 004.85:004.8:368.2

DOI:10.25729/ESI.2026.43.3.002

Neural network stacking for loss forecasting in Insurance

Mikhail I. Morozov

Yugra state university, Russia, Khanty-Mansiysk, mikhailmorozov99@bk.ru

Abstract. This study examines the primary limitations of tree-based decision models in the insurance sector, namely their sensitivity to noise and outliers, susceptibility to overfitting, and limited generalization capabilities. To overcome these challenges, we propose leveraging advanced machine learning techniques, neural networks, and stacking ensemble methods. We hypothesize that the quality of loss prediction can be slightly improved by combining the predictions of multiple base models (XGBoost, CatBoost, Logistic Regression, Random Forest, and TabTransformer), alongside the original dataset features, into a more robust ensemble. Specifically, the meta-model is constructed as a fully connected neural network with an additional input layer for the original features. The empirical study utilizes Russian auto insurance data from 2023 to 2025, detailing the experimental framework for training heterogeneous machine learning and neural network models. The data preprocessing pipeline is described, with a particular focus on a specific categorical encoding scheme based on the European car classification by luxury segment. Special attention is given to the severe imbalance of the target class, as losses account for less than 3% of the total dataset. Based on the trained base models, ensemble models are constructed using five stacking techniques: Simple Average, Constrained Logistic Regression, Adaptive Regression by Mixing (ARM), ARM-Tweedie, and Multilayer Perceptron (MLP). A comprehensive comparative analysis of the base models and the ensembles is conducted using both classification metrics (ROC-AUC, PR-AUC) and probabilistic metrics (Log Loss, Brier Score). The results indicate that the best-performing base model is TabTransformer (Brier Score = 0.037, Log Loss = 0.185, ROC-AUC = 0.635, PR-AUC = 0.054). The MLP-based ensemble outperformed it across all key metrics (Brier Score = 0.026, Log Loss = 0.131, ROC-AUC = 0.649, PR-AUC = 0.055). These findings confirm the initial hypothesis: the fully connected neural network ensemble yields a slightly improvement in the primary evaluation metrics. This demonstrates the methodological value of the proposed approach: even under conditions of a weak data signal, the neural network ensemble, augmented with original features, ensures a more effective aggregation of predictions compared to traditional predictive analytics and alternative stacking ensemble methods.

Keywords: decision trees, neural networks, machine learning, vehicle insurance, loss ratio, stacking ensemble

Acknowledgements: The author expresses gratitude to the scientific supervisor, Honored Scientist of the Russian Federation, Doctor of Technical Sciences, Professor Andrey Vitalievich Melnikov for his professional guidance, assistance and active participation in the development of scientific research.

References

1. Akhvlediani Yu.T. Aktual'nyye napravleniya razvitiya strakhovogo rynka v sovremennykh usloviyakh [Current directions of insurance market development in modern conditions]. Moscow, Moscow State University Publ., 2023, pp. 26–30.
2. Ritho B.M., Simiyu E., Omagwa J. The impact of loss ratio on the financial stability of insurance firms in Kenya. *Journal of Finance and Accounting*, 2023, vol. 7, no. 4, pp. 22–41.
3. Bronskaya T.A. Iskusstvennyy intellekt v strakhovoy sfere kak instrument povysheniya konkurentosposobnosti [Artificial intelligence in the insurance sector as a tool for increasing competitiveness]. Minsk, Belarusian State University Publ., 2023, pp. 389–391.
4. Kirichenko A.O., Zolkin A.L., Poskryakov I.A. Prognozirovaniye strakhovykh riskov s ispol'zovaniyem iskusstvennogo intellekta [Forecasting insurance risks using artificial intelligence]. *Prikladnyye ekonomicheskiye issledovaniya* [Applied Economic Research], 2024, no. 3, pp. 196–203.
5. So B. Enhanced gradient boosting for zero-inflated insurance claims and comparative analysis of CatBoost, XGBoost, and LightGBM. arXiv:2307.07771, 2024, DOI: 10.48550/arXiv.2307.07771.
6. Singh U. et al. Comparative analysis of transformers for modeling tabular data: a casestudy using industry scale dataset. arXiv:2311.14335, 2023, DOI: 10.48550/arXiv.2311.14335.
7. Morozov M.I. Predskazaniye nastupleniya strakhovogo sluchaya s pomoshch'yu transformernykh neyrosetey [Prediction of the occurrence of an insured event using transformer neural networks]. *Sistemnaya inzheneriya i informatsionnyye tekhnologii* [System Engineering and Information Technology], 2025, vol. 7, no. 2(21), pp. 96–102.
8. Morozov M.I. Prognozirovaniye strakhovykh sluchayev na osnove steking-ansamblirovaniya [Forecasting insurance claims based on stacking ensemble]. *Sistemnaya inzheneriya i informatsionnyye tekhnologii* [System Engineering and Information Technology], 2026, vol. 8, no. 1(25), pp. 93–100.
9. Chen T., Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794, DOI: 10.1145/2939672.2939785.
10. Prokhorenkova L. et al. CatBoost: unbiased boosting with categorical features. arXiv:1706.09516, 2019, DOI: 10.48550/arXiv.1706.09516.
11. Breiman L. Random forests. *Machine Learning*, 2001, vol. 45, no. 1, pp. 5–32, DOI: 10.1023/A:1010933404324.
12. Huang X. et al. TabTransformer: tabular data modeling using contextual embeddings. arXiv:2012.06678, 2020, DOI: 10.48550/arXiv.2012.06678.
13. Chenglong Ye, Zhang L., Han M. et al. Combining predictions of auto insurance claims. *Econometrics*, 2022, vol. 10, iss. 2, 19, DOI: 10.3390/econometrics10020019.
14. Kakiye klassy avtomobiley sushchestvuyut, i kak ikh opredelit' [What car classes exist and how to determine them]. *Major Auto*. Available at: <https://www.major-auto.ru/blog/132/> (accessed: 03.29.2026).
15. Baran S., Rola P. Prediction of motor insurance claims occurrence as an imbalanced machine learning problem. arXiv, 2022, DOI:10.48550/arXiv.2204.06109.
16. Mohamed Hanafy, Ruixing Ming. Machine learning approaches for auto insurance big data. *Risks*, 2021, vol. 9(2), no. 42, DOI: 10.3390/risks9020042.

Morozov Mikhail Iliich. Post-graduate student of the specialty "Systems analysis, management and information processing., Yugra State University, SPIN: 7231-0420, ORCID: 0009-0004-1217-9439, mikhailmorozov99@bk.ru, Khanty-Mansi Autonomous Okrug - Yugra, Khanty-Mansiysk, Chekhov Street, 16.

Статья поступила в редакцию 06.04.2026; одобрена после рецензирования 20.06.2026; принята к публикации 26.07.2026.

The article was submitted 04/06/2026; approved after reviewing 06/20/2026; accepted for publication 07/26/2026.

УДК 004.934.2

DOI:10.25729/ESI.2026.43.3.003

Адаптивная гибридная нейросетевая архитектура на основе внимания для идентификации дикторов в корпоративных совещаниях

Потапов Арсений Денисович^{1,2}, Лукьянов Никита Дмитриевич²

¹ПАО Сбербанк России, Россия, Иркутск

²Иркутский национальный исследовательский технический университет,
Россия, Иркутск, lukyanovnd@istu.edu

Аннотация. Актуальность исследования обусловлена ростом потребности в автоматизации ведения протоколов корпоративных встреч, которое является трудоёмким и затяжным процессом, сопровождающимся высоким риском ошибок. Современные системы распознавания речи (ASR – Automatic Speech Recognition) работают нестабильно при многопользовательских диалогах, перекрытиях голосовых фрагментов, фоновом шуме и использовании специализированной лексики. Особенно остро стоит проблема распознавания дикторов (Speaker Identification, SID), когда на одного сотрудника приходится мало эталонных записей, что вполне типично для офисной среды. Дополнительные трудности создаёт и акустическая обстановка: гулкие кабинеты, одновременные высказывания, отсутствие чёткого разделения голосов. Таким образом, цель данной работы – создание гибридной нейросетевой модели для точной транскрипции и надёжной атрибуции речи в групповых дискуссиях. В рамках работы проанализированы слабые места ASR-систем и методов SID, предложена трёхуровневая архитектура: сегментация аудиопотока VAD + SCD (VAD – Voice Activity Detection – определение активности речи), SCD – Speaker Change Detection – детектирование смены диктора), идентификация по голосовым эмбедингам с механизмом внимания к эталонам, транскрибация¹ с учётом личности говорящего, так же приведено математическое обоснование ключевых компонентов. Теоретическая значимость заключается в построении методики SID на основе механизма внимания, позволяющая эффективно решать задачи обучения с малым числом примеров (few-shot learning). В отличие от традиционных подходов, требующих обширных размеченных корпусов, новая архитектура учится сопоставлять речевые фрагменты с небольшим числом эталонных записей. Эффективность метода эмпирически подтверждена на реальных записях корпоративных совещаний (свыше 20 часов, пять участников). Модуль сегментации показал F1-меру VAD на уровне 0,885, а модуль идентификации диктора точность 62,5%, что втрое выше случайного уровня (20%). Транскрибация выполнена через внешнюю систему Whisper, гарантирующую высокую достоверность распознавания. Результаты позволяют построить систему автоматического ведения протоколов совещаний, а архитектура позволяет работать и с большим числом участников, а также легко адаптируется под разные форматы коммуникаций: от переговоров до встреч с участием клиентов. Функциональные возможности гибридной нейросетевой модели могут быть использованы при создании различных инструментов для протоколирования корпоративных встреч.

Ключевые слова: автоматическое распознавание речи, идентификация диктора, гибридная нейросетевая архитектура, механизм внимания, голосовые эмбединги, сегментация аудиопотока, транскрибация, корпоративные совещания

Цитирование: Потапов А.Д. Адаптивная гибридная нейросетевая архитектура на основе внимания для идентификации дикторов в корпоративных совещаниях / А.Д. Потапов, Н.Д. Лукьянов // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 31-41. – DOI:10.25729/ESI.2026.43.3.003.

Введение. Запись совещаний, презентаций и рабочих обсуждений давно стала привычной практикой в большинстве компаний. Архив таких аудиофайлов – это слепок деловой активности: в нём остаются детали проектных решений, логика управленческих шагов и контекст принятых договорённостей. Однако выделить из этих данных структурированное содержание, кто именно, когда и что произнёс, до сих пор является многочасовым кропотливым трудом [1]. Цифровая трансформация бизнеса заставляет крупные компании искать способы автоматической фиксации всего, что происходит на

¹ Транскрибация – процесс перевода устной речи из аудио или видеозаписи в текстовый формат.

встречах – от еженедельных совещаний команды разработки до многочасовых заседаний архитектурного комитета.

Отметим существенное различие решаемой задачи от задач, решаемых системами автоматического распознавания речи, ориентированных на расшифровку монологов, записанных в звукоизолированной кабине [2]. При использовании ASR в реальных корпоративных дискуссиях возникают ошибки: несколько участников говорят по очереди или одновременно, перебивают друг друга, паузы между репликами минимальны, что в конечном счете приводит к ухудшению точности стандартных ASR-решений [3]. Недостаток таких систем – это неумение надёжно поймать момент передачи слова и безошибочно привязать сказанное к конкретному сотруднику [4].

Идентификация диктора (SID) и разделение аудиопотока по голосам (диаризация) долгое время рассматривались как самостоятельные задачи, решаемые отдельно от распознавания текста [5]. Применительно к записям рабочих встреч эта разобщённость создаёт дополнительные трудности. Классические подходы к SID, опирающиеся на гауссовы смеси с универсальной фоновой моделью и технику i-векторов, строятся на выделении устойчивых акустических признаков и их статистическом описании [6]. Подобные методы плохо справляются с короткими фрагментами речи, требуют внушительных объёмов записей на каждого человека и чувствительны к шумам, которыми обычно наполнены офисные переговоры [7].

Прорыв в этой области обеспечило глубокое обучение. Свёрточные и рекуррентные сети, а затем и трансформерные архитектуры [8] позволили заметно улучшить качество как самой транскрипции – достаточно вспомнить успехи модели Whisper [9], – так и извлечения голосовых эмбеддингов, где стандартом стали x-vector и d-vector [10]. Впрочем, даже передовые системы диаризации, например, TS-VAD [11], по-прежнему привязаны к предварительной кластеризации эмбеддингов и раскрывают свой потенциал лишь при наличии масштабных обучающих выборок.

Для корпоративной среды характерна иная ситуация. Здесь на каждого сотрудника приходится считанное количество эталонных записей, и традиционные методы машинного обучения в таких условиях теряют точность [12]. Подобное положение дел вынуждает обращаться к архитектурам, рассчитанным на работу при острой нехватке примеров, то есть в режиме *few-shot learning*. Соответственно, требуется построить единую модель, которая совмещала бы транскрипцию и идентификацию говорящего, причём делала это в условиях ограниченных обучающих данных, столь привычных для офисной среды. Оценку такой системы естественно вести по трём направлениям: насколько точно реплика приписывается автору (*Speaker Accuracy*), как уверенно вычленяются осмысленные речевые отрезки из шумов и пауз (*VAD F1*) и насколько безошибочно фиксируется переход слова от одного участника к другому (*SCD Precision*).

В рамках работы предлагается многомодульная нейросетевая архитектура, берущая на себя одновременно задачи сегментации звукового потока, расшифровки сказанного и установления личности каждого говорящего. Достигнутые на сегодня результаты могут стать основой для разработки средств автоматической фиксации совещаний – инструментов, которые позволят упорядочить документирование корпоративных встреч и снять часть рутинной нагрузки с сотрудников.

Материалы и методы. В основе предлагаемого метода лежит архитектура, объединяющая модули сегментации на основе свёрточной сети, идентификации диктора с механизмом внимания и транскрипции на основе трансформера, предназначенная для последовательной обработки аудиосигнала многопользовательского совещания. Общая схема метода включает три ключевых модуля, работающих последовательно: модуль сегментации и

выделения признаков, модуль идентификации диктора и модуль транскрибации. Обучение сети проводится в два этапа: предварительное обучение на больших публичных наборах данных и тонкая настройка на целевых голосах сотрудников.

Пусть исходными данными является аудиозапись совещания X , представленная в виде временного ряда. Первоначально выполняются предобработка и выделение признаков. Входной сигнал $X(t)$ преобразуется в последовательность логарифмических мел-спектрограмм [13] $S = (s_1, s_2, \dots, s_T)$, где $s_t \in R^F$ – вектор из F мел-частотных коэффициентов в момент времени t , где T – длительность записи [14]. Данное представление сохраняет как спектральные, так и временные характеристики сигнала и является стандартным для современных моделей распознавания речи.

Первым модулем является «Сеть сегментации и выделения универсальных речевых признаков», который реализован на основе стандартной сверточной нейронной сети (CNN) с одномерными свертками во временной области [15]. Его задача – это анализ спектрограммы S и определение двух ключевых аспектов, а именно, наличие речевой активности (Voice Activity Detection – VAD) и признаков смены диктора (Speaker Change Detection – SCD).

Обозначим сверточный модуль как C_{CNN} , который преобразует входную спектрограмму S в последовательность высокоуровневых признаков $H_{speech} = (h_1, h_2, \dots, h_{T'})$, где каждый компонент последовательности $h_t \in R^D$ является вектором признаков размерности D и извлечён из соответствующего временного интервала. Здесь $T' \leq T$ число временных шагов после операций субдискретизации (например, из-за шага свёртки больше 1), которое может быть меньше исходного числа кадров T . Таким образом, последовательность H_{speech} сохраняет временную упорядоченность признаков, но с пониженным разрешением.

$$H_{speech} = C_{CNN}(S), \quad (1)$$

Над этим представлением обучаются два параллельных полносвязных классификатора [16]:

$F_{VAD}: h_t \rightarrow \{0, 1\}$, где 1 – наличие речи в сегменте вокруг t .

$F_{SCD}: (h_t, h_{t+1}) \rightarrow \{0, 1\}$, где 1 – вероятность смены говорящего между t и $t + 1$.

Выходом этого модуля является разбиение исходного аудиопотока $X(t)$ на сегменты $Seg_1, Seg_2, \dots, Seg_K$, внутри которых речевая активность присутствует и диктор, предположительно, не меняется. Каждый сегмент Seg_k ассоциирован с соответствующим фрагментом признаков $H_k \subseteq H_{speech}$.

Вторым модулем является «Модуль идентификации диктора с механизмом внимания к эталонным голосам». Для каждого сегмента Seg_k необходимо определить, какому из целевых дикторов (сотрудников) он принадлежит. Вместо построения отдельной модели для каждого диктора мы используем парадигму мета-обучения (meta-learning) и механизм внимания [17].

Пусть, имеются эталонные записи голосов N целевых дикторов $D_1, D_2, \dots, D_N$. Для каждого диктора $D_i, i = 1(1)N$ из его эталонной записи (например, несколько произнесенных фраз) извлекаются эталонные эмбеддинги голоса. Для получения эталонных эмбеддингов каждого диктора в предложенной архитектуре используется предобученная модель Wav2Vec2, представляющая собой трансформерную архитектуру, обученную на крупномасштабных корпусах для задач распознавания речи [18]. Данная модель преобразует входной аудиосигнал в последовательность высокоуровневых признаков, эффективно кодирующих акустические характеристики голоса. Процедура извлечения эмбеддинга для каждого эталонного аудиофайла включает несколько последовательных этапов. Исходный сигнал приводится к единой частоте дискретизации 16 кГц и подвергается нормализации для устранения вариаций громкости. Нормализованный сигнал подаётся на вход модели Wav2Vec2, которая возвращает тензор признаков размерности $T \times G$, где T соответствует

количеству временных шагов, а G – размерности скрытого представления (для конфигурации base она составляет 768). Для получения фиксированного векторного представления всего аудиофрагмента выполняется операция усреднения по временной оси (mean pooling), в результате чего формируется вектор размерности G . Заключительным этапом является L2-нормализация полученного вектора, обеспечивающая устойчивость при последующих вычислениях косинусного сходства.

Таким образом, каждый эталонный файл диктора преобразуется в эмбединг – точку в пространстве фиксированной размерности, отражающую тембральные характеристики голоса. При наличии нескольких эталонных записей для одного диктора они образуют множество векторов, которое может быть использовано как непосредственно в механизме внимания, так и для вычисления усреднённого прототипа. В предложенной архитектуре механизм внимания оперирует всем множеством эталонных эмбедингов для каждого диктора, что позволяет учитывать естественную вариативность произношения и повышает робастность идентификации в условиях ограниченных данных.

Пусть $E_i = e_i^1, e_i^2, \dots, e_i^M$ – множество M_i векторов-эмбедингов диктора i , полученных с помощью предобученной сети (например, на основе x-vector архитектуры) [19]. Для сегмента Seg_k с признаками $H_k = (h_1, \dots, h_L)$, где L – длина сегмента во временных шагах после свёртки, в работе используется механизм внимания для вычисления взвешенного контекстного вектора, ориентированного на эталонные голоса.

Для каждого вектора признаков h_l из сегмента Seg_k вычисляется его совместимость со всеми эталонными эмбедингами e_j^m с помощью функции $score(h_l, e_j^m)$, где j – номер диктора ($j = 1, \dots, N$), а m – номер эталонной записи ($m = 1, \dots, M_j$). В качестве функции совместимости используется скалярное произведение (при этом векторы предполагаются нормированными). Затем веса внимания $a_{l,j,m}$ получаются применением функции $softmax$ по всем дикторам и их эталонам, что гарантирует их нормировку. Таким образом, элементы $a_{l,j,m}$ образуют трёхмерный тензор, отражающий вклад каждого эталонного эмбединга каждого диктора e_j^m в описание текущего фрагмента речи h_l :

$$a_{l,j,m} = \frac{\exp(score(h_l, e_j^m))}{\sum_{p=1}^N \sum_{q=1}^{M_p} \exp(score(h_l, e_p^q))} \quad (2)$$

где M_p – количество эталонных эмбедингов для диктора D_p , p – индекс перебора всех дикторов, а q – индекс эталонной записи для диктора p .

Затем для каждого диктора D_j вычисляется вектор внимания, элементы которого c_k^j , рассчитываются на основе признаков сегмента Seg_k :

$$c_k^j = \sum_{l=1}^L \sum_{m=1}^{M_j} a_{l,j,m} * e_j^m \quad (3)$$

где M_j – число эталонных эмбедингов диктора D_j , j – индекс целевого диктора, а k – индекс сегмента.

Здесь e_j^m эталонные эмбединги диктора D_j , а веса $a_{l,j,m}$ определяют вклад каждого эталонного вектора в описание текущего сегмента. Итоговый вектор z_k диктор-зависимых признаков для сегмента Seg_k формируется как конкатенация усредненных признаков сегмента $avg(H_k)$ и векторов внимания векторов внимания c_k^j вычисленных для каждого диктора D_j :

$$z_k = [avg(H_k); c_k^1; c_k^2; \dots; c_k^N] \quad (4)$$

где H_k – усреднённый по времени вектор признаков сегмента Seg_k , а N – общее число целевых дикторов.

Вектор z_k подается на полносвязный классификатор $F_{SPK}: z_k \rightarrow R^N$ с N выходами (по числу дикторов), который определяет вероятность принадлежности сегмента к каждому из

целевых дикторов. Идентифицированный диктор D_k для сегмента k определяется как $\operatorname{argmax}(F_{SPK}(z_k))$.

Третьим модулем является «Модуль транскрибации на основе преобразователя (Transformer)». Для непосредственного преобразования аудио в текст сегмента Seg_k используется архитектура типа Transformer-Encoder [20], которая показала высокую эффективность в ASR. Однако, чтобы улучшить распознавание для конкретных дикторов, мы конкретизируем работу декодера на идентифицированном дикторе. Признаки H_k сегмента Seg_k обрабатываются энкодером преобразователя:

$$O_k = \text{Transformer_Encoder}(H_k) \quad (5)$$

Декодер преобразователя, который генерирует последовательность текстовых токенов $Y_k = (y_1, y_2, \dots, y_T)$, модифицируется таким образом, что его начальное скрытое состояние инициализируется (conditioned) вектором эмбеддинга идентифицированного диктора e_{d_k} (усредненный эталонный эмбеддинг для d_k), где T – это длина сгенерированной последовательности, определяемая моментом появления специального токена конца последовательности ($\langle \text{eos} \rangle$) или достижением максимальной длины (в работе $T \geq 500$). Это позволяет декодеру адаптировать стиль генерации под акустические особенности конкретного говорящего.

$$y_t = \text{Transformer_Decoder}(y_{t-1}, O_k, e_{d_k}) \quad (6)$$

Обучение сети проводится в два этапа.

Предварительное обучение: Модуль сегментации предобучается на задаче VAD/SCD. Модуль идентификации диктора предобучается на большом наборе данных с множеством дикторов для обучения качественного эмбеддинг-слоя. Модуль транскрибации предобучается на массовом ASR-корпусе (например, LibriSpeech).

Тонкая настройка (Fine-tuning): Вся архитектура целиком дообучается на целевом наборе данных, содержащем записи совещаний с участием N целевых сотрудников. Используется комбинированная функция потерь $L_{total} = \alpha * L_{ASR} + \beta * L_{SPK}$, где L_{ASR} – распознавания речи (Connectionist Temporal Classification – CTC или Cross-Entropy), L_{SPK} – идентификации диктора (Cross-Entropy).

Результаты и обсуждение. В результате исследования разработана формальная архитектура нейронной сети (рисунок 1 и таблица 1). Основные теоретические результаты заключаются в следующем:

1. Предложена трехмодульная математическая модель обработки многопользовательского аудиопотока, формализованная системой уравнений (1-6);
2. Разработан и формально описан механизм идентификации диктора на основе внимания к эталонным эмбеддингам, который позволяет эффективно использовать обучения с малым количеством примеров;
3. Предложен метод условного управления декодером транскриптора на основе идентифицированного диктора. Как показано в ряде исследований [21–22], conditioning декодера на информацию о дикторе позволяет модели фокусироваться на акустических характеристиках целевого говорящего и улучшает точность распознавания, особенно в условиях перекрывающейся речи и множества дикторов;
4. Формализован двухэтапный процесс обучения, обеспечивающий устойчивость модели при ограниченности целевых данных.

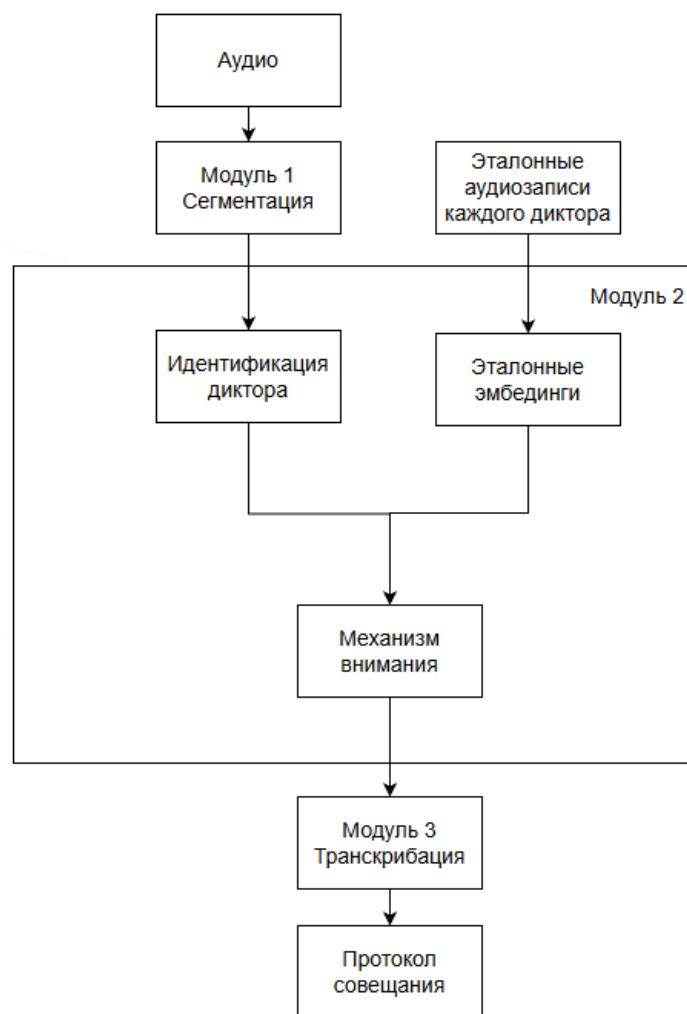


Рис. 1. Общая архитектура предлагаемой гибридной нейронной сети

Таблица 1. Сравнение функций модулей предлагаемой архитектуры

Модуль	Программный аппарат	Входные данные	Выходные данные
1. Сегментации	Глубокая CNN, VAD, SCD	Мел-спектрограмма S	Сегменты речи Seg_k , признаки H_k
2. Идентификации диктора	Механизм внимания, мета-обучение	H_k , эталонные эмбеддинги E_i	Идентификатор диктора d_k
3. Транскрибации	Transformer-Encoder/Decoder	H_k, e_{d_k}	Текстовая транскрипция Y_k

Для оценки эффективности предложенной трёхмодульной гибридной архитектуры были проведены эксперименты на реальных записях корпоративных совещаний общей длительностью более 20 часов. В качестве целевых дикторов выступили пять участников дискуссий, для каждого из которых было подготовлено от 7 до 11 эталонных аудиозаписей длительностью 10–30 секунд. Обучение проводилось в режиме тонкой настройки предобученных компонентов (Wav2Vec2, Whisper) на собранном корпусе.

В таблице 2 представлены полученные метрики качества для каждого модуля разработанной системы.

Таблица 2. Метрики качества модулей гибридной архитектуры

Модуль	Метрика	Значение
Сегментация	VAD Accuracy	0,7940
	VAD F1-score	0,8852
	SCD Precision	0,6245
	SCD Recall	0,4167
	SCD F1-score	0,5000
Идентификация диктора	Speaker Accuracy	0,6254

Модуль сегментации продемонстрировал высокую эффективность в задаче обнаружения речевой активности (VAD F1 = 0,8852), что подтверждает корректность выбора глубокой свёрточной сети для выделения универсальных речевых признаков. Высокий VAD Accuracy (0,7940) свидетельствует о способности модели надёжно выделять речь в условиях реальных совещаний.

В задаче детекции смены диктора (SCD) наблюдается характерный для таких задач дисбаланс классов, при умеренном значении полноты (Recall = 0,4167) точность детекции (Precision = 0,6245) указывает на то, что примерно 62% определяемых смен спикера являются истинными, однако при этом часть реальных смен не идентифицируется. Данный эффект объясняется разреженностью событий смены диктора в обучающей выборке. Улучшение этой метрики требует применения методов борьбы с дисбалансом, таких как взвешенные функции потерь или аугментация данных, что выходит за рамки данной работы.

Модуль идентификации диктора, реализующий предложенный механизм внимания к эталонным эмбедингам, показал результат 62,5% точности при идентификации пяти дикторов. Данный результат следует признать высоким, учитывая следующие факторы:

- обучение проводилось в условиях *few-shot learning* (до 11 эталонных записей на диктора);
- предложенная архитектура не требует переобучения при добавлении новых дикторов;
- идентификация выполнялась в условиях естественных диалогов с фоновым шумом и перекрытием речи.

Тем не менее, полученное значение существенно превосходит случайное угадывание (20%) и подтверждает эффективность предложенного подхода с конкатенацией усреднённых признаков сегмента и взвешенных сумм векторов внимания.

Транскрибация в разработанной архитектуре изначально предполагалась, как встроенный модуль на базе трансформера. Однако экспериментально установлено, что для достижения приемлемого качества распознавания речи требуются значительно большие объёмы текстовых транскрипций (50–100 часов), что не всегда доступно в корпоративной среде. В связи с этим в практической реализации системы принято решение использовать гибридный подход: модули сегментации и идентификации работают в предложенной архитектуре, а транскрибация каждого выделенного сегмента выполняется предобученной моделью Whisper. Это обеспечивает высокое качество распознавания речи без необходимости дополнительного обучения и полностью соответствует корпоративным сценариям использования.

Заключение. Полученные результаты дают основание утверждать, что разработанная архитектура работоспособна в обозначенных ранее условиях.

Интеграция задач в единой модели обеспечила согласованную работу модулей сегментации и идентификации. Так, достаточно высокое качество детекции речевой активности (VAD F1 = 0,885) позволяет с приемлемым качеством выделять реплики, а точность идентификации диктора (62,5% для пяти участников) демонстрирует, что механизм внимания к эталонным эмбедингам успешно обучается и способен сопоставлять признаки

сегмента с голосовыми эталонами, что подтверждает, необходимость совместного обучения, позволяющего избегать накопления ошибок, характерного для отдельных конвейеров.

Устойчивость к ограниченным данным является ключевым преимуществом предложенного подхода, так как применение механизма внимания к эталонным записям является эффективной реализацией *few-shot learning*, ввиду того, что модель не требует сбора больших размеченных наборов для каждого нового диктора. Всего 7-11 эталонных записей на одного спикера обеспечивают точность идентификации равной 62,5%, что значительно выше случайного уровня (20%). Это подтверждает применимость архитектуры в корпоративной среде, где количество данных для новых сотрудников весьма ограничено.

Детекция смены диктора (SCD) показала умеренные результаты ($F1 \approx 0,5$). Причина этого – сильный дисбаланс классов: в реальных диалогах смены происходят относительно редко, и модель испытывает трудности с их обнаружением. Для улучшения этого компонента целесообразно использовать взвешенные функции потерь и методы аугментации (улучшения, расширения или добавления новых качеств данных), например, искусственное создание смен путём склейки сегментов разных дикторов.

Транскрибация речи в исходной архитектуре предполагала встроенный декодер на трансформере, однако проведенные исследования показали, что для качественного обучения потребуется значительное количество размеченных текстовых данных (более 50 часов), что не всегда доступно. В связи с этим в практической реализации было принято решение использовать внешнюю ASR-систему Whisper для распознавания текста выделенных сегментов. Такой гибридный подход сохраняет модульность, не снижая качества идентификации, и позволяет получать точные транскрипции без дополнительного обучения.

Основным ограничением остаётся вычислительная сложность механизма внимания при большом числе дикторов, однако для корпоративных совещаний с типичным числом участников до 10 это не столь критично. Кроме того, качество идентификации напрямую зависит от чистоты эталонных записей – в реальных условиях может потребоваться предварительная очистка аудио.

Список источников

1. Ермоленко Т.В. Система автоматического распознавания слитной русской речи на основе глубоких нейросетей / Т.В. Ермоленко, Я.С. Пикалёв // Речевые технологии, 2021. – № 1-2. – С. 3-18. – DOI: 10.58633/2305-8129_2021_1-2_3.
2. Mujtaba D., Mahapatra N., Arney M. et al. Lost in transcription: identifying and quantifying the accuracy biases of automatic speech recognition systems against disfluent speech. Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), 2024, pp. 4795-4809, DOI: 10.18653/v1/2024.naacl-long.269.
3. Cornell S., Wiesner M.S., Watanabe S. et al. The CHiME-7 DASR challenge: distant meeting transcription with multiple devices in diverse scenarios. 7th International Workshop on Speech Processing in Everyday Environments (CHiME 2023), 2023, DOI: 10.21437/CHiME.2023-1.
4. Sharma M., Joshi S., Chatterjee T. et al. A comprehensive empirical review of modern voice activity detection approaches for movies and TV shows. Neurocomputing, 2022, vol. 494, pp. 116-131, DOI: 10.1016/j.neucom.2022.04.084.
5. Суин А.Е. Использование каскадов моделей для резюмирования встреч / А.Е. Суин, В.О. Серебряков, Т.Ю. Чернышева // Промышленный искусственный интеллект (ПИИ'2025): Всероссийская научно-практическая конференция. – СПб.: ПОЛИТЕХ-ПРЕСС, 2025. – С. 439–447. – DOI: 10.18720/SPBPU/2/id25-548.
6. Khadar K.V. A., Kumar R.K. S., Sreekanth N.S. Adapting to noise in forensic speaker verification using GMM-UBM I-vector method in high-noise backgrounds. Computational Sciences and Sustainable Technologies (Communications in Computer and Information Science), 2024, pp. 275-287, DOI: 10.1007/978-3-031-50993-3_22.
7. Abd El-Moneim S., El-Shafai W., Hammam H. et al. Effect of interference on text-independent speaker recognition based on deep learning. Multimedia Tools and Applications, 2025, vol. 84, pp. 47425–47458, DOI: 10.1007/s11042-024-19493-1.

8. An K., Zhang S., Yan Z. Are transformers in pre-trained LM a good ASR encoder? an empirical study. arXiv (Cornell University), 2024, DOI: 10.48550/arXiv.2409.17750.
9. Hirano Y., Sakti S. SC-SOT: conditioning the decoder on diarized speaker information for end-to-end overlapped speech recognition. arXiv (Cornell University), 2025, DOI: 10.48550/arXiv.2506.12672.
10. Jakubec M., Jarina R., Jakubcova E. et al. Deep speaker embeddings for speaker verification: review and experimental comparison. Engineering Applications of Artificial Intelligence, 2023, vol. 123, art. 107025, DOI: 10.1016/j.engappai.2023.107232.
11. Krizan S., Beliaev S., Ginsburg B. et al. Quartznet: deep automatic speech recognition with 1D time-channel separable convolutions. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 6124-6128, DOI: 10.1109/ICASSP40776.2020.9053889.
12. Latif S., Zaidi A., Cuayahuitl H. et al. Transformers in speech processing: a survey. arXiv (Cornell University), 2025, DOI: 10.48550/arXiv.2303.11607.
13. Mavaddati S. A voice activity detection algorithm using deep learning in time–frequency domain. Neural Computing and Applications, 2025, vol. 37, pp. 2581–2595, DOI: 10.1007/s00521-024-10795-x.
14. McFee B., Raffel C., Liang D. et al. Librosa: audio and music signal analysis in python. 14th Conference: Python in Science Conference, 2015, pp. 18-25, DOI: 10.25080/Majora-7b98e3ed-003.
15. Polok A., Klemen D., Kocour M. et al. DiCoW: diarization-conditioned whisper for target speaker automatic speech recognition. Computer Speech and Language, 2026, vol. 95, no. 1, pp. 1-19, DOI: 10.1016/j.csl.2025.101841.
16. Rafidison M.A., Rakotomihamina A.H., Rajaonarison F.T.M. et al. Intervention of light convolutional neural network in document survey form processing. Multimedia Tools and Applications, 2023, vol. 82, pp. 32583–32605, DOI: 10.1007/s11042-023-16076-4.
17. Snyder D., Garcia-Romero D., Sell G. et al. X-vectors: robust DNN embeddings for speaker recognition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5329-5333, DOI: 10.1109/ICASSP.2018.8461375.
18. Baeviski A., Zhou H., Mohamed A., Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations. arXiv preprint arXiv:2006.11477, 2020, DOI: 10.48550/arXiv.2006.11477.
19. Su H., Zhao D., Dang L. et al. A multitask learning framework for speaker change detection with content information from unsupervised speech decomposition. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 8087-8091, DOI: 10.1109/ICASSP43922.2022.9746116.
20. Szwoch G. Crosstalk suppression in a multi-channel, multi-speaker system using acoustic vector sensors. MDPI Sensors, 2025, vol. 23, issue 21, pp. 6731, DOI: 10.3390/s25216731.
21. Wang W., Zhao H., Yang Y. et al. Few-shot short utterance speaker verification using meta-learning. PeerJ Computer Science, 2023, 9:e1276, DOI: 10.7717/peerj-cs.1276.
22. Cui C., Sheikh I., Sadeghi M., Vincent E. End-to-end multichannel speaker-attributed ASR: speaker guided decoder and input feature analysis. IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2023, DOI: 10.1109/ASRU57964.2023.10389729.

Потапов Арсений Денисович. Студент магистратуры Иркутского национального исследовательского технического университета, эксперт по цифровым технологиям аудита Байкальского банка ПАО Сбербанк. ORCID: 0009-0008-5949-1293, arseny.potapov2012@yandex.ru, 664074, г. Иркутск, ул. Лермонтова 83.

Лукьянов Никита Дмитриевич. Кандидат технических наук, руководитель центра электронного обучения, доцент лаборатории сетевых систем и ИТ-инфраструктуры, Иркутского национального исследовательского технического университета. Направления исследований: алгоритмы машинного обучения и искусственного интеллекта, автоматизация в промышленности, математическое моделирование. AuthorID: 732737, SPIN: 8971-7440, ORCID: 0000-0002-8826-3424, lukyanovnd@ex.istu.edu, 664074, г. Иркутск, ул. Лермонтова 83.

Adaptive hybrid neural network architecture with attention for speaker identification in corporate meetings

Arseniy D. Potapov¹, Nikita D. Lukyanov²

¹PJSC Sberbank of Russia, Russia, Irkutsk

²Irkutsk National Research Technical University, Russia, Irkutsk, lukyanovnd@istu.edu

Abstract. The relevance of this research stems from the growing need to automate corporate meeting minutes, a labor-intensive and time-consuming process with a high risk of errors. Current speech recognition systems (ASR – Automatic Speech Recognition) perform poorly in multi-user conversations, with overlapping voices, background noise, and the use of specialized vocabulary. Speaker identification (SID – Speaker Identification) is particularly challenging when there are few reference recordings per employee, which is typical in an office environment. Acoustic environments also pose additional challenges: echoing offices, simultaneous speech, and poor voice separation. Therefore, the goal of this study is to develop a hybrid neural network model for accurate transcription and reliable speech attribution in group discussions. This paper analyzes the weaknesses of ASR systems and SID methods and proposes a three-tier architecture: audio stream segmentation (VAD – Voice Activity Detection, SCD – Speaker Change Detection), voice embedding identification with a reference attention mechanism, and speaker-aware transcription. A mathematical justification for the key components is also provided. The theoretical significance lies in the development of an attention-based SID method that effectively solves few-shot learning problems. Unlike traditional approaches that require large labeled corpora, the new architecture learns to match speech fragments with a small number of reference recordings. The effectiveness of the method has been empirically confirmed using real corporate meeting recordings (over 20 hours, five participants). The segmentation module demonstrated a VAD F1 score of 0.885, and the speaker identification module achieved an accuracy of 62.5%, which is three times higher than chance (20%). Transcription was performed using the external Whisper system, which guarantees high recognition accuracy. The results enable the development of a system for automatic meeting minutes, and the architecture supports a large number of participants and is easily adapted to various communication formats: from negotiations to client meetings. The functionality of the hybrid neural network model can be used to create various tools for corporate meeting minutes.

Keywords: Automatic speech recognition, speaker identification, hybrid neural network architecture, attention mechanism, voice embeddings, audio stream segmentation, transcription, corporate meetings

References

1. Ermolenko T.V., Pikalev Ya.S. Sistema avtomaticheskogo raspoznavaniya slitnoy russkoy rechi na osnove glubokikh neyrosetey [System for automatic continuous Russian speech recognition based on deep neural networks]. *Rechevyye tekhnologii* [Speech Technologies], 2021, no. 1-2, pp. 3-18, DOI: 10.58633/2305-8129_2021_1-2_3.
2. Mujtaba D., Mahapatra N., Arney M. et al. Lost in transcription: identifying and quantifying the accuracy biases of automatic speech recognition systems against disfluent speech. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, 2024, pp. 4795-4809, DOI: 10.18653/v1/2024.naacl-long.269.
3. Cornell S., Wiesner M.S., Watanabe S. et al. The CHiME-7 DASR challenge: distant meeting transcription with multiple devices in diverse scenarios. *7th International Workshop on Speech Processing in Everyday Environments (CHiME 2023)*, 2023, DOI: 10.21437/CHiME.2023-1.
4. Sharma M., Joshi S., Chatterjee T. et al. A comprehensive empirical review of modern voice activity detection approaches for movies and TV shows. *Neurocomputing*, 2022, vol. 494, pp. 116-131, DOI: 10.1016/j.neucom.2022.04.084.
5. Suin A.E., Serebryakov V.O., Chernysheva T.Yu. Ispol'zovanie kaskadov modeley dlya rezyumirovaniya vstrech [Using cascades of models for meeting summarization]. *Promyshlennyy iskusstvennyy intellekt (PII'2025): Vserossiyskaya nauchno-prakticheskaya konferentsiya [Industrial Artificial Intelligence (IAI'2025): All-Russian Scientific and Practical Conference]*, St. Petersburg, POLYTECH-PRESS, 2025, pp. 439-447. DOI: 10.18720/SPBPU/2/id25-548.
6. Khadar K.V. A., Kumar R.K. S., Sreekanth N.S. Adapting to noise in forensic speaker verification using GMM-UBM I-vector method in high-noise backgrounds. *Computational Sciences and Sustainable Technologies (Communications in Computer and Information Science)*, 2024, pp. 275-287, DOI: 10.1007/978-3-031-50993-3_22.

7. Abd El-Moneim S., El-Shafai W., Hammam H. et al. Effect of interference on text-independent speaker recognition based on deep learning. *Multimedia Tools and Applications*, 2025, vol. 84, pp. 47425–47458, DOI: 10.1007/s11042-024-19493-1.
8. An K., Zhang S., Yan Z. Are transformers in pre-trained LM a good ASR encoder? an empirical study. *arXiv (Cornell University)*, 2024, DOI: 10.48550/arXiv.2409.17750.
9. Hirano Y., Sakti S. SC-SOT: conditioning the decoder on diarized speaker information for end-to-end overlapped speech recognition. *arXiv (Cornell University)*, 2025, DOI: 10.48550/arXiv.2506.12672.
10. Jakubec M., Jarina R., Jakubcova E. et al. Deep speaker embeddings for speaker verification: review and experimental comparison. *Engineering Applications of Artificial Intelligence*, 2023, vol. 123, art. 107025, DOI: 10.1016/j.engappai.2023.107232.
11. Krizan S., Beliaev S., Ginsburg B. et al. Quartznet: deep automatic speech recognition with 1D time-channel separable convolutions. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 6124-6128, DOI: 10.1109/ICASSP40776.2020.9053889.
12. Latif S., Zaidi A., Cuayahuitl H. et al. Transformers in speech processing: a survey. *arXiv (Cornell University)*, 2025, DOI: 10.48550/arXiv.2303.11607.
13. Mavaddati S. A voice activity detection algorithm using deep learning in time–frequency domain. *Neural Computing and Applications*, 2025, vol. 37, pp. 2581–2595, DOI: 10.1007/s00521-024-10795-x.
14. McFee B., Raffel C., Liang D. et al. Librosa: audio and music signal analysis in python. *14th Conference: Python in Science Conference*, 2015, pp. 18-25, DOI: 10.25080/Majora-7b98e3ed-003.
15. Polok A., Klemen D., Kocour M. et al. DiCoW: diarization-conditioned whisper for target speaker automatic speech recognition. *Computer Speech and Language*, 2026, vol. 95, no. 1, pp. 1-19, DOI: 10.1016/j.csl.2025.101841.
16. Rafidison M.A., Rakotomihamina A.H., Rajaonarison F.T.M. et al. Intervention of light convolutional neural network in document survey form processing. *Multimedia Tools and Applications*, 2023, vol. 82, pp. 32583–32605, DOI: 10.1007/s11042-023-16076-4.
17. Snyder D., Garcia-Romero D., Sell G. et al. X-vectors: robust DNN embeddings for speaker recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 5329-5333, DOI: 10.1109/ICASSP.2018.8461375.
18. Baevski A., Zhou H., Mohamed A., Auli M. wav2vec 2.0: a framework for self-supervised learning of speech representations. *arXiv preprint arXiv:2006.11477*, 2020, DOI: 10.48550/arXiv.2006.11477.
19. Su H., Zhao D., Dang L. et al. A multitask learning framework for speaker change detection with content information from unsupervised speech decomposition. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 8087-8091, DOI: 10.1109/ICASSP43922.2022.9746116.
20. Szwoch G. Crosstalk suppression in a multi-channel, multi-speaker system using acoustic vector sensors. *MDPI Sensors*, 2025, vol. 23, issue 21, pp. 6731, DOI: 10.3390/s25216731.
21. Wang W., Zhao H., Yang Y. et al. Few-shot short utterance speaker verification using meta-learning. *PeerJ Computer Science*, 2023, 9:e1276, DOI: 10.7717/peerj-cs.1276.
22. Cui C., Sheikh I., Sadeghi M., Vincent E. End-to-end multichannel speaker-attributed ASR: speaker guided decoder and input feature analysis. *IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2023, DOI: 10.1109/ASRU57964.2023.10389729.

Potapov Arseniy Denisovich. Master's student at Irkutsk National Research Technical University, expert in digital technologies of audit of Baikal Bank of Sberbank. ORCID: 0009-0008-5949-1293, arseny.potapov2012@yandex.ru, 664074, Irkutsk, 83 Lermontov Street.

Lukyanov Nikita Dmitriyevich. PhD in Engineering, Head of the E-Learning Center, Associate Professor at the Laboratory of Network Systems and IT Infrastructure at Irkutsk National Research Technical University. Research areas: machine learning and artificial intelligence algorithms, industrial automation, mathematical modeling. AuthorID: 732737, SPIN: 8971-7440, ORCID: 0000-0002-8826-3424, lukyanovnd@ex.istu.edu, 664074, Irkutsk, 83 Lermontov Street.

Статья поступила в редакцию 07.04.2026; одобрена после рецензирования 25.05.2026; принята к публикации 29.07.2026.

The article was submitted 04/07/2026; approved after reviewing 05/25/2026; accepted for publication 07/29/2026.

Интеллектуальная автоматизированная обучающая система AESU и лежащие в её основе концепции, механизмы и методы

Углев Виктор Александрович^{1,2}, Смирнов Георгий Артёмович¹

¹Сибирский федеральный университет, Россия, Железногорск, vauglev@sfu-kras.ru

²Московский университет МВД России им. В.Я. Кикотя, Россия, Москва

Аннотация. Проблема системного и научно-обоснованного подхода к организации электронного обучения повышает свою актуальность в условиях массового применения больших языковых моделей. В статье рассматривается проблема концептуального методического обоснования работы интеллектуальных автоматизированных обучающих систем (ИАОС). Показана проблема описания и раскрытия в научных источниках внутренних механизмов их работы, позволяющих верифицировать и автоматически объяснять принимаемые ими решения. В качестве примера рассмотрена авторская экспериментальная обучающая система AESU. Для неё рассматриваются базовые концепции, механизмы и методические решения. В качестве основных концепций приведены теория функциональных систем П.К. Анохина (модель афферентного синтеза), прикладная семиотика Д.А. Поспелова и теория рефлексивного управления В.А. Лефевра. Показана роль когнитивной визуализации (картирования) при синтезе решений. Для этого объединяется логика продукционных экспертных систем, нечёткой логики и проверки гипотез по Шортлиффу – интеллектуальных механизмов решателя обучающей системы. Эмерджентный эффект, возникающий от использования приведённых концепций и механизмов, позволяет предложить авторский подход к реализации модульной концентрации данных и сквозного анализа учебной ситуации. Системообразующим элементом в синтезе решений ИАОС становятся структурная и функциональная нотация карт: метод когнитивных карт диагностики знаний и метод унифицированного графического воплощения активности. Показана возможность взаимного перехода параметрического и визуального отображения учебной ситуации между этими типами карт и их использование при выработке решений как решателем обучающей системы, так и учащимся в процессе работы в электронной обучающей среде. В заключении отмечены архитектурные особенности ИАОС AESU. Кроме типовых функций систем управления обучением, предложенная ИАОС реализует диалоговый механизм взаимодействия с учащимся в естественно-языковой форме без привлечения возможностей больших языковых моделей: гибкий механизм обратной связи позволяет поддерживать интерактивный диалог, опираясь на методически обоснованные выводы. Это позволяет повысить уровень доверия человека-учащегося к рекомендациям обучающей системы и способствует вовлечению в образовательный процесс на базе ИАОС.

Ключевые слова: электронное обучение, интеллектуальные автоматизированные обучающие системы, когнитивная визуализация, принятие решений, метод

Цитирование: Углев В.А. Интеллектуальная автоматизированная обучающая система AESU и лежащие в её основе концепции, механизмы и методы / В.А. Углев, Г.А. Смирнов // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 42-55. – DOI:10.25729/ESI.2026.43.3.004.

Введение. Функциональные возможности *автоматизированных обучающих систем*, опирающихся на методы искусственного интеллекта, во многом определяются теми концепциями и механизмами, которые лежат в их основе. Большие языковые модели, принимающие решения по принципу «чёрного ящика», ещё недостаточно изучены и несут в себе ряд скрытых рисков: от методической необоснованности и до нежелательных когнитивных воздействий на учащихся [1]. Поэтому базовая технология, лежащая в основе обучающих систем, должна позволять аргументировано принимать решения по выработке педагогических воздействий и аргументировать их учащемуся, если такой запрос от него поступит. Выбор технологий, отражающих крайне противоположные подходы к организации процесса выработки решений, на практике не оправдывают себя. Так, фанатичное использование нейронных сетей (например, в [2]) не даёт эффекта объяснимости решений, а использование онтологий (например, в [3]) заставляет скрупулезно прорабатывать все варианты решений на уровне каждой учебной задачи. Очевидно, что проработка логики решений должна базироваться не только на частных

случаях, но и опираться на определенный блок методических знаний. Для этого эти знания необходимо как-то хранить, организовывать их обработку с опорой на специальные модели.

Интегрированные в состав интеллектуальных автоматизированных обучающих систем (ИАОС) модели определяют потенциальную гибкость в логике принимаемых решений. Далеко не все обучающие системы включают, помимо стандартных моделей ученика и электронного курса [4], модели учителя, тьютора [5] и учебной ситуации [6]. Поэтому от выбранных разработчиками механизмов и моделей будет зависеть как архитектура ИАОС, так и гибкость и результативность принятия решений интеллектуальным планировщиком. В некоторых программных решениях информация о механизмах и моделях явно описывается разработчиками в научной литературе [7,8]. По этой информации можно оценить возможности и конкурентные преимущества системы. К сожалению, не для каждой ИАОС можно легко найти четко изложенные данные о лежащих в их основе механизмах и реализованных методах. Занимаясь разработкой собственной ИАОС AESU, мы также раскрывали в своих публикациях различные аспекты её работы. Но сейчас становится очевидным, что без последовательного раскрытия этого вопроса специфика её функционирования может быть не ясна или понята неправильно. Поэтому нами было принято решение описать базовые концепции, механизмы и методические подходы к построению ИАОС AESU, определяющие её архитектуру.

ИАОС AESU была сформирована из ранее созданных программных модулей, изначально реализованных в виде самостоятельных настольных приложений, а затем в виде единой сетевой системы (браузерный вариант). К 2020 году она приобрела вид экспериментальной системы управления обучением (Learning Management System, LMS) с расширенными функциональными возможностями. В её основу легли теории П.К. Анохина о поведенческом акте, прикладная семиотика Д.А. Поспелова, теории В.А. Лефевра о рефлексивном управлении, А.А. Зенкина о когнитивной компьютерной графике. Основными механизмами стали теория продукционных экспертных систем Э. Поста, нечёткой логики Л. Заде, проверки экспертных мнений Шортлиффа и Бьюкенена, когнитивных карт и пиктографики. В связи с необходимостью детальной проработки логики взаимодействия с моделями, постоянно модифицировались базы знаний и механизмы обобщения. Для упрощения процесса обновления логики работы ИАОС и объединения всего многообразия подходов в единую систему, потребовалось предложить ряд оригинальных решений: модели этапной и модульной концентрации данных, а также сквозной подход к оценке учебной ситуации. Дадим краткую характеристику каждой из концепций и механизмов в рамках их использования в рамках ИАОС AESU, а также раскроем суть предложенных методов и отметим их место в архитектуре системы.

Концепции. Базовые концепции и теории ориентированы на организацию данных и знаний, позволяя обеспечить процесс объяснимого принятия решений. Само принятие решений опирается на модель M , которая для решателя ИАОС является чёрным ящиком и входит в состав модульной базы знаний. Для извлечения и концентрации исходных данных используется модель *афферентного синтеза* из теории функциональных систем П.К. Анохина [9]. Суть её заключается в том, что для принятия решений необходимо подготовить исходные данные, включающие описание учебной ситуации (обстановочная афферентация); цифровой образовательный след с прецедентами (имеющийся опыт); номенклатуру событий (пусковую афферентацию); а также работу с целями (акцептор результата). Такая «физиологическая» интерпретация механизма разумного поведения была предложена самим Анохиным в качестве основы для работы интеллектуальных систем [10] и нами лишь адаптирована к специфике ИАОС [11]. Такой массив исходных данных D формирует «поле» признакового пространства, которое должно в дальнейшем подвергаться обработке планировщиком.

Процедура сбора исходных данных не решает проблему её упорядочивания непосредственно для принятия решений. По этой причине декомпозиция данных, включенных в процесс принятия решения, дополнительно делится на шаблонное представление семантики, синтактики и прагматики, дополненного конкретизированными для текущей учебной ситуации данными (соответствует *квадрату Пospelова* [12] из прикладной семиотики [13]). Роль синтактики (структуры данных) в таком представлении играет модель дидактического материала, иерархии целей и нормативов (тех же компетенций учебного плана). Роль семантики выполняют логические связи между различными элементами (как с моделями, так и внутри структур данных), а также фрагменты баз знаний. Прагматикой в этом подходе стали модели реакций моделей учащегося, учителя и тьютора (соответствующие фрагменты базы знаний). Непосредственно оперативные данные, их динамика («выжимка» из цифрового образовательного следа) и параметры модели ученика формируют денотат [14]. В результате такого представления все данные после загрузки соотносятся с элементами семиотического представления (размечаются в поле признакового пространства) и могут быть подгружены при обработке события класса k при принятии решений.

Характерной чертой развитых систем искусственного интеллекта, среди прочего, является их способность к рефлексии собственных решений и оперированию расширенной моделью человека, с которым происходит взаимодействие [15]. Эту задачу в своей *теории рефлексивного управления* замечательно формализовал В.А. Лефевр [16]. Мы использовали элементы его подхода не только к обоснованию выделения различных сущностей (модели ученика, учителя, тьютора, методиста), но и различную интерпретацию ими учебной ситуации. Очевидно, что все эти модели имеют не только собственные цели, но и логику оценки важности, нужности, полезности и пр. как различных элементов дидактического материала, так и решений со стороны ИАОС [17]. Кроме того, предложенный подход позволил выделить ряд аспектов анализа (знаниевый, компетентностный, целевой), учёт которых углубляет возможности по анализу учебной ситуации. Так появляются признаки точки зрения на проблему i и аспекта анализа j .

Ещё одна концепция, которую вобрала в себя ИАОС AESU, стала когнитивная визуализация, относимая А.А. Зенкиным и Д.А. Поспеловым к разделу *когнитивной компьютерной графики* [18]. Применительно к обучающим системам Д.А. Поспелов писал следующее: «даже весьма робкие попытки в этом направлении, ... привлекающие сейчас пристальное внимание специалистов (особенно тех, кто занят созданием интеллектуальных обучающих систем), показывает перспективность подобных исследований» [19]. Именно эту идею в виде концепции функционально-структурного картирования *Map* мы взяли на вооружение как при концентрации данных в процессе принятия решений, так и при предъявлении её человеку-учащемуся. Принцип наглядности и понятности решений за счёт графической концентрации был взят из Data Mining (раздела Data Visualization [20]) и является востребованным в рамках современных ИАОС [21].

Использование системного подхода к объединению всех этих концепций в рамках одного инструмента (подсистемы принятия решений) позволило выйти на обобщённую последовательность обобщения данных и выработки решений (1).

$$M_k : D \rightarrow \{D^{ij}\} \rightarrow Map\{D^{ij}\} \rightarrow F \rightarrow r \quad (1)$$

где M – модель принятия решений для события типа k ; D – признаковое пространство; i – точка зрения на решение; j – аспект анализа учебной ситуации; *Map* – механизм картирования; F – процедуры принятия решений; а r – решение (выбранная альтернатива из множества решений R_k). Эта модель предполагает наличие частных механизмов.

Механизмы. Частные механизмы, которые обеспечивают функциональность ИАОС, делятся на штатные процедуры (выборки, статистические обобщения, процедуры расчёта оценок

отдельных показателей) и специфические. Первые используются на всех этапах обобщения данных, согласно настройкам модели принятия решений *M*. Они не опираются на методы искусственного интеллекта и подключаются не из базы знаний, а из репозитория алгоритмов обучающей системы.

Специфические механизмы ориентированы на то, чтобы планировщик мог проверять различные гипотезы, опираясь на многообразие методов искусственного интеллекта. Кратко опишем основные из них.

Механизм проверки гипотез на основе продукционной модели представления знаний Э. Поста реализован на базе конструктора FLM_Builder [22] и позволяет формировать внешние фрагменты базы знаний в формате flm (рис. 1, справа). Они размещаются в репозитории баз знаний и ассоциируются с конкретными событиями образовательной среды через модель *M*. Машина вывода (решатель) для просчёта flm-файлов реализована на языке Python, ядро которого интегрировано в серверную часть ИАОС AESU. Таким образом, в обучающую систему встроен механизм проверки гипотез с помощью *продукционных экспертных систем*, формализующих граф вывода знаний на базе коллекции эвристик.

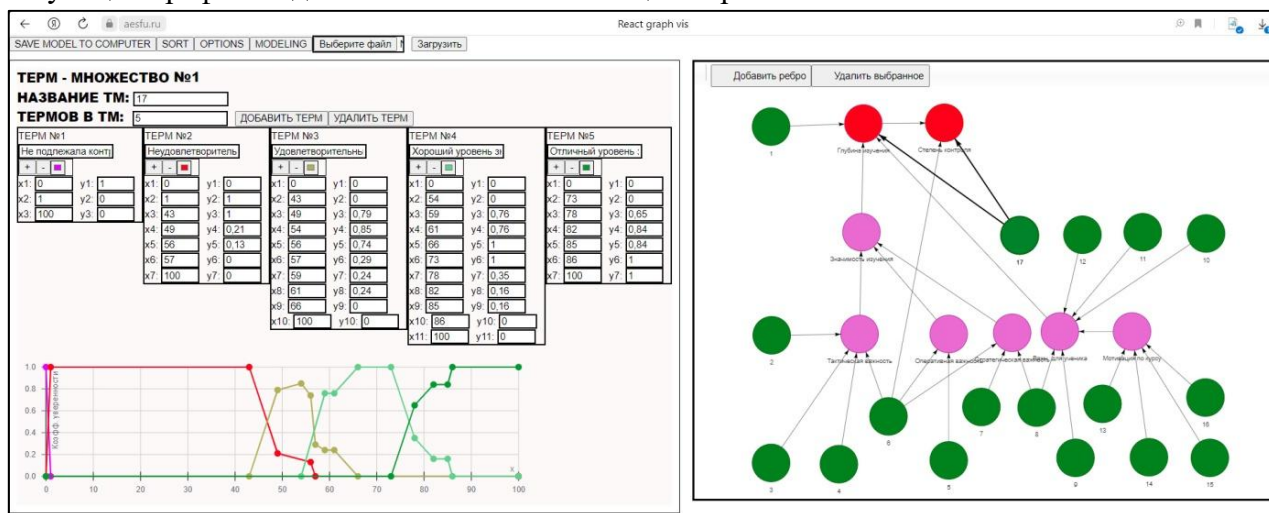


Рис. 1. Окно конструктора экспертных систем FLM_Builder: формирование графа решений (справа) и характеристических функций терм-множеств (слева)

Самостоятельным механизмом, интегрированным с flm-моделями, является метод *нечёткой логики* Л. Заде [23]. Он может быть подключён на любом этапе работы модели *M* и реализует функции фаззификации и дефаззификации. Настройка термов и терм-множеств обеспечивается в том же конструкторе FLM_Builder, а запуск просчёта с учётом выбора предпочтительного алгоритма разрешения конфликтов реализован через соответствующий API (рис. 1, слева).

Ещё одним механизмом проверки гипотез является использование *критерия уверенности Шортлиффа и Бьюкенена* [24]. Он предполагает, что каждая сложная по своей сути реакция/событие в ИАОС могут быть учтены при анализе с разных сторон: одни гипотезы в процессе появления очередного события могут повышать свой уровень достоверности, а другие – понижать. Для этого для анализируемого события формируется множество гипотез и для каждой из них эксперт выставляет свои оценки, которые в дальнейшем позволяют оценить, насколько эти гипотезы подтвердились на множестве событий. В дальнейшем результаты проверки используются как при формировании векторов входов в flm-файлы, так и для осуществления картирования. В первую очередь этот механизм применяется при формировании компетентного профиля учащихся [25].

Переходя от концепции к механизмам когнитивной компьютерной графики, следует остановиться на механизмах картирования. Собственно, механизм картирования в ИАОС

AESU представлен двумя видами карт: первые ориентированы на свёртку и концентрацию данных относительно структуры дидактического материала, а вторые – на свёртку и концентрацию относительно активности учащегося (функциональный аспект). Оба они опираются на подготовленную в $Map\{D^{ij}\}$ структуру данных (параметрическую карту), но обрабатывают и анализируют её по-разному (в зависимости от специфики решаемой задачи).

Структурное картирование реализовано в виде *когнитивных карт диагностики знаний* (ККДЗ) [26] и используется при анализе учебной ситуации с позиции обучающей системы (рефлексия). Она представляет из себя граф вида модели «маленького мира» [27], который соединяет в себе все семиотические элементы учебной ситуации из D^{ij} , используя в качестве подложки структуру дидактического материала. На карту дополнительно вводится цветовая маркировка и возможность отображения данных из связанных структурных элементов учебного плана (сводная карта) [28]. Например, если в качестве масштаба взята учебная дисциплина, то её структура для учащегося на определённый момент обучения будет выглядеть так, как показано на рис. 2 (слева).

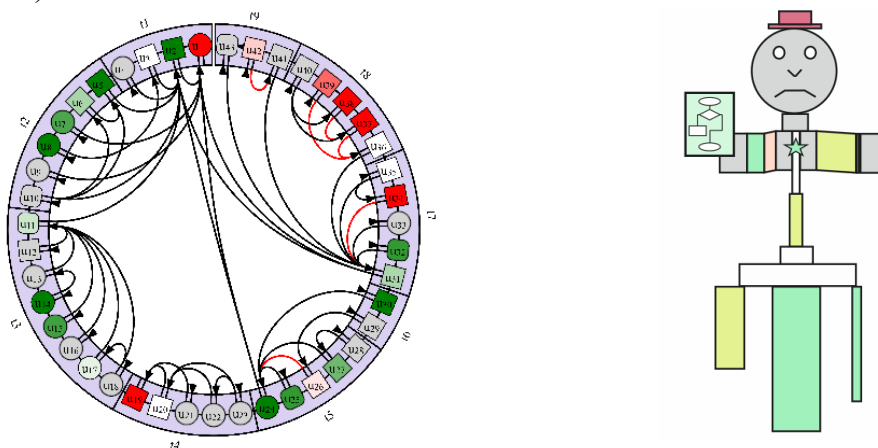


Рис. 2. Примеры ККДЗ (слева) и антропоморфного образа в нотации UGVA (справа)

Функциональные карты реализованы с помощью метода *унифицированного графического воплощения активности* (UGVA) [29]. Они базируются на пиктографии из Data Mining, где метод лиц Чернова [30] был развит до антропоморфного образа (работы В.А. Филимонова [31]), а затем адаптирован к отображению активных объектов [32]. За основу представления данных в этом методе взяты части условной человеческой фигуры, которые по своим размерам и цвету выбираются в соответствии с отдельными показателями из параметрической карты. Такой образ ориентирован для рефлексии своих достижений со стороны учащегося и, как правило, формируется относительно текущей учебной ситуации по всем изучаемым на текущий момент времени дисциплинам учебного плана (пример приведен на рис. 2, справа).

Методические решения. Объединить описанные выше механизмы в рамках одной интеллектуальной системы можно только на базе целостного методологического подхода. Поэтому кратко опишем те методические решения, которые позволяют выстроить логику принятия решений в ИАОС AESU.

Базовым методом для принятия решений планировщиком ИАОС является *модель четырехэтапной концентрации* для проверки гипотез. Суть её заключается в том, что в предельном варианте цепочка обобщения исходных данных для принятия решений включает четыре этапа:

1. Статистическая концентрация позволяет реализовать в виде самостоятельных функций три типа операций над источником данных: сгруппировать данные из выбранных источников в нужной последовательности (сформировать вектор); выполнить статистическое

- обобщение для значений одного из источников (например, поиск минимальных значений, максимальных, средних, приращений и пр.); дать оценку значения в качественной форме (например, используя метод нечёткой логики [23]).
2. Метрическая концентрация, позволяющая свести вектор предобработанных данных в единую структуру (параметрическую карту *Map*), дать оценку отдельным структурным элементам карты в количественном или качественном форматах и сформировать интерактивное графическое изображение в структурном (ККДЗ) или функциональном (UGVA) аспектах.
 3. Семантическая концентрация позволяет на ранее полученных картах выделить закономерности характерных подструктур и контрастировать их относительно специфики решаемой задачи и одной из альтернативных точек зрения (мнений моделей) на учебную ситуацию.
 4. Логическая концентрация комплексно анализирует различные точки зрения в доступных для рассмотрения аспектах и позволит получить компромиссное решение для сложившейся учебной ситуации (используется единый граф принятия решений или их множество для итерационного уточнения по модели *black board*).

Цепочки принятия решений формируются и диспетчеризируются в составе модели *M*. В процессе выработки решений планировщиком таких цепочек обобщения может быть произвольное количество, их длина может варьироваться (включать только часть цепочки). Для реализации такой функциональности был адаптирован метод *модульной концентрации данных* [33], предложенной специально для обработки нетипичных логических цепочек рассуждений со стороны интеллектуального решателя (здесь – планировщика ИАОС). Его суть заключается в том, что диспетчеризирующая модель *M* формирует трек рассуждений в виде взаимосвязанных цепочек методом, аналогичным составлению графических моделей в модуле Simulink пакета Matlab. Моделирующее ядро позволяет подключать из репозитория статистические и аналитические функции, а также графы рассуждений в виде *flm*-файлов, которые представляются в виде графических блоков, требующих объединения в последовательности, отвечающей потребностям задачи принятия решений. Подробное рассмотрение данного метода, реализованных механизмов конструирования моделей и их просчёта выходит за рамки данной статьи и будет представлено в виде самостоятельной работы.

Вся совокупность указанных методов, реализующих системный подход в принятии решений для задач электронного обучения, получила название *сквозного подхода* к анализу учебной ситуации [34]. Его спецификой является не только опора на гибкую логику принятия решений с помощью управляющей модели *M*, но и смещение акцента на обработку запросов, поступивших по инициативе учащегося. Центральным механизмом для этого выбрано взаимодействие учащегося с интерактивными картами (см. пример диалогового окна на левом изображении рис. 3), позволяющими человеку не только самостоятельно обследовать учебную ситуацию, но и в любой момент вызвать диалог методического типа и прояснить причины принятых планировщиком решений, а именно этот эффект и является определяющим в успешности применения ИОС. С результатами наших экспериментов по применению ИАОС AESU, демонстрирующих использование в учебном процессе сквозного подхода на базе карт, можно ознакомиться в работе [35].

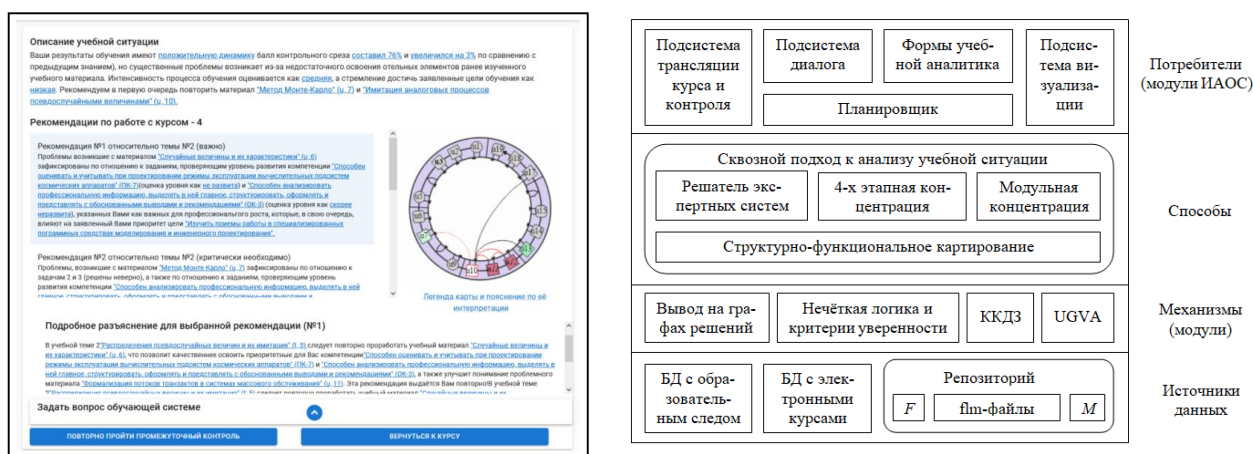


Рис. 3. Пример диалоговой формы с интегрированной интерактивной ККДЗ (слева) и архитектурная схема отвечающих за решения компонентов ИАОС (справа)

Завершая краткий обзор подходов, механизмов и методов, на базе которых организована ИАОС AESU, следует сказать несколько слов о нейронных сетях. Они не применялись в качестве ведущего механизма для данной системы: это в первую очередь связано с тем, что обучающая система должна не только принимать решения, но и объяснять их в диалоге с учащимся. Концепция ХАИ [36] в этой области приложения не показывает высоких результатов. Но мы всё же экспериментируем с возможностью использования свёрточных нейронных сетей для анализа содержимого карт (имитация визуальной интерпретации – поиск шаблонов, выделения закономерностей и аномалий).

AESU как LMS (Learning Management System) и как ИАОС. Описание ИАОС AESU будет недостаточным, если мы не затронем вопросы её функциональности на уровне архитектуры. Констатируем основные функции, которые позволяют нам причислить экспериментальную ИАОС к классу LMS (Learning Management System – система управления обучением):

- выполнение базовых функций обучающей системы (трансляция учебного материала, проведение контрольно-измерительных мероприятий);
- развита система обратной связи в виде диалога на события в учебной среде;
- наличие инструментария по компоновке дидактического материала в электронный курс, его настройке и дополнению его контрольно-измерительными материалами;
- фиксация цифрового образовательного следа в базе данных;
- наличие личных кабинетов учащихся и преподавателей.

Эти, и ряд других возможностей, позволяют реализовать типовую функциональность LMS [37] при развертывании ИАОС AESU в рамках образовательной организации без углубления в вопросы работы интеллектуального решателя. Возможности современных ИАОС [38], опираясь на описанные выше концепции, механизмы и методические решения, в предлагаемой экспериментальной обучающей системе будут следующие:

- формирование индивидуального состава дидактического материала курса и индивидуальной образовательной траектории;
- осуществление гибкого методического диалога в естественно-языковой форме как для события в обучающей среде, по инициативе учащегося, дополненного интерактивными средствами когнитивной визуализации (картами);
- адаптивный контроль и специальные механизмы проверки гипотез об уровне развития предметных и надпредметных компетентностей;
- использование механизма фаззификации и дефаззификации при необходимости работы с качественными значениями различных параметров учебной ситуации;

- имитация процесса рефлексии и поиска обоснованных компромиссов (каждое решение может быть поставлено под сомнение и затребованы аргументы);
- использование механизмов экспертных систем (на эвристиках и с доской объявлений) при обработке всех основных событий в системе;
- гибкий механизм проверки гипотез интеллектуальным планировщиком ИАОС.

Если встраивать специализированные функции через элементы репозитория F , сопровождая каждую моделью M , то общий состав компонентов ИАОС, отвечающих за принятие решений, можно представить в виде архитектурной схемы [39] с рис. 3 (справа). Как уже было показано выше, благодаря методу модульной концентрации данных, обновление логики работы отдельных функций из F в экспериментальной ИАОС будет обладать всеми преимуществами модульного подхода [40], изменяя стратегию проверки гипотез вплоть до модификации отдельных фрагментов баз знаний [41].

Особо следует отметить в архитектуре ИАОС AESU логику реализации стратегии экспертного принятия решений на базе механизма доски объявлений (модель black board). Она базируется на том, что модели ученика, учителя, методиста и тьютора имеют собственные фрагменты баз знаний (источники знаний или knowledge source по [24]). В качестве общей модели предметной области выступает параметрическая карта, визуальная основа для представления которой относительно всех моделей едина – это подложка к ККДЗ (базовая карта по [26]). На неё накладываются индивидуализированные «взгляды» моделей в разных аспектах, а затем, используя механизм итерационного согласования, вырабатывается компромиссное решение, отвечающее задаче анализа текущей учебной ситуации.

Характерной чертой ИАОС является то, что учащийся не только активно взаимодействует с обучающей средой, но и критически относится к её решениям [42]. Для снижения уровня недоверия к этим решениям используются три механизма. Первые два мы уже упомянули: первый – это картирование учебной ситуации и её обследование в интерактивном режиме, второй – методический диалог, направляемый учащимся. Третий механизм – это согласование решения с учащимся-человеком. После решения сложной задачи (например, синтеза индивидуализированной сборки электронного курса и индивидуальной образовательной траектории его изучения) учащемуся предъявляется решение, сопровождаемое той же ККДЗ, с возможностью «обсудить» его. При этом любой элемент решения может получить согласование, запрос на разъяснение, предложение альтернативы или запрос на исключение. Обработка этого взаимодействия также запускает механизм принятия решений с доской объявлений, где картирование играет одну из ключевых ролей. Такой подход позволяет вовлечь учащегося (инициируя рефлексивную деятельность [43]) в сам процесс принятия решений, повышая уровень принятия дальнейших педагогических воздействий.

Заключение. Системный подход к процессу обеспечения принятия объяснимых решений в ИАОС AESU позволил не только задействовать потенциал лежащих в её основе концепций и механизмов, но и за счёт их сочетания получить эмерджентный эффект в идее формирования нового подхода к организации работы обучающей системы. Если пользоваться классификацией уровней функциональных возможностей интеллектуальных систем В.К. Финна, то реализуемый в ИАОС AESU подход обеспечивает уровни $D1$ и $D2$, а за счёт механизма [34] обладает элементами $D3$ [15]. Исследования в этой области нами продолжаются, обучающая система AESU постепенно наращивает свои функциональные возможности.

Механизмы и подходы, описанные выше, предполагают не только комплексный подход к организации электронного обучения на базе интеллектуальных автоматизированных обучающих систем, но и специфическую архитектуру построения такой системы. Отмеченные в

данной статье архитектурные решения только в первом приближении описывают организацию нашей экспериментальной ИАОС AESU. Подробное описание архитектуры и логики работы планировщика в режиме доски объявлений будет представлено в виде самостоятельной публикации.

Список источников

1. Искусственный интеллект в образовании: Социальная практика современной России: Доклад по итогам стратегической сессии Института наследия и современного общества РГГУ. – М.: РГГУ, 2026. – 171 с. – URL: https://iniso.rgu.ru/publications/doklad_ii_v_obrazovanii/.
2. Giannakos M. et al. The promise and challenges of generative AI in education. *Behaviour & Information Technology*, 2024, pp. 1–27, DOI: 10.1080/0144929X.2024.2394886.
3. Денисов М.Е. Онлайн-тренажер "Как это работает: Алгоритмы" / М.Е. Денисов, О.А. Сычёв, Г.В. Терехов // Информационные технологии, 2023. – № 29(5). – С. 267–276. – DOI: 10.17587/it.29.267-276.
4. Карпенко А.П. Модельное обеспечение автоматизированных обучающих систем. Обзор / А.П. Карпенко, А.А. Добряков // Машиностроение и компьютерные технологии, 2011. – № 7. – С. 1-63.
5. Углев В.А. Автоматизированные обучающие системы нового поколения: основные черты и принципы организации / В.А. Углев // Перспективные методы и средства интеллектуальных систем: материалы Всероссийского научно-практического семинара и школы молодых ученых. – Новосибирск: Изд-во НГТУ, 2015. – С. 37–40.
6. Углев В.А. Представление учебной ситуации в интеллектуальных автоматизированных обучающих системах с позиций прикладной семиотики / В.А. Углев // Робототехника и искусственный интеллект: материалы XIV Всероссийской научно-технической конференции с международным участием. – Красноярск: ЛИТЕРА-принт, 2022. – С. 76-83.
7. Рыбина Г.В. Интеллектуальные обучающие системы на основе интегрированных экспертных систем: учеб. Пособие / Г.В. Рыбина. – М.: Директ-Медиа, 2023. – 132 с.
8. Troussas C., Krouska A., Sgouropoulou C. Enhancing human-computer interaction in digital repositories through a mcda-based recommender system. *Advances in Human-Computer Interaction*, 2021, vol. 2021, art. 7213246, DOI: 10.1155/2021/7213246.
9. Анохин П.К. Узловые вопросы теории функциональных систем / П.К. Анохин. – М.: Наука, 1980. – 200 с.
10. Анохин П.К. Проблема принятия решений в психологии и физиологии / П.К. Анохин // Проблемы принятия решений: сборник статей. – М.: Наука, 1974. – С. 7-15.
11. Углев В.А. Модель учебной ситуации для интеллектуального планировщика обучающей системы / В.А. Углев // XXIII национальная конференция по искусственному интеллекту с международным участием (КИИ-2025). – СПб., 2025. – С. 91-102. – DOI: 10.15622/rcai.2025.008.
12. Осипов Г.С. От ситуационного управления к прикладной семиотике / Г.С. Осипов // Новости искусственного интеллекта, 2002. – № 6. – С. 3-7.
13. Поспелов Д.А. Прикладная семиотика / Д.А. Поспелов, Г.С. Осипов // Новости искусственного интеллекта, 1999. – № 1. – С. 9–35.
14. Углев В.А. Представление учебной ситуации как знака и подходы к её визуализации с помощью когнитивной графики / В.А. Углев // Поспеловские чтения: искусственный интеллект - проблемы и перспективы: материалы Всероссийской конференции. – М.: Изд-во ФИЦ ИУ РАН, 2002. – С. 81-89.
15. Финн В.К. Искусственный интеллект: Методология. Применения. Философия / В.К. Финн. – М.: ЛЕНАНД, 2023. – 468 с.
16. Лефевр В.А. Лекции по теории рефлексивных игр / В.А. Лефевр. – М.: Когито-Центр, 2009. – 218 с.
17. Углев В.А. Базовая модель процессов рефлексии в интеллектуальных автоматизированных обучающих системах / В.А. Углев // Математические структуры и моделирование, 2018. – № 1. – С. 111–121.
18. Зенкин А.А. Когнитивная компьютерная графика / А.А. Зенкин. – М.: Наука, 1991. – 192 с.
19. Поспелов Д.А. Десять "горячих точек" в исследованиях по искусственному интеллекту / Д.А. Поспелов // Искусственный интеллект и принятие решений, 2019. – № 4. – С. 3–9.
20. Han J., Kamber M. *Data mining: concepts and techniques*. 2nd ed. University of Illinois at Urbana Champaign, Morgan Kaufmann, 2006, 560 p.
21. Ilves K., Leinonen J., Hellas A. Supporting self-regulated learning with visualizations in online learning environments. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 2018, pp. 257–262, DOI: 10.1145/3159450.3159509.
22. Болсуновский Н.А. Конструктор продукционных экспертных систем с элементами нечёткой логики FLM_Builder и интеграция его моделей в пользовательские проекты / Н.А. Болсуновский, А.Д. Пронин,

- В.А. Углев // Нейроинформатика, ее приложения и анализ данных: XXX Всероссийский семинар. – Красноярск: ИБМ СО РАН, 2022. – С. 24-33.
23. Zadeh L.A., Aliev R.A. Fuzzy logic theory and applications: part I and part II. New Jersey, USA, World Scientific Publishing, 2018.
24. Джексон П. Введение в экспертные системы / П. Джексон. – М.: Вильямс, 2001. – 624 с.
25. Uglev V., Shangina E. Assessment and visualization of course-level and curriculum-level competency profiles. Computational Science and Its Applications – ICCSA 2023 Workshops, LNCS, vol. 14104, Springer, Cham, 2023, pp. 478–493, DOI: 10.1007/978-3-031-37105-9_32.
26. Uglev V.A. Cognitive Maps of Knowledge Diagnosis (CMKD): the essence of the method, classification, characteristics and synthesis principles. Novel & Intelligent Digital Systems: Proceedings of the 4th International Conference (NiDS 2024), NiDS 2024, Lecture Notes in Networks and Systems, vol. 1170, Springer, Cham, 2024, pp. 594–605, DOI: 10.1007/978-3-031-73344-4_51.
27. Milgram S. et al. The small world problem. Psychology Today, 1967, vol. 2, no. 1, pp. 60–67.
28. Углев В.А. Визуализация динамики учебного процесса с помощью когнитивных карт диагностики знаний в интеллектуальных обучающих системах / В.А. Углев, Г.А. Смирнов // Информационные и математические технологии в науке и управлении, 2025. – № 1 (37). – С. 122-129. – DOI: 10.25729/ESI.2025.37.1.012.
29. Углев В.А. Поддержка процесса принятия решений с использованием нотации унифицированного графического воплощения активности (UGVA) / В.А. Углев // Вестник МГТУ им. Н.Э. Баумана. Сер. Приборостроение, 2023. – № 3(144). – С. 125-140. – DOI: 10.18698/0236-3933-2023-3-125-140.
30. Chernoff H. The use of faces to represent points in k-dimensional space graphically. Journal of the American Statistical Association, 1973, vol. 68, pp. 361-368.
31. Филимонов В.А. Способ когнитивной визуализации многопараметрических компонентов системы / В.А. Филимонов // Робототехника и искусственный интеллект: материалы XIII Всероссийской конференции с международным участием. – Красноярск: Литера-принт, 2021. – С. 136-140. – URL: https://aesu.ru/local/conference/_docs/2021/RAI-21_print.pdf.
32. Uglev V., Meshkov S., Kuznetsov M. Methodology of complex visual representation of human activity using cognitive visualization. Journal of Integrated Design and Process Science, 2024, vol. 27, no. 3-4, pp. 184-199, DOI: 10.1177/10920617241289768.
33. Uglev V. An approach to organizing intelligent tutoring systems with customizable decision-making logic. Intelligent Tutoring Systems, LNCS, vol. 15723, Springer, Cham, 2025, pp. 47-58, DOI: 10.1007/978-3-031-98281-1_4.
34. Uglev V., Gavrilova T. Cross-cutting visual support of decision making for forming personalized learning space. Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference (NiDS 2022), NiDS 2022, Lecture Notes in Networks and Systems, vol. 556, Springer, Cham, 2022, pp. 3-13, DOI: 10.1007/978-3-031-17601-2_1.
35. Uglev V., Smirnov G. A cross-cutting approach to analysis of the learning situation in ITS using a mapping mechanism. Journal of Integrated Design and Process Science, 2024, vol. 27, no. 3-4, pp. 288-303, DOI: 10.1177/10920617241289777.
36. Khosravi H., Shum S.B. et al. Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence, 2022, vol. 3, 100074, DOI: 10.1016/j.caeai.2022.100074.
37. Левин С.М. Не все LMS одинаково полезны (для обучения) / С.М. Левин // Национальная ассоциация ученых, 2021. – № 66-3. – С. 22-26.
38. Du Boulay B., Mitrovic A., Yacef K. Handbook of artificial intelligence in education. Edward Elgar Publishing, 2023, DOI: 10.4337/9781800375413.
39. Kossiakoff A. et al. Systems engineering principles and practice. vol. 83. John Wiley & Sons, 2011, 560 p.
40. Васильев А.Л. Модульный принцип формирования техники / А.Л. Васильев. – М.: Изд-во стандартов, 1989. – 238 с.
41. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2000. – 382 с.
42. Baker R.S. Stupid tutoring systems, intelligent humans. International Journal of Artificial Intelligence in Education, 2016, vol. 26, no. 2, pp. 600-614, DOI: 10.1007/s40593-016-0105-0.
43. Щенников С.А. Дидактика электронного обучения / С.А. Щенников // Высшее образование в России, 2010. – № 12. – С. 83-90.

Углев Виктор Александрович. Доцент, канд. техн. наук ФГАОУ ВО «Сибирский федеральный университет», Московский университет МВД России им. В.Я. Кикотя, доцент. Член Российской ассоциации искусственного интеллекта. Основные направления исследований: интеллектуальные автоматизированные обучающие

системы, инженерия знаний, системная инженерия, когнитивная компьютерная графика, имитационное моделирование, методология науки. AuthorID: 559875, SPIN: 7368-2910, ORCID: 0000-0003-1205-2652, uglev-v@yandex.ru. 662971 Красноярский край, г. Железногорск, ул. Кирова, д. 12а, ауд. 210.

Смирнов Георгий Артёмович. Аспирант ФГАОУ ВО «Сибирский федеральный университет». Основные направления исследований: искусственный интеллект, автоматизированные обучающие системы, инженерия знаний, системная инженерия. AuthorID: 1245841, SPIN: 9183-5051, ORCID: 0009-0003-7402-4419, Determaer2012@gmail.com. 662971 Красноярский край, г. Железногорск, ул. Кирова, д. 12а, ауд. 210.

UDC 004.89:004.832.3:378

DOI:10.25729/ESI.2026.43.3.004

AESU Intelligent Tutoring Systems and its underlying concepts, mechanisms and methods

Viktor A. Uglev^{1,2}, Georgy A. Smirnov¹

¹Siberian Federal University, Russia, Zheleznogorsk, vauglev@sfu-kras.ru

²Moscow University of the Ministry of Internal Affairs of Russia, Russia, Moscow

Abstract. The problem of a systematic and scientifically-based approach to the organization of e-learning is becoming increasingly relevant in the context of the widespread use of large language models. This article explores the issue of conceptual and methodological support for the operation of Intelligent Tutoring Systems (ITS). It highlights the challenge of describing and explaining the internal mechanisms of these systems in scientific publications, allowing for the verification and automatic explanation of their decisions. As an example, the author's experimental learning system, AESU, is examined. The article focuses on the fundamental concepts, mechanisms, and methodological solutions employed in this system. The main concepts include P.K. Anokhin's theory of functional systems (the model of afferent synthesis), D.A. Pospelov's applied semiotics and V.A. Lefebvre's theory of reflexive control. The role of cognitive visualization (mapping) in the synthesis of solutions is demonstrated. This involves combining the logic of production expert systems, fuzzy logic, and Shortliffe's hypothesis testing, which are the intellectual mechanisms of the learning system's solver. The emergent effect arising from the use of the above-mentioned concepts and mechanisms allows us to propose an author's approach to the implementation of modular data concentration and cross-cutting approach of the learning situation. The structural and functional notation of maps becomes the system-forming element in the synthesis of ITS solutions: the method of Cognitive Maps of Knowledge Diagnosis and the method of Unified Graphic Visualization of Activity. The article demonstrates the possibility of mutual transition between the parametric and visual representation of the learning situation between these types of maps and their use in the development of solutions by both the solver of the learning system and the student during their work in the electronic learning environment. In conclusion, the architectural features of the AESU ITS are noted. In addition to the typical functions of learning management systems, the proposed ITS implements a dialogic mechanism for interacting with students in a natural language form without using the capabilities of large language models. The flexible feedback mechanism allows for interactive dialogue based on methodologically sound conclusions, which increases the level of trust between the student and the learning system and promotes engagement in the educational process based on the ITS.

Keywords: electronic learning, Intelligent Tutoring Systems, cognitive visualization, decision-making, method

References

1. Iskusstvennyj intellekt v obrazovanii: Social'naya praktika sovremennoj Rossii: Doklad po itogam strategicheskoy sessii Instituta naslediya i sovremennogo obshchestva RGGU [Artificial Intelligence in Education: Social Practice in Contemporary Russia: Report on the Strategic Session of the Institute of Heritage and Contemporary Society]. Moscow, RSGI, 2026, 171 p, available at: https://iniso.rgg.ru/publications/doklad_ii_v_obrazovanii/.
2. Giannakos M. et al. The promise and challenges of generative AI in education. Behaviour & Information Technology, 2024, pp. 1–27, DOI: 10.1080/0144929X.2024.2394886.

3. Denisov M.E., Sychev O.A., Terekhov G.V. Onlayn-trenazher "Kak eto rabotayet: Algoritmy" [Online simulator "How it works: Algorithms"]. *Informatsionnyye tekhnologii* [Information Technology], 2023, vol. 29, no. 5, pp. 267–276, DOI: 10.17587/it.29.267-276.
4. Karpenko A.P., Dobryakov A.A. Model'noye obespecheniye avtomatizirovannykh obuchayushchikh sistem. *Obzor* [Model support of automated learning systems. Review]. *Mashinostroyeniye i kompyuternyye tekhnologii* [Mechanical Engineering and Computer Technologies], 2011, no. 7, pp. 1-63.
5. Uglev V.A. Avtomatizirovannyye obuchayushchiye sistemy novogo pokoleniya: osnovnyye cherty i printsipy organizatsii [New generation of Intelligent Tutoring Systems: main features and principles of organization]. *Perspektivnyye metody i sredstva intellektual'nykh sistem: materialy Vserossiyskogo nauchno-prakticheskogo seminar i shkoly molodykh uchenykh* [Advanced Methods and Tools of Intelligent Systems: Proceedings of the All-Russian Scientific and Practical Seminar and School of Young Scientists]. Novosibirsk, NSTU Publ., 2015, pp. 37–40.
6. Uglev V.A. Predstavleniye uchebnoy situatsii v intellektual'nykh avtomatizirovannykh obuchayushchikh sistemakh s pozitsiy prikladnoy semiotiki [Representation of the learning situation in Intelligent Tutoring Systems from the standpoint of applied semiotics]. *Robototekhnika i iskusstvennyy intellekt: materialy XIV Vserossiyskoy nauchno-tekhnicheskoy konferentsii s mezhdunarodnym uchastiyem* [Robotics and Artificial Intelligence: Proceedings of the XIV All-Russian Scientific and Technical Conference with International Participation]. Krasnoyarsk, LITERA-print Publ., 2022, pp. 76-83.
7. Rybina G.V. Intellektual'nyye obuchayushchiye sistemy na osnove integrirovannykh ekspertnykh sistem: ucheb. posobiye [Intelligent learning systems based on integrated expert systems: textbook]. Moscow, Direkt-Media Publ., 2023, 132 p.
8. Troussas C., Krouska A., Sgouropoulou C. Enhancing human-computer interaction in digital repositories through a mcda-based recommender system. *Advances in Human-Computer Interaction*, 2021, vol. 2021, art. 7213246, DOI: 10.1155/2021/7213246.
9. Anokhin P.K. Uzlovye voprosy teorii funktsional'nykh sistem [Key issues of the theory of functional systems]. Moscow, Nauka [Science] Publ., 1980, 200 p.
10. Anokhin P.K. Problema prinyatiya resheniy v psikhologii i fiziologii [The problem of decision making in psychology and physiology]. *Problemy prinyatiya resheniy: sbornik statey* [Problems of Decision Making: Collection of Articles]. Moscow, Nauka [Science] Publ., 1974, pp. 7-15.
11. Uglev V.A. Model' uchebnoy situatsii dlya intellektual'nogo planirovshchika obuchayushchey sistemy [A learning situation model for an intelligent tutoring systems planner]. *XXIII natsional'naya konferentsiya po iskusstvennomu intellektu s mezhdunarodnym uchastiyem (KII-2025)* [XXIII National Conference on Artificial Intelligence with International Participation (RCAI-2025)]. Saint Petersburg, 2025, pp. 91-102, DOI: 10.15622/rcai.2025.008.
12. Osipov G.S. Ot situatsionnogo upravleniya k prikladnoy semiotike [From situational control to applied semiotics]. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News], 2002, no. 6, pp. 3-7.
13. Pospelov D.A., Osipov G.S. Prikladnaya semiotika [Applied semiotics]. *Novosti iskusstvennogo intellekta* [Artificial Intelligence News], 1999, no. 1, pp. 9–35.
14. Uglev V.A. Predstavleniye uchebnoy situatsii kak znaka i podkhody k yeye vizualizatsii s pomoshch'yu kognitivnoy grafiki [Representation of the learning situation as a sign and approaches to its visualization using cognitive graphics]. *Pospelovskiye chteniya: iskusstvennyy intellekt - problemy i perspektivy: materialy Vserossiyskoy konferentsii* [Pospelov Readings: Artificial Intelligence - Problems and Prospects: Proceedings of the All-Russian Conference]. Moscow, FRC IU RAS Publ., 2002, pp. 81-89.
15. Finn V.K. *Iskusstvennyy intellekt: Metodologiya. Primeneniya. Filosofiya* [Artificial intelligence: Methodology. Applications. Philosophy]. Moscow, LENAND Publ., 2023, 468 p.
16. Lefevre V.A. *Leksii po teorii refleksivnykh igr* [Lectures on the theory of reflexive games]. Moscow, Kogito-Tsentr [Cogito-Center] Publ., 2009, 218 p.
17. Uglev V.A. Bazovaya model' protsessov refleksii v intellektual'nykh avtomatizirovannykh obuchayushchikh sistemakh [The base model of reflexive process in the intellectual automation educational systems]. *Matematicheskiye struktury i modelirovaniye* [Mathematical Structures and Modeling], 2018, no. 1, pp. 111–121.
18. Zenkin A.A. *Kognitivnaya komp'yuternaya grafika* [Cognitive computer graphics]. Moscow, Nauka [Science] Publ., 1991, 192 p.
19. Pospelov D.A. Desyat "goryachikh tochek" v issledovaniyakh po iskusstvennomu intellektu [Ten "hot spots" in artificial intelligence research]. *Iskusstvennyy intellekt i prinyatiye resheniy* [Artificial Intelligence and Decision Making], 2019, no. 4, pp. 3–9.
20. Han J., Kamber M. *Data mining: concepts and techniques*. 2nd ed. University of Illinois at Urbana Champaign, Morgan Kaufmann, 2006, 560 p.

21. Ilves K., Leinonen J., Hellas A. Supporting self-regulated learning with visualizations in online learning environments. *Proceedings of the 49th ACM Technical Symposium on Computer Science Education*, 2018, pp. 257–262, DOI: 10.1145/3159450.3159509.
22. Bolsunovsky N.A., Pronin A.D., Uglev V.A. Konstruktor produktsionnykh ekspertnykh sistem s elementami nechetkoy logiki FLM_Builder i integratsiya yego modeley v pol'zovatel'skiye proyekty [Constructor of production expert systems with fuzzy logic elements FLM_Builder and integration of its models into user projects]. *Neyroinformatika, yeye prilozheniya i analiz dannykh: XXX Vserossiyskiy seminar [Neuroinformatics, Its Applications and Data Analysis: XXX All-Russian Seminar]*. Krasnoyarsk, ICM SB RAS Publ., 2022, pp. 24–33.
23. Zadeh L.A., Aliev R.A. *Fuzzy logic theory and applications: part I and part II*. New Jersey, USA, World Scientific Publishing, 2018.
24. Jackson P. *Vvedeniye v ekspertnyye sistemy [Introduction to expert systems]*. Moscow, Williams Publ., 2001, 624 p.
25. Uglev V., Shangina E. Assessment and visualization of course-level and curriculum-level competency profiles. *Computational Science and Its Applications – ICCSA 2023 Workshops, LNCS*, vol. 14104, Springer, Cham, 2023, pp. 478–493, DOI: 10.1007/978-3-031-37105-9_32.
26. Uglev V.A. Cognitive Maps of Knowledge Diagnosis (CMKD): the essence of the method, classification, characteristics and synthesis principles. *Novel & Intelligent Digital Systems: Proceedings of the 4th International Conference (NiDS 2024), NiDS 2024, Lecture Notes in Networks and Systems*, vol. 1170, Springer, Cham, 2024, pp. 594–605, DOI: 10.1007/978-3-031-73344-4_51.
27. Milgram S. et al. The small world problem. *Psychology Today*, 1967, vol. 2, no. 1, pp. 60–67.
28. Uglev V.A., Smirnov G.A. Vizualizatsiya dinamiki uchebnogo protsessa s pomoshch'yu kognitivnykh kart diagnostiki znaniy v intellektual'nykh obuchayushchikh sistemakh [Visualization of the learning process dynamics using Cognitive Maps of Knowledge Diagnosis in Intelligent Tutoring Systems]. *Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii [Information and Mathematical Technologies in Science and Management]*, 2025, no. 1 (37), pp. 122–129, DOI: 10.25729/ESI.2025.37.1.012.
29. Uglev V.A. Podderzhka protsessa prinyatiye resheniy s ispol'zovaniyem notatsii unifikirovannogo graficheskogo voploshcheniya aktivnosti (UGVA) [Support of decision-making by Unified Graphic Visualization of Activity notation (UGVA)]. *Vestnik MGTU im. N.E. Bauman. Ser. Priborostroyeniye [Bulletin of Bauman MSTU. Ser. Instrument Engineering]*, 2023, no. 3(144), pp. 125–140, DOI: 10.18698/0236-3933-2023-3-125-140.
30. Chernoff H. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 1973, vol. 68, pp. 361–368.
31. Filimonov V.A. Sposob kognitivnoy vizualizatsii mnogoparametricheskikh komponentov sistemy [Method of cognitive visualization of multi-parametric system components]. *Robototekhnika i iskusstvennyy intellekt: materialy XIII Vserossiyskoy konferentsii s mezhdunarodnym uchastiyem [Robotics and Artificial Intelligence: Proceedings of the XIII All-Russian Conference with International Participation]*. Krasnoyarsk, Litera-print Publ., 2021, pp. 136–140. Available at: https://aesu.ru/local/conference/_docs/2021/RAI-21_print.pdf.
32. Uglev V., Meshkov S., Kuznetsov M. Methodology of complex visual representation of human activity using cognitive visualization. *Journal of Integrated Design and Process Science*, 2024, vol. 27, no. 3–4, pp. 184–199, DOI: 10.1177/10920617241289768.
33. Uglev V. An approach to organizing intelligent tutoring systems with customizable decision-making logic. *Intelligent Tutoring Systems, LNCS*, vol. 15723, Springer, Cham, 2025, pp. 47–58, DOI: 10.1007/978-3-031-98281-1_4.
34. Uglev V., Gavrilova T. Cross-cutting visual support of decision making for forming personalized learning space. *Novel & Intelligent Digital Systems: Proceedings of the 2nd International Conference (NiDS 2022), NiDS 2022, Lecture Notes in Networks and Systems*, vol. 556, Springer, Cham, 2022, pp. 3–13, DOI: 10.1007/978-3-031-17601-2_1.
35. Uglev V., Smirnov G. A cross-cutting approach to analysis of the learning situation in ITS using a mapping mechanism. *Journal of Integrated Design and Process Science*, 2024, vol. 27, no. 3–4, pp. 288–303, DOI: 10.1177/10920617241289777.
36. Khosravi H., Shum S.B. et al. Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, 2022, vol. 3, 100074, DOI: 10.1016/j.caeai.2022.100074.
37. Levin S.M. Ne vse LMS odinakovo polezny (dlya obucheniya) [Not all LMS are equally useful (for learning)]. *Natsional'naya assotsiatsiya uchenykh [National Association of Scientists]*, 2021, no. 66–3, pp. 22–26.
38. Du Boulay B., Mitrovic A., Yacef K. *Handbook of artificial intelligence in education*. Edward Elgar Publishing, 2023, DOI: 10.4337/9781800375413.
39. Kossiakoff A. et al. *Systems engineering principles and practice*. vol. 83. John Wiley & Sons, 2011, 560 p.

40. Vasiliev A.L. Modul'nyy printsip formirovaniya tekhniki [The modular principle of technology formation]. Moscow, Standards Publishing House, 1989, 238 p.
41. Gavrilova T.A., Khoroshevsky V.F. Bazy znaniy intellektual'nykh sistem [Knowledge bases of intelligent systems]. Saint Petersburg, Piter Publ., 2000, 382 p.
42. Baker R.S. Stupid tutoring systems, intelligent humans. International Journal of Artificial Intelligence in Education, 2016, vol. 26, no. 2, pp. 600-614, DOI: 10.1007/s40593-016-0105-0.
43. Shchennikov S.A. Didaktika elektronnoy obucheniya [Didactics of e-learning]. Vyssheye obrazovaniye v Rossii [Higher Education in Russia], 2010, no. 12, pp. 83-90.

Uglov Viktor Alexandrovich. Associate Professor, Candidate of Technical Sciences, Siberian Federal University, Moscow University of the Ministry of Internal Affairs of Russia, Associate Professor. Member of the Russian Association of Artificial Intelligence. Main research areas: Intelligent Tutoring Systems, knowledge engineering, systems engineering, cognitive computer graphics, simulation modeling, methodology of science. AuthorID: 559875, SPIN: 7368-2910, ORCID: 0000-0003-1205-2652. uglov-v@yandex.ru. 662971 Krasnoyarsk territory, Zheleznogorsk, Kirova str., h. 12a, aud. 210.

Smirnov Georgy Artemovich. Postgraduate student of the Siberian Federal University. Main research areas: artificial intelligence, Intelligent Tutoring Systems, knowledge engineering, systems engineering. AuthorID: 1245841, SPIN: 9183-5051, ORCID: 0009-0003-7402-4419. Determaer2012@gmail.com. 662971 Krasnoyarsk territory, Zheleznogorsk, Kirova str., h. 12a, aud. 210.

Статья поступила в редакцию 12.11.2025; одобрена после рецензирования 20.11.2025; принята к публикации 31.07.2026.

The article was submitted 11/12/2025; approved after reviewing 11/20/2025; accepted for publication 07/31/2026.

Математическое моделирование и его применение в научных исследованиях

УДК 517.95;535.3

DOI:10.25729/ESI.2026.43.3.005

Моделирование свойств критического значения показателя преломления при переходах пика интенсивности между нелинейной и градиентной оптическими средами

Савотченко Сергей Евгеньевич^{1,2}, Афанасьева Надежда Олеговна²

¹МИРЭА – Российский технологический университет,
Россия, Москва, savotchenkose@mail.ru

²Российский государственный геологоразведочный университет им. Серго Орджоникидзе,
Россия, Москва

Аннотация. В работе проведено моделирование пространственного перераспределения интенсивности оптического поля в планарной волноводной структуре, образованной границей раздела между нелинейной средой и средой с градиентным показателем преломления. В качестве модели градиентной среды была выбрана обобщенная зависимость диэлектрической проницаемости от расстояния от поверхности контакта сред. В качестве модельной нелинейной среды была выбрана среда с оптическим эффектом Поккельса, в которой диэлектрическая проницаемость линейно зависит от амплитуды напряженности электрического поля. В исследовании выявлены критические режимы работы, в которых максимум интенсивности направляемой моды может быть целенаправленно расположен внутри любой из составляющих сред или точно на границе их раздела путем регулирования оптических параметров системы, определяющих критическое значение эффективного показателя преломления. Основным результатом работы является доказательство существования критического эффективного показателя преломления независимо от выбора модели нелинейности, при котором максимум поля точно локализован на границе раздела. Показано, что значение зависит исключительно от параметров, определяющих профиль градиентного показателя преломления, что обеспечивает дисперсионное соотношение. Особый акцент сделан на выявлении того, как каждый из параметров градиентного профиля в отдельности влияет на величину порога перехода. Анализ выявил, что критическое значение эффективного показателя преломления монотонно изменяется с изменением характерного пространственного масштаба и диэлектрических параметров градиентного слоя. Показано, что характер этого влияния определяется знаком и величиной соответствующих диэлектрических констант, что позволяет целенаправленно формировать необходимую зависимость критического показателя преломления от оптических параметров планарной волноводной структуры. В отличие от большинства работ, ограничивающихся численным моделированием конкретных конфигураций, данное исследование предлагает строгий аналитический подход, позволяющий получить явные критерии существования критического режима. Дополнительно обсуждается возможность использования полученных результатов для калибровки экспериментальных установок, настройки условий фазового синхронизма при генерации второй гармоники, а также для оценки допустимых разбросов параметров при производстве интегральных схем. Результаты создают теоретическую основу разработки методов контролируемой локализации света в гибридных волноводных структурах путем изменения их материальных и геометрических параметров или рабочей длины волны.

Ключевые слова: математическое моделирование, нелинейные уравнения, оптический волновод, нелинейная среда, краевая задача

Цитирование: Савотченко С.Е. Моделирование свойств критического значения показателя преломления при переходах пика интенсивности между нелинейной и градиентной оптическими средами / С.Е. Савотченко, Н.О. Афанасьева // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 56-69. – DOI:10.25729/ESI.2026.43.3.005.

Введение. Проблемы управления пространственной локализацией оптического поля на границах раздела сред с принципиально различными физическими свойствами относятся к важным фундаментальным и прикладным задачам в области фотоники и волноводной оптики [1-3]. Эффективность нелинейных оптических процессов, таких, как генерация второй

гармоники [4], параметрическое усиление и самофокусировка, критически зависят от локализации максимума интенсивности поля относительно функциональных областей структуры [5]. Пространственное выравнивание областей высокой интенсивности с областями сильной нелинейной восприимчивости является необходимым условием для достижения высокой эффективности преобразования энергии [6]. Особую актуальность приобретает задача предсказуемого управления положением пика интенсивности без механического перемещения элементов или изменения состава нелинейного материала [7].

Несмотря на обширную литературу по нелинейной и волноводной оптике [8, 9], остается пробел в понимании условий, при которых гибридная волноводная структура из нелинейной и градиентной сред демонстрирует качественное изменение режима локализации света [10]. Отсутствует развитая систематическая теория, предсказывающая критическое значение параметра (эффективного показателя преломления), при котором происходит переход максимальной интенсивности от одного материала к другому. Предлагаемая работа восполняет этот пробел, выводя универсальное условие, определяемое оптическими характеристиками волноводной структуры.

Основная цель рассматриваемой работы заключается в моделировании и анализе критического значения эффективного показателя преломления, определяющего пространственный переход пика интенсивности между двумя различными оптическими средами (нелинейной средой и средой с градиентным показателем преломления) внутри планарной волноводной структуры. Новизна заключается в демонстрации того, что критическое условие локализации максимума поля на границе раздела определяется исключительно дисперсионными характеристиками градиентной среды, независимо от конкретного механизма нелинейного отклика в соседнем материале. Это отделяет критическое условие перехода от конкретной модели оптической нелинейности (что показано на примерах нелинейных сред с оптическими эффектами Поккельса и Керра [11]), предлагая более общую теоретическую основу по сравнению с анализом, сфокусированном на одном типе нелинейности. Эта универсальность представляет собой значительный концептуальный прогресс по сравнению с традиционными подходами к нелинейным волноводам, которые обычно связывают условия перехода непосредственно с нелинейными коэффициентами.

Применение методов математического моделирования также дает возможность получить явные аналитические выражения (где возможно) и четкие условия сопряжения, что позволяет глубже понять физику процесса, в отличие от чисто численных расчетов. Аналитические результаты и точные решения модельных уравнений служат эталоном для проверки более сложных численных моделей и коммерческих программ для моделирования (например, основанных на методе конечных элементов) [12, 13]. Кроме того, готовые аналитические решения позволяют проводить быстрый параметрический анализ и оптимизацию, что экономит вычислительные ресурсы, а их результаты создают теоретический базис для разработки нового поколения управляемых фотонных устройств, где пространственное положение световой энергии является управляющим параметром [14].

1. Основные уравнения модели и аналитические результаты. Будем рассматривать только поперечные электрические (ТЕ) волны, бегущие вдоль плоской границы раздела (вдоль оси Oz) двух сред с различными оптическими свойствами с компонентой напряженности электрического поля $E_y(x, z) = u(x) \exp(iknz)$, где k – волновое число, связанное с длиной волны в вакууме λ_0 : $k = 2\pi/\lambda_0$, n – эффективный показатель преломления, связанный с углом падения луча, возбуждающего волноводные моды или поверхностные волны. Как известно [15], в этом случае в рамках теории оптических волноводных систем получается скалярное уравнение

$$u''(x) + k^2(\varepsilon - n^2)u(x) = 0, \quad (1)$$

описывающее пространственное распределение электрического поля $u(x)$ в направлении, перпендикулярном границе раздела двух сред (расположенной в плоскости yOz) с различными оптическими свойствами (вдоль оси Ox). В (1) ε – диэлектрическая проницаемость, включающая свойства двух контактирующих сред. Поле затухает с ростом расстояния x от такой границы, причем на больших расстояниях убывание поля носит, как правило, экспоненциальный характер. Вдоль плоской границы раздела поле в виде бегущей (вдоль оси Oz) волны будет распространяться без затухания на большие расстояния в немагнитных средах с пренебрежительно малыми диэлектрическими потерями. Это означает, что поле локализуется вблизи такой границы раздела, поэтому она обладает волноводными свойствами, и может выступать в роли плоского волновода.

Поскольку наиболее перспективными с точки зрения управляемости дисперсионными характеристиками являются волноводные системы, сочетающие слои из кристаллов с нелинейным оптическим откликом, характеризуемым зависимостью диэлектрической проницаемости от амплитуды напряженности электрического поля (или интенсивности света) $\varepsilon = \varepsilon_N(u)$, и слои с неоднородным распределением показателя преломления, приводящего к зависимости диэлектрической проницаемости от расстояния $\varepsilon = \varepsilon_G(x)$ [16], то будем рассматривать волноводную систему, состоящую из разделенных плоской тонкой границей нелинейной среды при $x < 0$ и неоднородной (градиентной) среды при $x > 0$.

В ряде фотонно-кристаллических структур показатель преломления может меняться достаточно быстро с расстоянием от границы раздела [17, 18], поэтому при моделировании таких свойств резкой неоднородности можно использовать для модуляции пространственного профиля диэлектрической проницаемости функцию вида:

$$\varepsilon_G(x) = \varepsilon_{G0} + \frac{\varepsilon_{G1}}{x+h} + \frac{\varepsilon_{G2}}{(x+h)^2}, \quad (2)$$

где h , ε_{Gj} ($j=0, 1, 2$) – константы, называемые параметрами профиля. Так как функция (2) определена при $x > 0$, то для того, чтобы избежать сингулярности в модели, предполагается, что $h > 0$. Остальные параметры профиля ε_{Gj} могут иметь разные знаки, сочетание которых определяют его форму. Более подробный анализ соответствия значений параметров профиля (2) и его формы можно найти в [19], где показано, при каких знаках параметров функция (2) монотонная, имеет максимум или минимум.

Если считать, что искомое решение уравнения (1) $u(x) = u_G(x)$ при $x > 0$ и $u(x) = u_N(x)$ при $x < 0$ (здесь функции $u_G(x)$ и $u_N(x)$ описывают пространственные распределения напряженности поперечного электрического поля в неоднородной и нелинейной частях волновода, соответственно), и использовать представление диэлектрической проницаемости с учетом разделения на нелинейную и неоднородное полупространства и функцию (2), то получаются два уравнения

$$u_G'' + \left(\varepsilon_{G0} + \frac{\varepsilon_{G1}}{x+h} + \frac{\varepsilon_{G2}}{(x+h)^2} - n^2 \right) k^2 u_G = 0, \quad x > 0, \quad (3)$$

$$u_N'' + (\varepsilon_N(u) - n^2) k^2 u_N = 0, \quad x < 0. \quad (4)$$

Эти уравнения дополняются естественными требованиями непрерывности компонент поля на границе раздела сред, то есть решение u и его производные должны быть всюду непрерывными. Также решение u должно быть всюду ограниченным и исчезать на бесконечности. В результате получается краевая задача сопряжения для нахождения гладких, непрерывных и ограниченных решений u_G и u_N уравнений (3) и (4) на соответствующих полуосях, что составляет математическую формулировку модели рассматриваемой волноводной структуры.

Уравнение (3) представляет собой линейное дифференциальное уравнение второго порядка с переменными коэффициентами. Его точное решение выражается через специальные функции. В частности, можно воспользоваться представлением решения через функцию Уиттекера $W_{\mu,\nu}(x)$ [19, 20]:

$$u_G(x) = u_0 \frac{W_{\mu,\nu}(2k\sqrt{n^2 - \varepsilon_{G0}}(x+h))}{W_{\mu,\nu}(2kh\sqrt{n^2 - \varepsilon_{G0}})}, x > 0. \quad (5)$$

где $u_0 = u_G(0)$ – амплитуда поля на границе раздела сред,

$$\mu = \frac{\varepsilon_{G1}}{2\sqrt{n^2 - \varepsilon_{G0}}}, \quad (6)$$

$$\nu = \frac{1}{2}\sqrt{1 - 4k^2\varepsilon_{G2}}. \quad (7)$$

Уравнение (4) представляет собой нелинейное дифференциальное уравнение второго порядка. Его решение зависит от выбора модели нелинейного отклика среды. В частности, в средах с эффектом Погкелса нелинейная диэлектрическая проницаемость $\varepsilon_N(u)$ записывается в виде линейной функции поля [21]:

$$\varepsilon_N(u) = \varepsilon_{N0} + \varepsilon_{N1} \cdot u, \quad (8)$$

где ε_{N0} – невозмущенная диэлектрическая константа нелинейного кристалла, ε_{N1} – коэффициент нелинейности Погкельса. Тогда исчезающее на бесконечности решение нелинейного уравнения (4) с нелинейностью Погкельса (8) можно записать в виде:

$$u_N(x) = \frac{3(n^2 - \varepsilon_{N0})}{2\varepsilon_{N1}\text{ch}^2(k\sqrt{n^2 - \varepsilon_{N0}}(x+x_0)/2)}, x < 0. \quad (9)$$

В средах с эффектом Керра нелинейная диэлектрическая проницаемость $\varepsilon_N(u)$ записывается в виде квадратичной функции поля [21]:

$$\varepsilon_N(u) = \varepsilon_{N0} + \varepsilon_{N2} \cdot u^2, \quad (10)$$

где ε_{N2} – коэффициент нелинейности Керра. Тогда исчезающее на бесконечности решение нелинейного уравнения (4) с нелинейностью Керра (10) можно записать в виде:

$$u_N(x) = \sqrt{\frac{2(n^2 - \varepsilon_{N0})}{\varepsilon_{N2}}} \frac{1}{\text{ch}^2(k\sqrt{n^2 - \varepsilon_{N0}}(x+x_0))}, x < 0. \quad (11)$$

Оба решения (9) и (11) в обеих моделях нелинейностей определяются с точностью до сдвига на постоянную x_0 , которая определяется из условия непрерывности решения и его производной в точке $x=0$ при сшивке с решением на другой полуоси, заданным (5). Следовательно, можно считать, что решения нелинейного уравнения (4), порождаемые широким классом моделей нелинейностей [20], будут зависеть от аргумента со сдвижкой $x+x_0$. Поэтому будем далее использовать общее представление решения нелинейного уравнения (4) в виде:

$$u_N(x) = f(x+x_0). \quad (12)$$

где конкретный вид функции f определяется выбранной моделью нелинейности (считается далее известной). В частности, для модели нелинейности (8) решение задается функцией (9), а для модели нелинейности (10) решение задается функцией (11). Функция f ограниченная, непрерывная, дважды дифференцируемая и такая, что $|f(x)| \rightarrow 0$ при $x \rightarrow -\infty$, а также, не теряя общности, можно считать далее, что $f(x) > 0$ при $x \in (-\infty; 0]$.

Параметр x_0 определяет положение максимума решения (12), то есть максимума интенсивности поля. Этот максимум находится в нелинейной среде, если $x_0 > 0$. При значениях $x_0 < 0$ он перемещается из нелинейной среды в неоднородную, причем при $x_0 = 0$ его положение

максимума оказывается в точности на границе раздела сред. Поэтому такое условие будет определять критические значения параметров системы, при которых максимум интенсивности поля переходит из одной среды в другую. Обычно варьируемым в экспериментах параметром является эффективный показатель преломления, поэтому можно говорить о существовании его критического значения n_c , определяемого оптическими параметрами системы.

Значение величины x_0 находится из условия непрерывности решений (5) и (12) и их производных в точке $x=0$:

$$u_0 = f(x_0), \quad (13)$$

$$q_G u_0 = f'(x_0), \quad (14)$$

где:

$$q_G = 2k\sqrt{n^2 - \varepsilon_{G0}} \frac{W'_{\mu,v}(2kh\sqrt{n^2 - \varepsilon_{G0}})}{W_{\mu,v}(2kh\sqrt{n^2 - \varepsilon_{G0}})}. \quad (15)$$

Из (13) и (14) получается уравнение:

$$q_G = \frac{f'(x_0)}{f(x_0)}. \quad (16)$$

Если ввести функцию $F=f'/f$, то (16) примет вид:

$$q_G = F(x_0). \quad (17)$$

Если считать, что функция F монотонная, взаимнооднозначная и существует обратная $F^{[-1]}$, то из (17) можно найти положение максимума в явном виде:

$$x_0 = F^{[-1]}(q_G). \quad (18)$$

Если максимум поля расположен в точности на границе раздела сред, то из (18) получается, что $F^{[-1]}(q_G)=0$. Если F нечётная и глобально обратимая на симметричной области (область определения и область значений функции симметричны относительно нуля), то $F^{[-1]}$ тоже будет нечётной. Следовательно, в этом случае получается, что из того, что $x_0=0$, вытекает, что $q_G=0$.

Последнее условие следует рассматривать, как уравнение для нахождения критического значения эффективного показателя преломления n_c , при котором максимум интенсивности располагается на границе раздела сред. Важно отметить, что это условие определяется выражением (15) и не зависит от модели нелинейности, а определяется характеристиками неоднородной среды. Такое поведение выявляет универсальную особенность гибридных волноводных систем, где геометрические и дисперсионные свойства градиентной среды определяют порог перехода для локализации поля.

Очевидно, что существует область значений эффективного показателя преломления $n > n_c$, при которой q_G и x_0 , соответственно, будут иметь определенные знаки, а при $n < n_c$ их знак поменяется на противоположный. Это означает, что максимум интенсивности поля будет располагаться в одной среде при $n > n_c$, и в другой при $n < n_c$. Конкретизация места его расположения определяется выбором модели волноводной системы, жестко связывающей q_G и x_0 , зависящих от n .

Существование такого критического значения n_c свидетельствует о детерминированном переходе в механизме локализации энергии внутри гибридного волновода. В точке $n=n_c$ конкурирующие механизмы ограничения из обеих сред уравниваются, в результате чего максимум поля «фиксируется» на физической границе раздела. Это аналогично фазовому переходу в пространственной области оптического поля.

Результаты аналитического моделирования выявляют фундаментальное расхождение между геометрическими волноводными свойствами градиентной среды и механизмом

нелинейного отклика при определении условия перехода для локализации поля. Физически это происходит потому, что критическое условие соответствует сингулярной точке, где логарифмическая производная поля (16) совпадает по обе стороны границы раздела сред. Условие, вытекающее из (16), математически эквивалентно требованию нулевой первой производной поля точно на границе.

В градиентной среде эта производная зависит исключительно от локального градиента показателя преломления и кривизны через решение уравнения в виде функции Уиттекера. В нелинейной среде, хотя форма профиля поля зависит от типа нелинейности, требование стационарной точки на границе накладывает универсальное ограничение, не зависящее от конкретного нелинейного коэффициента. Следовательно, порог перехода становится свойством только линейной градиентной структуры.

Это физически объясняет то, почему, как нелинейность типа Поккельса (линейная электрооптическая), так и нелинейность типа Керра (квадратичная, зависящая от интенсивности), приводят к одинаковым критическим условиям, несмотря на принципиально различную зависимость амплитуды поля. Данный результат имеет важное значение для проектирования реконфигурируемых фотонных устройств, где локализация поля должна контролироваться без изменения нелинейного состава материала.

2. Анализ влияния оптических параметров на критическое значение показателя преломления перехода. Для того, чтобы наглядно продемонстрировать эффект перехода интенсивности поля из одной среды в другую при варьировании оптических параметров системы, выберем конкретную модель нелинейности и получим в явном виде решение краевой задачи для уравнений (3) и (4), описывающее поперечное распределение поля в поверхностной волне.

В качестве примера выберем модель нелинейности (8). Тогда при использовании решения (9) можно получить явный вид уравнения (18):

$$x_0 = -\frac{2}{k\sqrt{n^2 - \varepsilon_{N0}}} \operatorname{arcth} \left(\frac{2q_G}{k\sqrt{n^2 - \varepsilon_{N0}}} \right). \quad (19)$$

Для этого случая из (13) с учетом (19) можно найти амплитуду поля на границе раздела сред:

$$u_0 = \frac{3(n^2 - \varepsilon_{N0} - 4q_G^2 / k^2)}{2\varepsilon_{N1}}. \quad (20)$$

Тогда можно записать решение краевой задачи для уравнений (3) и (4) с нелинейностью (8) в явном виде:

$$u(x) = \frac{3}{2\varepsilon_{N1}} \begin{cases} (n^2 - \varepsilon_{N0} - 4q_G^2 / k^2) \frac{W_{\mu,\nu}(2k\sqrt{n^2 - \varepsilon_{G0}}(x+h))}{W_{\mu,\nu}(2kh\sqrt{n^2 - \varepsilon_{G0}})}, & x > 0, \\ \frac{3(n^2 - \varepsilon_{N0})}{2\operatorname{ch}^2(k\sqrt{n^2 - \varepsilon_{N0}}(x+x_0)/2)}, & x < 0. \end{cases} \quad (21)$$

где x_0 определяется (19), а q_G – (15). Для вещественности данного решения должны выполняться условия $n^2 > \max\{\varepsilon_{N0}, \varepsilon_{G0}\}$, $\varepsilon_{G2} < 1/4k^2$ и $|q_G(\lambda)| < k\sqrt{n^2 - \varepsilon_{N0}}$.

Решение (21) описывает поперечное распределение поля в бегущей вдоль границы раздела сред поверхностной волне. Максимум интенсивности этого распределения может находиться в зависимости от значений параметров системы как в нелинейной среде, так и в неоднородной (Рис.1).

При выборе модели нелинейности (8) максимум интенсивности поля будет располагаться в неоднородной среде при $n < n_c$ (линия 1 на Рис.1), поскольку при таких значениях $x_0 < 0$, как следует из (19). Здесь профиль с градиентным показателем преломления действует как более эффективный оптический потенциал, ограничивая волну.

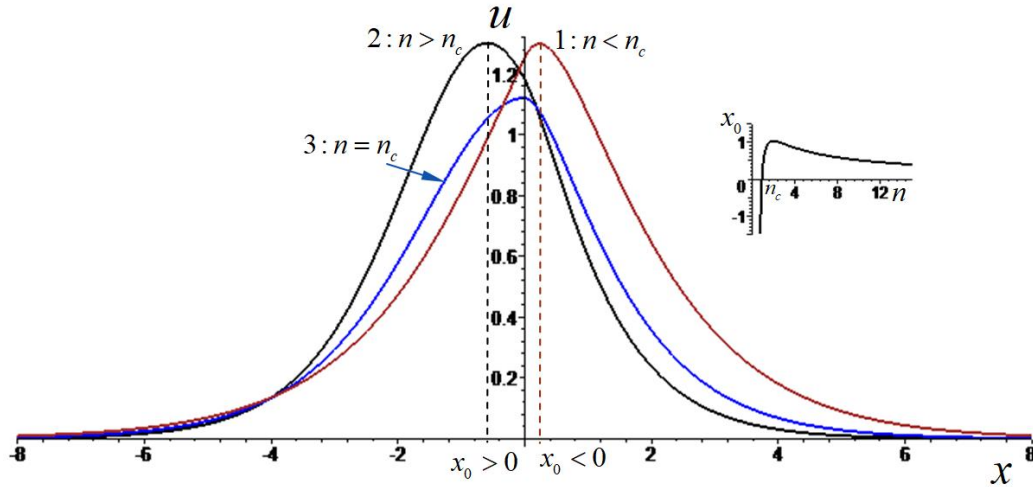


Рис. 1. Пространственные распределения интенсивности поля (21) при значениях параметров системы (в условных безразмерных единицах): $k=1, h=0.5, \varepsilon_{G0}=-0.1, \varepsilon_{G1}=1, \varepsilon_{G2}=0.4, \varepsilon_{N0}=0.1, \varepsilon_{N1}=0.4, n=0.75$ (1), $\varepsilon_{N1}=1.28, n=1.1$ (2), $\varepsilon_{N1}=1, n=0.92$ (3), а также зависимость x_0 от эффективного показателя преломления n (19) при $k=1, h=0.5, \varepsilon_{G0}=-0.1, \varepsilon_{G1}=1, \varepsilon_{G2}=0.2, \varepsilon_{N0}=0.1, \varepsilon_{N1}=1$.

При $n > n_c$ максимум интенсивности поля будет располагаться в нелинейной среде (линия 2 на Рис.1), так как при таких значениях будет $x_0 > 0$. В этом режиме доминирует нелинейный эффект самофокусировки (или линейное волноводное распространение в модели Поккельса), который обеспечивает более сильное ограничение для более высоких эффективных показателей преломления.

При $n = n_c$ будет $x_0 = 0$ и максимум интенсивности располагается на границе раздела сред, хотя и распределение поля при этом по обе стороны от нее не будет симметричным (линия 3 на Рис.1), но на больших расстояниях оно будет спадать экспоненциально по обе стороны от границы раздела сред. Из (19) видно, что $x_0 = 0$ будет при $q_G = 0$, как и ожидалось. Отсюда получается уравнение для нахождения критического значения эффективного показателя преломления n_c , при котором максимум интенсивности располагается на границе раздела сред:

$$W'_{\mu(n_c), \nu}(2kh\sqrt{n_c^2 - \varepsilon_{G0}}) = 0. \quad (22)$$

Здесь учтено, что нули производной функции Уиттекера не совпадают с нулями самой функции Уиттекера при одинаковых значениях параметров.

Если обозначить $s = 2kh\sqrt{n_c^2 - \varepsilon_{G0}}$ – корень уравнения:

$$W'_{(kh\varepsilon_{G1}/s), \nu}(s) = 0, \quad (23)$$

то из (23) вытекает критическое значение, определяющее критическое значение эффективного показателя преломления

$$n_c^2 = \varepsilon_{G0} + \left(\frac{s}{2kh}\right)^2. \quad (24)$$

Следует отметить, что уравнение (23) является сложным трансцендентным уравнением, поскольку искомая величина s входит как в аргумент, так и в индекс производной функции Уиттекера. Его решение можно найти только численно. Оно зависит только от параметров

градиентного профиля (2), моделирующего неоднородность оптических свойств среды: $s=s(kh, \varepsilon_{G0}, \varepsilon_{G1}, \varepsilon_{G2})$. Следовательно, от этих же параметров зависит критическое значение эффективного показателя преломления в соответствии с (24) $n_c=n_c(kh, \varepsilon_{G0}, \varepsilon_{G1}, \varepsilon_{G2})$. В этом случае (24) можно назвать законом дисперсии поверхностных волн, максимум интенсивности которых располагается в точности на границе раздела сред.

На рис. 2 представлены результаты анализа закона дисперсии (24) на основе решения уравнения (23). Критическое значение эффективного показателя преломления демонстрирует систематическую зависимость от геометрических и оптических параметров, определяющих профиль диэлектрической проницаемости градиентной среды. В частности, увеличение характерного расстояния профиля неоднородности h (он определяет пространственную скорость изменения показателя преломления в градиентной среде, а его большие значения подразумевают более медленное и плавное изменение показателя преломления) приводит к уменьшению критического значения n_c (рис. 2a).

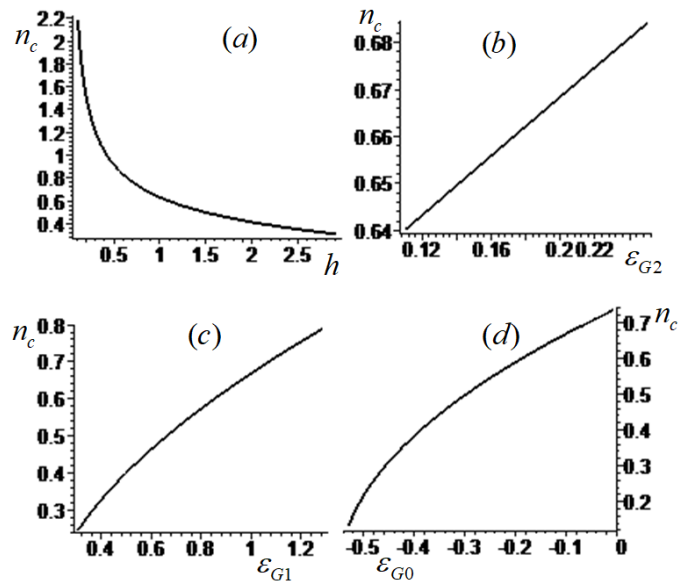


Рис. 2. Зависимости критического значения эффективного показателя преломления, рассчитанные из уравнения (24):

- a – от h при $\varepsilon_{G0}=-0.1, \varepsilon_{G1}=1, \varepsilon_{G2}=0.2, k=1$;
- b – от ε_{G2} при $\varepsilon_{G0}=-0.1, \varepsilon_{G1}=1, h=0.5, k=1$;
- c – от ε_{G1} при $\varepsilon_{G0}=-0.1, \varepsilon_{G2}=0.2, h=0.5, k=1$;
- d – от ε_{G0} при $\varepsilon_{G1}=1, \varepsilon_{G2}=0.2, h=0.5, k=1$.

Это обусловлено тем, что постепенный, более плавный, пространственный переход оптической плотности ослабляет волноводную силу вблизи границы раздела, что требует меньшего эффективного контраста показателя преломления для поддержания локализации поля на границе. Для поддержания пика интенсивности точно на границе раздела сред эффективный показатель преломления направляемой моды должен быть ниже, что уменьшает фазовое рассогласование и позволяет полю распространяться более симметрично относительно границы.

При увеличении остальных параметров профиля неоднородности (2) ε_{G2} (рис. 2b), ε_{G1} (рис. 2c) и $|\varepsilon_{G0}|$ (рис. 2d) наблюдается уменьшение критического значения эффективного показателя преломления. Характер всех зависимостей – монотонный. Этот набор параметров определяет базовое значение и форму (например, монотонную, с экстремумом) профиля диэлектрической проницаемости в градиентной среде. Монотонная зависимость n_c от параметров системы обеспечивает предсказуемую «кривую настройки».

Увеличение величины этих параметров, как правило, соответствует более сильному оптическому контрасту или более крутой эффективной потенциальной яме в градиентной среде. Это можно интерпретировать, как усиление волноводных свойств на градиентной стороне. Более крутые градиенты усиливают эффективный потенциал локализации в градиентной области, смещая положение равновесия максимума интенсивности, если это не компенсируется уменьшением эффективного показателя преломления. Для уравнивания сил, притягивающих максимум поля с обеих сторон и фиксирующих его на границе раздела, необходимо уменьшить волноводную силу эффективного показателя преломления (которая притягивает нелинейную сторону при более высоких значениях n), следовательно, требуется меньшее значение n_c .

Кроме того, смещение фоновой диэлектрической проницаемости ε_{G0} влияет на критическое условие за счет своего вклада в локальное волновое число в градиентной среде. Его более высокие значения увеличивают рассогласование фазовых скоростей на границе раздела, что требует корректировки эффективного показателя преломления для поддержания тонкого баланса, необходимого для локализации на границе раздела сред.

Следует отметить, что закон дисперсии, определяющий критическое значение эффективного показателя преломления, задается параметрами профиля (2), которые соответствуют модели градиентной среды, моделирующей ее неоднородность. Следовательно, регулирование их значений, а значит и критического значения эффективного показателя преломления, может осуществляться за счет формирования профиля неоднородности с такими характеристиками, значения которых будут попадать в нужную область для получения заданного распределения поля.

Однако такого управления профилем поля можно также добиться варьированием величины k , связанной с длиной волны в вакууме λ_0 лазерного излучения, возбуждающего поверхностную волну. Следовательно, в экспериментах в технических системах, составляющей частью которых будет являться рассматриваемая планарная волноводная структура, можно будет контролировать условия перехода максимума интенсивности поля из одной среды в другую путем подбора соответствующей длины волны рабочего лазера.

В рассматриваемой модели, как следует из анализа закона дисперсии (24), к уменьшению критического значения эффективного показателя приводит увеличение волнового числа k , и соответственно, уменьшение длины волны лазерного излучения. Следовательно, использование более короткой длины волны возбуждения облегчает переход пика интенсивности в градиентную среду при заданном фиксированном эффективном показателе преломления.

3. Возможное применение и практическое значение результатов моделирования.

Полученные результаты имеют непосредственное значение для прикладной фотоники и оптической инженерии. Предложенная теоретическая модель позволяет точно контролировать пространственное распределение интенсивности оптического поля в гибридных волноводных системах, что находит прямое применение в интегрированных фотонных схемах, требующих динамического перераспределения поля [22].

Одно из непосредственных применений заключается в разработке оптических переключателей, где переход максимума интенсивности через границу раздела материалов вызывает нелинейный отклик, например, усиление генерации второй гармоники, когда максимум поля пересекает нецентросимметричный кристалл, демонстрирующий нелинейность Поккельса [21]. Волноводная структура, работающая вблизи критической точки n_c , может функционировать, как сверхчувствительный оптический переключатель. Путем изменения параметров градиентного профиля для установки критического показателя

преломления на рабочей длине волны можно создавать селективные по длине волны переключающие элементы, где незначительная настройка длины волны вызывает резкое пространственное перемещение областей высокой интенсивности, тем самым модулируя эффективность нелинейного преобразования без механических компонентов. Такое перераспределение интенсивности может быть использовано для модуляции передачи сигнала между соседними выходными каналами с высоким контрастом в оптических модуляторах [23].

Кроме того, эти принципы применимы к проектированию поверхностных плазмон-поляритонов на границах раздела металл-диэлектрик с градиентными наложениями показателей преломления, где контроль глубины проникновения электромагнитных полей определяет чувствительность датчиков в безмаркерных биосенсорах. Это также может служить основой для разработки и усовершенствования датчиков. Высокая чувствительность положения пика к параметрам градиентного слоя (например, h , ε_{G1}) делает эту структуру перспективной для сенсорных приложений. Любой физический или химический процесс, изменяющий профиль показателя преломления (например, адсорбция молекул, изменения давления или температуры), будет сдвигать критическое значение перехода n_c . Мониторинг длины волны или угла ввода, необходимых для удержания пика на границе раздела, может обеспечить высокоточное измерение этих внешних изменений.

Возможность точного позиционирования максимумов интенсивности на функциональных границах раздела также полезна для фотонных кристаллических устройств и систем оптического захвата, где градиентные силы критически зависят от локализации поля [24]. Наиболее значимым является то, что зависимость критического состояния от длины волны позволяет осуществлять полностью оптическое управление распределением поля с использованием перестраиваемых лазерных источников, что дает возможность создавать реконфигурируемые фотонные схемы, в которых маршрутизация информации происходит за счет пространственного перераспределения оптической интенсивности, а не за счет традиционных механизмов направленной связи.

В фотонной инженерии возможность контролировать пространственное перекрытие максимума оптической интенсивности с нелинейным или линейным градиентным материалом позволяет оптимизировать конкретные взаимодействия. Например, для эффективного генерирования второй гармоники можно спроектировать систему таким образом, чтобы в рабочей точке пик был локализован внутри нелинейного кристалла Поккельса. И, наоборот, для приложений, основанных на линейных градиентных эффектах (например, для формирования определенных мод), пик может быть расположен внутри градиентного слоя.

Изложенные принципы предоставляют собой рекомендации по управлению перекрестными помехами и межслойной связью в многослойных фотонных интегральных схемах. Проектируя свойства интерфейса для установки определенного значения n_c , инженеры могут контролировать, будет ли оптическая мода преимущественно локализована в определенном слое или распределена на границе раздела, что имеет решающее значение для создания сложных многофункциональных чипов.

Заключение. В работе предложена математическая модель планарного волновода с границей раздела нелинейной и градиентной сред. Сформулирована и решена краевая задача, описывающая распространение поверхностных волн вдоль границы раздела между нелинейной и градиентной оптическими средами. Основным результатом работы является получение в явном виде трансцендентного уравнения, определяющего критический эффективный показатель преломления, при котором максимум интенсивности точно находится на границе. При переходе величины эффективного показателя преломления через

данное критическое значение происходит переход максимума интенсивности поля из одной среды в другую. Аналитически доказано, что это критическое значение не зависит от модели нелинейного отклика.

На примере нелинейности Поккельса было получено явное решение для распределения поля, которое использовалось для моделирования поведения критического эффективного показателя преломления в зависимости от варьирования оптических параметров системы. На основе проведенного анализа было установлено, что зависимости этого критического значения от всех параметров, характеризующих профиль градиентного показателя преломления, носят монотонный (возрастающий/убывающий) характер.

Было обсуждено возможное применение и практическое значение результатов моделирования. Практическая значимость подтверждается возможностью переноса полученных аналитических соотношений в инженерные методики расчета. В частности, установленные закономерности позволяют оценивать, какое изменение длины волны или толщины градиентного слоя вызовет смещение максимума поля из нелинейной области в градиентную. Таким образом, предложенная модель формирует не только теоретическую основу для понимания физики процесса, но и инструментарий для практической реализации управляемых волноводных структур с заданными свойствами.

Список источников

1. Meng Y., Chen Y., Lu L. et al. Optical meta-waveguides for integrated photonics and beyond. *Light: Science & Applications*, 2021, vol. 10, p. 235, DOI: 10.1038/s41377-021-00655-x.
2. Wang Y., Zhong L., Lau K.Y. et al. Precise mode control of laser-written waveguides for broadband, low-dispersion 3D integrated optics. *Light: Science & Applications*, 2024, vol. 13, p. 130, DOI: 10.1038/s41377-024-01473-7.
3. Muis D., Li Y., Barczyk R. et al. Broadband localization of light at the termination of a topological photonic waveguide. *Science Advances*, 2025, vol. 11, no. 16, p. eadr9569, DOI: 10.1126/sciadv.adr9569.
4. Caselli N., Riboli F., Intonti F. et al. Spatial steadiness of individual disorder modes upon controlled spectral tuning. *APL Photonics*, 2016, vol. 1, no. 4, p. 041301, DOI: 10.1063/1.4946852.
5. Krupa K., Fona R., Tonello A. et al. Spatial beam self-cleaning in second-harmonic generation. *Scientific Reports*, 2020, vol. 10, p. 7204, DOI: 10.1038/s41598-020-64080-7.
6. Wang Z.Y., Wu X., Xiong X. et al. Toward ultimate-efficiency frequency conversion in nonlinear optical microresonators. *Science Advances*, 2025, vol. 11, no. 18, p. eadu7605, DOI: 10.1126/sciadv.adu7605.
7. Echarrri A.R., Cox J.D., Yu R. et al. Enhancement of nonlinear optical phenomena by localized resonances. *ACS Photonics*, 2018, vol. 5, no. 4, pp. 1521-1527, DOI: 10.1021/acsphotonics.7b01562.
8. Wang X., Yin C., Cao Z. *Progress in planar optical waveguides*. Berlin, Springer, 2016, 241 p., DOI: 10.1007/978-3-662-48984-0.
9. *Fundamentals of optical waveguides*. Cambridge, Elsevier Inc., Academic Press, 2022, DOI: 10.1016/C2017-0-02432-1.
10. Agrawal G.P. *Physics and engineering of graded-index media*. New York, Cambridge University Press, 2023, 348 p., DOI: 10.1017/9781009282086.
11. Беспрозванных В.Г. *Нелинейная оптика* / В.Г. Беспрозванных, В.П. Первачук. – Пермь: Изд-во Перм. гос. техн. ун-та, 2011. – 200 с.
12. *Введение в моделирование систем*. – СПб.: РГГМУ, 2023. – 182 с.
13. Хакимзянов Г.С. *Математическое моделирование. Ч. 1: Общие принципы математического моделирования* / Г.С. Хакимзянов, Л.Б. Чубаров, П.В. Воронина. – Новосиб.: Новосиб. гос. ун-т., 2010. – 148 с.
14. Kivshar Y. All-dielectric meta-optics and non-linear nanophotonics. *National Science Review*, 2018, vol. 5, no. 2, pp. 144-158, DOI: 10.1093/nsr/nwy017.
15. Chen C.-L. *Foundations for guided-wave optics*. New York, John Wiley & Sons, 2005, 462 p., DOI: 10.1002/0470042222
16. Савотченко С.Е. Модели волноводов, сочетающих градиентные и нелинейно-оптические слои / С.Е. Савотченко // *Russian Technological Journal*, 2023. – Т. 11, № 4. – С. 84-93. – DOI: 10.32362/2500-316X-2023-11-4-84-93.

17. Singh B.K., Bijalwan A., Pandey P.C. et al. Photonic bandgaps engineering in double graded hyperbolic, exponential and linear index materials embedded one-dimensional photonic crystals. *Engineering Research Express*, 2019, vol. 1, no. 2, p. 025004, DOI: 10.1088/2631-8695/ab48a0.
18. Singh B.K., Bambole V., Rastogi V. et al. Multi-channel photonic bandgap engineering in hyperbolic graded index materials embedded one-dimensional photonic crystals. *Optics & Laser Technology*, 2020, vol. 129, p. 106293, DOI: 10.1016/j.optlastec.2020.106293.
19. Савотченко С.Е. Моделирование поверхностных волн в фотонных кристаллических структурах с профилем показателя преломления, убывающим с расстоянием от поверхности // *Russian Technological Journal*, 2026. – Т. 14, № 1. – С. 91-102. – DOI: 10.32362/2500-316X-2026-14-1-91-102.
20. Савотченко С.Е. Точное решение контактной краевой задачи для волноводной структуры из нелинейной и неоднородной оптических сред // *Письма в журнал технической физики*, 2026. – Т. 52, № 5. – С. 24-27. – DOI: 10.21883/PJTF.2026.05.59631.20008.
21. Boyd R.W. *Nonlinear optics*. 4th ed. London, Academic Press, 2020, 609 p.
22. Башкатов А.С. Состояние и потенциал развития фотонных интегральных схем на фосфиде индия, ниобате лития и кремнии / А.С. Башкатов, С.В. Корнилов, О.Н. Морозова и др. // *Радиоэлектронная отрасль: проблемы и их решения*, 2023. – № 11. – С. 2-10.
23. Li Y., Sun M., Miao T. et al. Towards high-performance Pockels effect-based modulators: review and projections. *Micromachines*, 2024, vol. 15, no. 7, p. 865, DOI: 10.3390/mi15070865.
24. Крайский А.А. О резонансном увеличении поля внутри фотонного кристалла в окнах прозрачности вблизи запрещенной зоны и некоторых его применениях / А.А. Крайский, А.В. Крайский, М.А. Казарян и др. // *Фотоника*, 2018. – Т. 12, № 8(76). – С. 802-822. – DOI: 10.22184/1993-7296.2018.12.8.802.822.

Савотченко Сергей Евгеньевич. Доктор физико-математических наук, доцент, профессор кафедры высшей математики-3 ФГБОУ ВО «МИРЭА – Российский технологический университет» (РТУ МИРЭА), профессор кафедры высшей математики и физики ФГБОУ ВО «Российский государственный геологический университет им. Серго Орджоникидзе» (МГРИ). AuthorID: 137272, SPIN: 2552-4344, ScopusID: 6603577988, ORCID: 0000-0002-7158-9145, Researcher ID: N-9227-2018, savotchenkose@mail.ru. 119454, Россия, Москва, пр. Вернадского, 78

Афанасьева Надежда Олеговна. Аспирант кафедры высшей математики и физики ФГБОУ ВО «Российский государственный геологический университет им. Серго Орджоникидзе» (МГРИ). AuthorID: 1278367, SPIN: 5458-9250, ORCID: 0009-0009-0044-7930, noafanasieva@mail.ru. 117997, Россия, Москва, ул. Миклухо-Маклая, д. 23.

UDC 517.95;535.3

DOI:10.25729/ESI.2026.43.3.005

Modeling the properties of the critical refractive index during peak intensity transitions between nonlinear and gradient optical media

Sergey E. Savotchenko^{1,2}, Nadezhda O. Afanasyeva²

¹MIREA – Russian technological university, Russia, Moscow, savotchenkose@mail.ru

²Sergo Ordzhonikidze Russian state university for geological prospecting, Russia, Moscow

Abstract. This paper presents a simulation of the spatial redistribution of optical field intensity in a planar waveguide structure formed by the interface between a nonlinear medium and a medium with a graded refractive index. A generalized dependence of permittivity on the distance from the contact surface of the media was chosen as a model for the gradient medium. A medium with a linear optical effect (Pockels effect), in which permittivity is linearly dependent on the amplitude of the electric field, was chosen as a model for the nonlinear medium. The study identifies critical operating modes in which the maximum intensity of the guided mode can be deliberately located within any of the constituent media or precisely at their interface by adjusting the optical parameters of the system that determine the critical value of the effective refractive index. The main result of the study is proof of the existence of a critical effective refractive index regardless of the choice of nonlinearity model, at which the

field maximum is precisely localized at the interface. It is shown that the value depends solely on the parameters determining the profile of the graded refractive index, which ensures the dispersion relation. Particular emphasis is placed on identifying how each of the gradient profile parameters individually affects the magnitude of the transition threshold. Aanalysis revealed that the critical value of the effective refractive index varies monotonically with changes in the characteristic spatial scale and dielectric parameters of the graded layer. It is shown that the nature of this influence is determined by the sign and magnitude of the corresponding dielectric constants, allowing for the targeted formation of the desired dependence of the critical refractive index on the optical parameters of a planar waveguide structure. Unlike most studies, which are limited to numerical modeling of specific configurations, this study proposes a rigorous analytical approach that allows for the derivation of explicit criteria for the existence of a critical regime. The potential use of the obtained results for calibrating experimental setups, adjusting phase-matching conditions for second-harmonic generation, and estimating acceptable parameter tolerances in integrated circuit manufacturing is also discussed. The results provide a theoretical basis for the development of methods for controlled light localization in hybrid waveguide structures by varying their material and geometric parameters or the operating wavelength.

Keywords: mathematical modeling, nonlinear equations, optical waveguide, nonlinear medium, boundary value problem

References

1. Meng Y., Chen Y., Lu L. et al. Optical meta-waveguides for integrated photonics and beyond. *Light: Science & Applications*, 2021, vol. 10, p. 235, DOI: 10.1038/s41377-021-00655-x.
2. Wang Y., Zhong L., Lau K.Y. et al. Precise mode control of laser-written waveguides for broadband, low-dispersion 3D integrated optics. *Light: Science & Applications*, 2024, vol. 13, p. 130, DOI: 10.1038/s41377-024-01473-7.
3. Muis D., Li Y., Barczy R. et al. Broadband localization of light at the termination of a topological photonic waveguide. *Science Advances*, 2025, vol. 11, no. 16, p. eadr9569, DOI: 10.1126/sciadv.adr9569.
4. Caselli N., Riboli F., Intonti F. et al. Spatial steadiness of individual disorder modes upon controlled spectral tuning. *APL Photonics*, 2016, vol. 1, no. 4, p. 041301, DOI: 10.1063/1.4946852.
5. Krupa K., Fona R., Tonello A. et al. Spatial beam self-cleaning in second-harmonic generation. *Scientific Reports*, 2020, vol. 10, p. 7204, DOI: 10.1038/s41598-020-64080-7.
6. Wang Z.Y., Wu X., Xiong X. et al. Toward ultimate-efficiency frequency conversion in nonlinear optical microresonators. *Science Advances*, 2025, vol. 11, no. 18, p. eadu7605, DOI: 10.1126/sciadv.adu7605.
7. Echarrri A.R., Cox J.D., Yu R. et al. Enhancement of nonlinear optical phenomena by localized resonances. *ACS Photonics*, 2018, vol. 5, no. 4, pp. 1521-1527, DOI: 10.1021/acsphotonics.7b01562.
8. Wang X., Yin C., Cao Z. *Progress in planar optical waveguides*. Berlin, Springer, 2016, 241 p., DOI: 10.1007/978-3-662-48984-0.
9. *Fundamentals of optical waveguides*. Cambridge, Elsevier Inc., Academic Press, 2022, DOI: 10.1016/C2017-0-02432-1.
10. Agrawal G.P. *Physics and engineering of graded-index media*. New York, Cambridge University Press, 2023, 348 p., DOI: 10.1017/9781009282086.
11. Besprozvannykh V.G., Pervadchuk V.P. *Nelinejnaya optika [Nonlinear Optics]*. Perm, Publishing House of Perm State Technical University, 2011. 200 p.
12. *Vvedenie v modelirovanie sistem [Introduction to System Modeling]*. SPb.: RSUHMU, 2023. 182.
13. Khakimzyanov G. S., Chubarov L. B., Voronina P. V. *Matematicheskoe modelirovanie. CH. 1: Obshchie principy matematicheskogo modelirovaniya [Mathematical Modeling Part 1: General Principles of Mathematical Modeling]*. Novosibirsk State University, Novosibirsk, 2010, 148.
14. Kivshar Y. All-dielectric meta-optics and non-linear nanophotonics. *National Science Review*, 2018, vol. 5, no. 2, pp. 144-158, DOI: 10.1093/nsr/nwy017.
15. Chen C.-L. *Foundations for guided-wave optics*. New York, John Wiley & Sons, 2005, 462 p., DOI: 10.1002/0470042222
16. Savotchenko S.E. Modeli volnovodov, sochetayushchikh gradientnyye i nelineyno-opticheskiye sloi [Models of waveguides combining gradient and nonlinear optical layers]. *Russian Technological Journal*, 2023, vol. 11, no. 4, pp. 84-93, DOI: 10.32362/2500-316X-2023-11-4-84-93.
17. Singh B.K., Bijalwan A., Pandey P.C. et al. Photonic bandgaps engineering in double graded hyperbolic, exponential and linear index materials embedded one-dimensional photonic crystals. *Engineering Research Express*, 2019, vol. 1, no. 2, p. 025004, DOI: 10.1088/2631-8695/ab48a0.

18. Singh B.K., Bambole V., Rastogi V. et al. Multi-channel photonic bandgap engineering in hyperbolic graded index materials embedded one-dimensional photonic crystals. *Optics & Laser Technology*, 2020, vol. 129, p. 106293, DOI: 10.1016/j.optlastec.2020.106293.
19. Savotchenko S.E. Modelirovaniye poverkhnostnykh voln v fotonnykh kristallicheskikh strukturakh s profilem pokazatelya prelomleniya, ubyayushchim s rasstoyaniyem ot poverkhnosti [Modeling of surface waves in photonic crystal structures with a refractive index profile decreasing with distance from the surface]. *Russian Technological Journal*, 2026, vol. 14, no. 1, pp. 91-102, DOI: 10.32362/2500-316X-2026-14-1-91-102.
20. Savotchenko S.E. Tochnoye resheniye kontaktnoy krayevoy zadachi dlya volnovodnoy struktury iz nelineynoy i neodnorodnoy opticheskikh sred [Exact solution of the contact boundary value problem for a waveguide structure made of nonlinear and inhomogeneous optical media]. *Technical Physics Letters*, 2026, vol. 52, no. 5, pp. 24-27, DOI: 10.21883/PJTF.2026.05.59631.20008.
21. Boyd R.W. *Nonlinear optics*. 4th ed. London, Academic Press, 2020, 609 p.
22. Bashkatov A.S., Kornilov S.V., Morozova O.N. et al. Sostoyaniye i potentsial razvitiya fotonnykh integral'nykh skhem na fosfide indiya, niobate litiya i kremnii [State and development potential of photonic integrated circuits on indium phosphide, lithium niobate and silicon]. *Radioelektronnaya otrasl': problemy i ikh resheniya* [Radioelectronic Industry: Problems and Solutions], 2023, no. 11, pp. 2-10.
23. Li Y., Sun M., Miao T. et al. Towards high-performance Pockels effect-based modulators: review and projections. *Micromachines*, 2024, vol. 15, no. 7, p. 865, DOI: 10.3390/mi15070865.
24. Kraiskii A.A., Kraiskii A.V., Kazaryan M.A. et al. O rezonansnom uvelichenii polya vnutri fotonnogo kristalla v oknakh prozrachnosti vblizi zapreshchennoy zony i nekotorykh yego primeneniyyakh [On resonant field enhancement inside a photonic crystal in transparency windows near the band gap and some of its applications]. *Fotonika* [Photonics], 2018, vol. 12, no. 8(76), pp. 802-822, DOI: 10.22184/1993-7296.2018.12.8.802.822.

Savotchenko Sergey Evgenievich, D.Sc. (Phys. and Mathematics), Associate Professor, Professor, Department of Higher Mathematics-3, Federal State Budgetary Educational Institution of Higher Education "MIREA – Russian Technological University" (RTU – MIREA), Professor The Department of Higher Mathematics and Physics, Federal State Budgetary Educational Institution of Higher Education "Sergo Ordzhonikidze Russian State University for Geological Prospecting" (MGRI). AuthorID: 137272, SPIN: 2552-4344, ScopusID: 6603577988, ORCID: 0000-0002-7158-9145, Researcher ID: N-9227-2018, savotchenkose@mail.ru. 119454, Russia, Moscow, Vernadsky Ave. 78.

Afanasieva Nadezhda Olegovna, Postgraduate student, The Department of Higher Mathematics and Physics, Federal State Budgetary Educational Institution of Higher Education "Sergo Ordzhonikidze Russian State University for Geological Prospecting" (MGRI). AuthorID: 1278367, SPIN: 5458-9250, ORCID: 0009-0009-0044-7930, noafanasieva@mail.ru. 117997, Russia, Moscow, Miklukho-Maklaya str., 23.

Статья поступила в редакцию 03.03.2026; одобрена после рецензирования 22.07.2026; принята к публикации 15.08.2026.

The article was submitted 03/03/2026; approved after reviewing 07/22/2026; accepted for publication 08/15/2026.

УДК 519.872.4

DOI:10.25729/ESI.2026.43.3.006

Перспективы использования систем массового обслуживания для моделирования работы маркетплейсов

Туренова Ирина Алексеевна, Моисеева Светлана Петровна, Шкленник Мария Александровна

Национальный исследовательский Томский государственный университет,
Россия, Томск, irenaturena@yandex.ru

Аннотация. В условиях постоянно растущего числа онлайн-платформ и усиливающейся конкуренции между маркетплейсами, оптимизация их работы и повышение эффективности управления становятся актуальными задачами. Математическое моделирование процессов продаж на маркетплейсах позволяет предсказать изменения условий и минимизировать риски. Модели должны учитывать различные факторы, такие, как сезонность спроса, конкуренция, изменения в логистике и политике маркетплейсов. Это помогает поставщикам оптимизировать цены, запасы и логистические процессы, что снижает финансовые потери и повышает устойчивость бизнеса. Маркетплейсы, такие, как Amazon, Wildberries или Ozon, ежедневно обрабатывают огромное количество транзакций и взаимодействий с клиентами, что требует разработки математических моделей систем управления и прогнозирования. В настоящей статье рассмотрены математические модели на основе систем массового обслуживания (СМО) с описанием их принципов и возможного применения в электронной коммерции. Для прогнозирования потоков клиентов, а также для оптимизации воронки продаж на маркетплейсах предлагаются модели массового обслуживания одно- и двухфазные с повторными обращениями, позволяющие спрогнозировать параметры планируемой маркетинговой акции и оценить ее влияние на доход и клиентскую базу. Для моделирования воронки продаж предлагаются многофазные модели, оценивающие время нахождения потенциального покупателя на соответствующем уровне воронки, что дает возможность оценить эффективность различных этапов воронки и выявить узкие места, которые могут снижать конверсию; разработать рекомендации по оптимизации воронки продаж, например, путём изменения дизайна интерфейса, улучшения системы поиска или предложения более выгодных условий оплаты; прогнозировать объём продаж и оптимизировать запасы товаров на основе анализа потоков заявок на разных этапах воронки. Статья закладывает основу для дальнейших исследований по адаптации этих моделей к реальным процессам, включая эмпирическую валидацию, определение параметров и разработку алгоритмов для оптимизации стратегий продаж селлеров.

Ключевые слова: маркетплейс, воронка продаж, экономико-математическое моделирование, вероятностные методы, системы массового обслуживания, потоки заявок, повторные обращения, многофазные СМО

Цитирование: Туренова И.А. Перспективы использования систем массового обслуживания для моделирования работы маркетплейсов / И.А. Туренова, С.П. Моисеева, М.А. Шкленник // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 70-82. – DOI:10.25729/ESI.2026.43.3.006.

Введение. В современном мире маркетплейсы становятся всё более популярными и востребованными. Они представляют собой платформы, где продавцы (селлеры) и покупатели могут взаимодействовать друг с другом, обеспечивая тем самым эффективный канал сбыта продукции [1]. Одним из ключевых преимуществ маркетплейсов является возможность сокращения расходов на рекламу. Благодаря таргетированным рекламным инструментам продавцы могут более эффективно привлекать целевую аудиторию, что особенно важно в условиях конкуренции. Кроме того, маркетплейсы предлагают удобные каналы дистрибуции для малого и микробизнеса. Это особенно актуально в рамках политики импортозамещения, когда малые предприятия могут использовать маркетплейсы для расширения своего присутствия на рынке [2]. Конкуренция на маркетплейсах среди предприятий малого бизнеса способствует поиску уникальных производственных и экономических решений.

Для потребителей маркетплейсы также предлагают ряд преимуществ, включая:

- Время. Покупки на маркетплейсах могут быть более удобными и быстрыми, поскольку предлагают широкий выбор товаров и возможность сравнить цены от разных продавцов.
- Выбор. На маркетплейсах представлен огромный ассортимент товаров, что позволяет потребителям найти именно то, что им нужно.
- Удобство. Особенно это актуально для людей с ограниченными возможностями и семей с детьми. Удобство совершения покупок может повысить качество жизни и социальную удовлетворённость населения.

Эффективная работа таких платформ требует оптимизации множества процессов, включая обработку заказов, логистику и управление ресурсами. В условиях высокой конкуренции на рынке маркетплейсы должны обеспечивать быструю и эффективную обработку заказов, чтобы удовлетворить потребности клиентов и сохранить свою репутацию, и для их успешного функционирования необходимо применять различные методы моделирования. Одним из ключевых аспектов является прогнозирование спроса на товары с помощью алгоритмов машинного обучения, таких, как регрессионный анализ, временные ряды, ансамблевые методы и нейронные сети [3, 4]. Эти инструменты позволяют анализировать исторические данные о продажах, сезонность и тренды, что помогает маркетплейсам оптимизировать запасы, устанавливать цены и планировать маркетинговые кампании. Наконец, модели машинного обучения для персонализации предложений помогают увеличить конверсию и удовлетворённость клиентов, создавая более индивидуальный опыт для покупателей.

Оптимизация ассортимента достигается с помощью методов линейного, целочисленного программирования и многокритериальной оптимизации [5]. Они помогают определить оптимальный набор товаров, максимизирующий прибыль или удовлетворяющий другие критерии эффективности. Управление запасами осуществляется с помощью моделей EOQ (Economic Order Quantity – «модель экономически обоснованного размера заказа»), ROP (Reorder Point – модель "точки заказа") и MRP (Material Requirements Planning – модель планирования потребностей в материалах), которые определяют оптимальные объёмы заказов и точки предзаказа. Это минимизирует затраты на хранение и предотвращает дефицит товаров [6]. Для оптимизации логистических процессов используется аппарат транспортного моделирования, маршрутизации и управления цепями поставок (SCM – Supply Chain Management) [7, 8]. Эти модели минимизируют затраты на логистику и улучшают обслуживание клиентов. Ценообразование на маркетплейсах оптимизируется с помощью моделей динамического ценообразования, учитывающих данные о спросе, конкуренции и затратах [9, 10]. Также применяются модели на основе предельной полезности и эластичности спроса [11]. Интегрированные модели объединяют методы прогнозирования спроса, оптимизации ассортимента, управления запасами и логистики, что позволяет достичь более высокой эффективности работы маркетплейса [12]. В некоторых работах используются модели на основе теории игр для понимания поведения участников маркетплейса и оптимизации стратегий [13, 14].

Таким образом, в руках маркетолога в наши дни может находиться широкий спектр инструментов для анализа и оптимизации различных сфер работы маркетплейса, однако многие из них теряют свою актуальность в связи с трансформацией экономических процессов и поведения клиентов. Например, популярные аналитические модели ценообразования основаны на предположениях относительно характера спроса на товары, очевидно не соответствующих реальной обстановке на цифровых платформах. Это закономерно, если принимать во внимание, что разрабатывались они в иное время и для иного рынка. Большинство используемых классических моделей и методов обладают преимуществом

легкости расчетов, и это сделало их популярными. Однако в современных коммерческих процессах уже невозможно игнорировать необходимость учитывать множество разнообразных факторов, влияющих на спрос (сезонность, трендовость, спрос на товары-компаньоны, транзакционный оффер, производство рекламного контента и пр.), так что быстрое преимущество сегодня попросту лишает такие модели адекватности. Однако все те же классические модели зачастую ложатся в основу алгоритмов разрабатываемых современных инструментов, что незамедлительно лишает их прогнозы желаемой точности [12].

В то же время, очевидной отличительной чертой современных процессов в сфере электронной торговли является массовость. Современные цифровые инструменты прекрасно приспособлены для обработки и работы с большими данными. На сегодняшний день такие инструменты наиболее актуальны и востребованы, так как позволяют анализировать информацию о поведении клиента на всех этапах воронки продаж. Однако даже на основании сверхточных исходных данных, прогноз, выдаваемый такими инструментами, будет основан на статистических методах и моделях, обладающих значительной степенью абстракции.

Обратимся к истокам статистических методов и заметим, что в том случае, когда наблюдается массовость явлений в условиях некоторой неопределенности, статистические методы преобразуются в вероятностные, а значит, и для моделирования процессов маркетинга вероятностные модели и методы окажутся адекватными.

1. Преимущества математических моделей на основе СМО. Теория массового обслуживания (ТМО) – это раздел прикладной математики, который изучает процессы, связанные с массовым обслуживанием. Она может быть использована для анализа и моделирования работы маркетинговых платформ, где множество покупателей и продавцов взаимодействуют друг с другом [15]. Для анализа различных процессов и явлений в работе платформы существует возможность выбора модели, оптимально отражающей необходимые аспекты и учитывающей значимые особенности поведения клиентов. Например, маркетинговые платформы часто сталкиваются с пиковыми нагрузками в праздничные сезоны, выходные дни и другие периоды повышенного спроса. Случайный характер потока заявок и время обслуживания приводят к неравномерной загрузке системы, что может вызвать как перегрузку каналов, так и их недогрузку. Для оптимизации работы СМО необходимо разработать стратегии управления потоками заявок, включая автоматическое масштабирование ресурсов в зависимости от текущей нагрузки [16, 17]. Теория массового обслуживания (ТМО) позволяет определить оптимальный вариант работы системы, минимизируя затраты от простоев каналов обслуживания, потерь времени и ресурсов. На маркетинговых платформах это включает в себя определение нужного количества касс, пунктов выдачи заказов, а также оптимальное распределение нагрузки между ними. Например, на крупных маркетинговых платформах, таких как Amazon, используется динамическое распределение заказов между складами в зависимости от текущей загрузки, что позволяет снизить время доставки и повысить удовлетворенность клиентов. Прогнозирование потоков клиентов на маркетинговых платформах позволяет оптимизировать работу системы, повысить эффективность бизнеса и улучшить качество обслуживания [18].

Итак, математические модели на основе СМО имеют ряд преимуществ, которые делают их незаменимыми инструментами для прогнозирования потоков клиентов на маркетинговых платформах:

1. Точность прогнозирования. Модели позволяют учесть множество факторов, влияющих на потоки клиентов, таких, как сезонность, праздничные дни, конкуренция, экономические условия, поведение пользователей и другие. Это обеспечивает более точное прогнозирование по сравнению с интуитивными методами или простыми статистическими моделями. Результаты анализа приведенных далее в качестве примеров СМО проходили сравнение с

результатами, полученными с помощью имитационного моделирования, и показали хорошую точность [19].

2. Гибкость. Модели могут быть адаптированы под конкретные условия работы маркетплейса, включая структуру системы, параметры потока заявок и другие факторы. Далее на примерах некоторых систем будет показано, как изменение структуры системы и введение отдельных параметров может быть связано с задачами анализа [20].

3. Возможность оптимизации. Модели позволяют определить оптимальные параметры работы СМО, такие, как количество каналов обслуживания, количество сотрудников, распределение нагрузки и другие. Это способствует снижению затрат, повышению производительности и улучшению качества обслуживания [21].

4. Прогнозирование поведения системы в различных условиях. Модели позволяют предсказать поведение системы в различных сценариях, таких, как изменение спроса, увеличение конкуренции, введение новых продуктов и другие. Это помогает принимать обоснованные решения и эффективно планировать ресурсы. Например, можно использовать модель для прогнозирования влияния введения новых категорий товаров на потоки клиентов и оптимизировать структуру системы для их обработки [22].

2. Некоторые модели массового обслуживания для процессов электронной коммерции.

Рассмотрим некоторые примеры возможного применения моделей в виде СМО. Наиболее простой моделью, позволяющей прогнозировать поведение покупателей, является система $M|GI|\infty$ с повторными обращениями [23]. Она позволяет анализировать отдельно потоки уже зарегистрированных пользователей маркетплейса и новых клиентов. В этой модели моменты совершения покупок новыми клиентами представляют собой события входящего потока, в то время как поток повторных покупок постоянных пользователей формируется по завершении стадии обслуживания, моделирующей время удовлетворенности покупателя до появления желания оформления следующего заказа. Система имеет неограниченное число обслуживающих линий, то есть число клиентов никак не ограничено моделью и может также прогнозироваться. Учитывается и вероятность того, что клиент может остаться неудовлетворенным и выберет другую платформу для дальнейшего обслуживания, то есть учитываются потери покупателей. (рис.1)

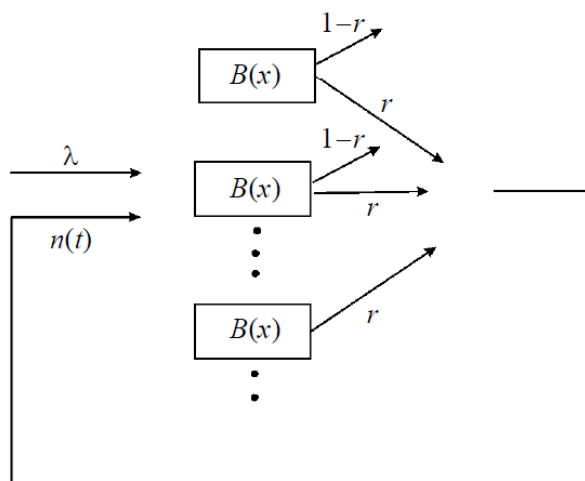


Рис. 1. Система $M|GI|\infty$ с повторными обращениями

Здесь предварительной статистической оценки требует интенсивность входящего потока λ , а также вероятность совершения повторной покупки r . Эта модель подойдет для анализа продаж широкого спектра товаров и услуг, так как распределение вероятностей времени

удовлетворенности клиента $B(x)$ произвольное, то есть может выбираться в зависимости от специфики реализуемого товара. Для данной модели получено аналитическое выражение для производящей функции распределения вероятностей двумерного случайного процесса, характеризующего потоки новых клиентов и постоянных пользователей маркетплейса за время наблюдения t :

$$G(x, y, t) = \exp \left\{ (x - 1)\lambda t + r(y - 1) \int_0^t h(x, y, s) ds \right\} \quad (1)$$

где функция $h(x, y, s)$ определяется из интегрального уравнения:

$$h(x, y, t) = \frac{\lambda}{1-r} + \left(\lambda x - \frac{\lambda}{1-r} \right) B(t) + ry \int_0^t b(t-s) h(x, y, s) ds, \quad (2)$$

а функция

$$b(t-s) = \left. \frac{\partial B(z+t-s)}{\partial z} \right|_{z=0}. \quad (3)$$

Таким образом, в работе [23] показано, что, так как полученная производящая функция $G(x, y, t)$ исследуемого двумерного распределения не равна произведению производящих функций одномерных распределений, то, очевидно, потоки новых и постоянных клиентов являются зависимыми и их анализ необходимо проводить только совместно.

На примере этой модели также можно продемонстрировать достигаемую точность моделирования. Полученное аналитическим путем распределение числа повторных обращений сравнивалось с таковым, полученным в результате имитационного моделирования. Например, для входящего простейшего потока с параметром $\lambda=3$, времени обслуживания, имеющего Гамма-распределение с параметром формы 3,5, вероятности возвращения $r=0,56$ и времени наблюдения $t=1$, получены эмпирическая и теоретическая функции распределения, представленные на рисунке 2.

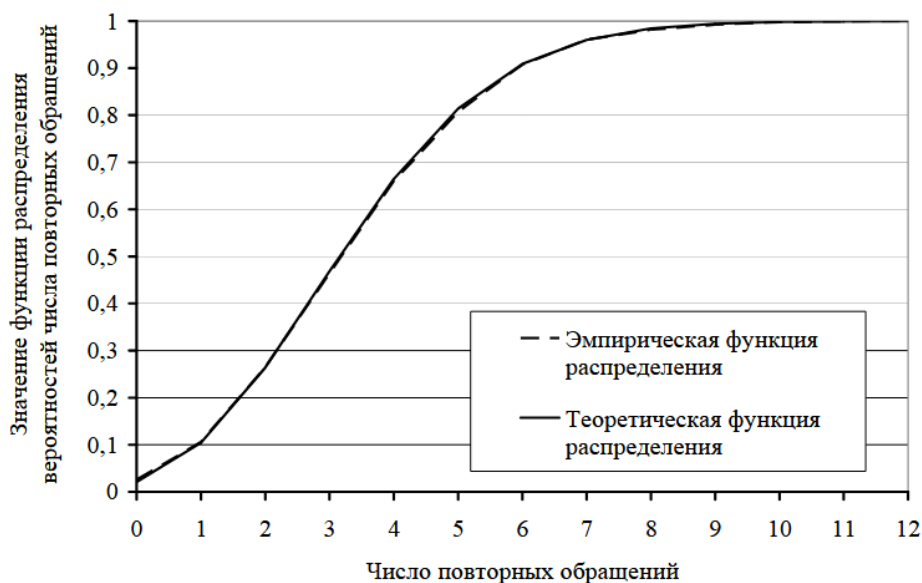


Рис. 2. Теоретическая и эмпирическая функции распределения вероятностей числа повторных обращений в системе

В приведенном примере значение статистики Хи-квадрат составило 10,8, что не превосходит порогового значения 67,5 при заданном уровне значимости $\alpha = 0,05$ и количестве степеней свободы 50. Для случая с входящим МАР-потокм были построены доверительные интервалы для точечных оценок распределения вероятностей числа повторных обращений и получено, что, согласно правилу трех сигм, с вероятностью 0,9973 абсолютная величина отклонения оценок не превышает 0,003.

Рассмотрим другой пример, когда нам необходимо спрогнозировать параметры планируемой маркетинговой акции и оценить ее влияние на доход и клиентскую базу. В такой

ситуации можно воспользоваться моделью в виде двухфазной СМО с неограниченным числом линий и реализацией повторных обращений к фазам [24]. В этой модели более детально рассматриваются потоки постоянных пользователей и происходит их условное деление на две категории. Например, совершая покупку на определенную сумму или просто приобретая VIP-аккаунт, клиент приобретает особый статус, то есть переходит на другую фазу обслуживания. На каждой из фаз клиент может оставаться неограниченное число раз, как бы возвращаясь для повторного обслуживания. Схематично образующиеся потоки клиентов можно изобразить, как показано на рисунке 3. Модель учитывает тот факт, что время между покупками обычных клиентов и имеющих VIP-подписку будет распределено по-разному, эти функции распределения времени удовлетворенности клиента отмечены на рисунке как $B_1(x)$ и $B_2(x)$. Таким образом, в модели помимо входящего потока новых клиентов, рассматривается формирование еще двух потоков постоянных клиентов двух категорий, с вероятностью r_1 и r_2 оформления повторных заказов соответственно, а также вероятностью q покупки VIP-аккаунта, то есть перехода из первой во вторую категорию.

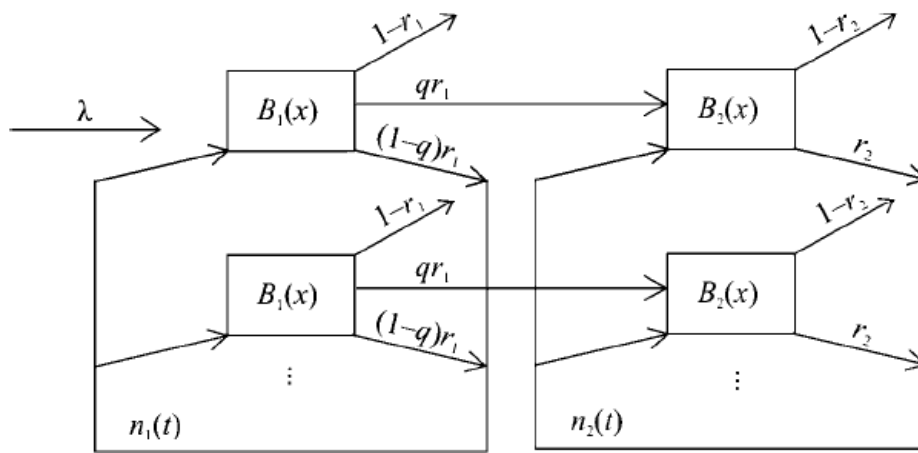


Рис. 3. Двухфазная СМО с повторными обращениями к фазам

Эта модель также не имеет ограничений на специфику товара или размер клиентской базы, то есть является достаточно универсальной и легко настраивается. Она позволяет определять и прогнозировать процесс изменения дохода компании, его среднее значение и дисперсию, оценивать влияние на доход транзакционных офферов¹.

Например, выражение для нахождения среднего дохода в общем случае имеет вид:

$$MS(t) = a \left\{ \lambda t + r_1 \int_0^t f_1(1, s) ds + r_2 \int_0^t f_2(1, s) ds \right\}, \quad (4)$$

где функции $f_1(x, t)$ и $f_2(x, t)$ определяются решением интегральных уравнений:

$$f_1(x, t) = \frac{\lambda}{1 - r_1(1 - q)} + \lambda \left(x - \frac{1}{1 - r_1(1 - q)} \right) B_1(t) + x r_1(1 - q) \int_0^t b_1(t - s) f_1(x, s) ds,$$

$$f_2(x, t) = \frac{\lambda}{1 - r_1(1 - q)(1 - r_2)} (1 - B_2(t)) + x r_1 q \int_0^t b_2(t - s) f_1(x, s) ds +$$

$$+ x r_2 \int_0^t b_2(t - s) f_2(x, s) ds,$$

а $b_1(t - s)$ и $b_2(t - s)$ определяются аналогично (3).

В частности, в работе [24] приведен численный пример по продажам в некоторой компании. Планировалась акция «Подарок за покупку» и необходимо было оптимизировать

¹ Оффер – конкретное, выгодное предложение о сотрудничестве или сделке.

стоимость предлагаемого подарка в зависимости от цели проведения акции. При средней стоимости покупки 3000 рублей и около 10 обращений новых клиентов в день, за месяц работы компания, не проводя акции, получит доход 1,2 млн рублей с продажи данного вида товара. Анализ показал, что максимальная прибыль от проведения акции достигается при установлении стоимости подарка в 1000 рублей. Число продаж при этом возрастает примерно в 1,5 раза. Если же компания не желает получать никакой дополнительной прибыли от проведения акции, а стремится только увеличить клиентскую базу и число продаж, то эта цель достигается при стоимости подарка равной 1700 рублей, число продаж компании возрастет более чем в два раза. Дальнейшее же увеличение стоимости подарка будет временно уменьшать доход компании. Для построения прогноза модель предварительно настраивалась: оценивались параметр входящего потока, вероятности перехода на вторую фазу и потери клиента, выбирался вид распределения, характерный для данного вида продаж.

3. Моделирование воронки продаж на маркетплейсах. Воронка продаж – это модель, которая описывает путь клиента от первого контакта с товаром до совершения покупки. На маркетплейсах воронка продаж может быть сложной и многоуровневой, включая этапы поиска товара, добавления его в корзину, оформления заказа, оплаты и доставки.

Моделирование воронки продаж на основе методов ТМО позволяет: оценить эффективность различных этапов воронки и выявить узкие места, которые могут снижать конверсию; разработать рекомендации по оптимизации воронки продаж, например, путём изменения дизайна интерфейса, улучшения системы поиска или предложения более выгодных условий оплаты; прогнозировать объём продаж и оптимизировать запасы товаров на основе анализа потоков заявок на разных этапах воронки.

Методы ТМО могут быть использованы для моделирования различных параметров воронки продаж, таких, как: интенсивность потока заявок на каждом этапе воронки; среднее время, затрачиваемое клиентами на переход от одного этапа к другому; вероятность потери клиента; прогноз числа клиентов на каждом уровне воронки в интересующий момент времени. Это позволяет более точно оценить эффективность воронки продаж и разработать стратегии для её оптимизации.

В качестве уже существующей модели СМО, хорошо подходящей под эту задачу, можно рекомендовать СМО с неограниченным числом фаз и линий [25], схема которой приведена на рисунке 4.

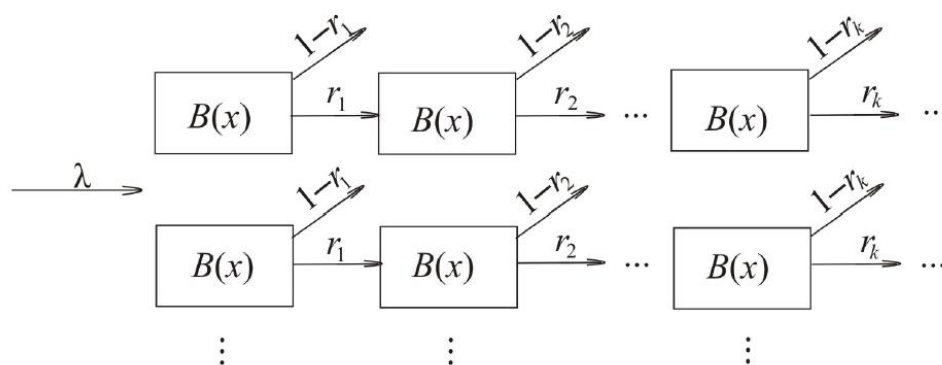


Рис. 4. СМО с неограниченным числом фаз и линий

В этой модели время обслуживания на каждой фазе моделирует время нахождения потенциального покупателя на соответствующем уровне воронки. Распределение времени вновь произвольное и может задаваться, исходя из конкретной реальной ситуации. Вероятности r_i представляют вероятности продвижения клиента с i -го уровня воронки вглубь. Так как модель не накладывает ограничений на число фаз, она универсальна и может как

настраиваться под уже известную структуру воронки, так и дает возможность определения оптимального числа уровней в рамках конкретной коммерческой задачи. Однако модель не учитывает различий в распределении времени пребывания клиента на различных уровнях воронки и должна быть модифицирована.

В случае необходимости анализа с учетом дальнейших повторных покупок система также может быть модифицирована до многофазной бесконечнолинейной СМО с неограниченным числом линий и мгновенной обратной связью на последней фазе (рис. 5) [26]. В этой модели число фаз фиксировано, но может выбираться соответственно исследуемой воронке продаж. Формирование потока повторных обращений на последней фазе позволяет проводить анализ не только процесса продвижения клиентов по воронке, но и дальнейшее их поведение в качестве постоянных клиентов. Недостатком модели является относительная узость ее применения в связи с экспоненциальным временем удовлетворенности клиента. Это может накладывать ограничения на специфику товара или временные рамки проводимого анализа.

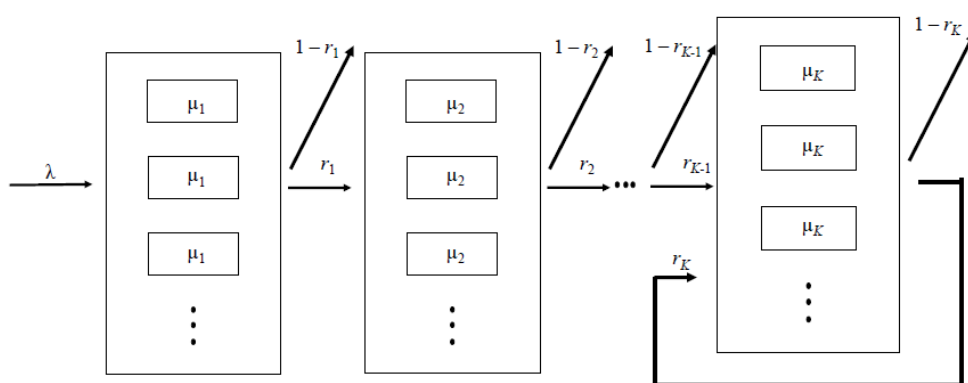


Рис. 5. СМО с неограниченным числом линий и повторными обращениями на последней фазе

В работе [27] рассматривается СМО, которая позволяет не только моделировать потоки потенциальных покупателей, продвигающихся вглубь воронки продаж, но и учитывать вероятность возврата клиента на некоторый уровень, после которого клиент может снова углубляться для совершения покупки или совсем покидать воронку. Здесь также рассматривается экспоненциальное время удовлетворенности клиента, что сужает область возможного применения этой модели (рис.6).

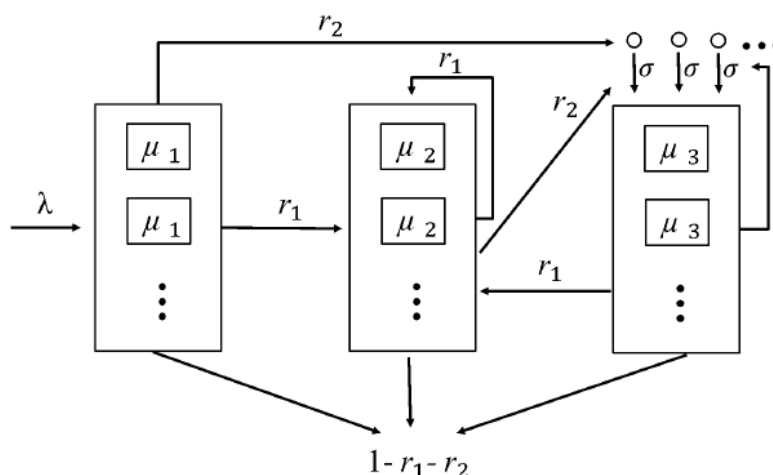


Рис.6. Система массового обслуживания с неограниченным числом обслуживающих устройств с мгновенной и отсроченной обратной связью

Такая модель применима для оптимизации параметров транзакционного оффера для контент-воронки или при построении триал-воронки. При этом параметр σ в модели характеризует случайное время потребления пробного товара или контента, предоставляемого на очередном этапе воронки, учитываются возможности движения потенциального клиента между уровнями воронки, а также многократное потребление пробного товара или рекламного контента, что является отличительной чертой этой модели.

В работах [25-27] получены аналитические выражения для распределения вероятностей исследуемых многомерных случайных процессов, и они могут непосредственно использоваться в рамках исследования процесса продаж, если соответствующая модель применима без дополнительных модификаций, или могут служить ориентиром для проведения похожих исследований новых модифицированных моделей.

Заключение. В условиях растущей конкуренции на рынке маркетплейсы должны постоянно искать пути оптимизации своих процессов. Модели на основе СМО могут стать ценным инструментом для достижения этой цели, обеспечивая более эффективное использование ресурсов и повышение качества обслуживания клиентов. Они дают возможность учесть специфику работы маркетплейсов, связанную с большим количеством участников и разнообразием товаров, обеспечивают достаточно точное моделирование потоков заказов и ресурсов, что позволяет оптимизировать их распределение и использование; учитывают динамический характер потоков заявок при моделировании воронки продаж; позволяют проводить глубокий анализ влияния различных параметров на работу системы. Это дает возможность выявлять узкие места и оптимизировать их, повышая общую эффективность работы маркетплейса и воронки продаж.

Стоит отметить, что все приведенные в работе модели являются аналитическими и обладают всеми преимуществами и универсальностью последних. Они могут как использоваться самостоятельно, так и быть основой алгоритмов разрабатываемых цифровых инструментов и имитационных моделей, позволяя существенно повышать точность последних.

В настоящей статье приведены примеры моделей, которые в дальнейшем будут адаптированы под конкретные задачи маркетплейсов с учетом особенностей поведения пользователей, наличия трендов, социальной коммерции и иных запросов на оптимизацию. Запланировано проведение комплекса исследований на тему адаптации систем и методов ТМО для моделирования процессов электронной коммерции. В рамках данного комплекса будут определены оптимальные методы оценки параметров моделей на основе реальных данных, сформирован набор моделей, востребованных для программной реализации и создания пользовательского интерфейса, с целью дальнейшей апробации на рынке.

Список источников

1. Бувич А.П. Маркетплейсы как ключевое звено современных каналов сбыта // Экономика, предпринимательство и право, 2024. – № 6. – С. 2875-2884. – DOI: 10.18334/epp.14.6.121249
2. Панова Е.А. Возможности маркетплейсов в развитии и повышении конкурентоспособности малых и средних предприятий // Государственное управление. Электронный вестник, 2021. – № 89. – С. 52–61. – DOI: 10.24412/2070-1381-2021-89-52-61
3. Вишневский В.М. Методы машинного обучения в решении задачи прогнозирования спроса на отдельные виды товаров / В.М. Вишневский, Ю.Л. Леохин, Т.Дж. Фатхулин, А.В. Занегин // Т-Comm, 2024. – № 10. – С. 34-43. –10.36724/2072-8735-2024-18-10-34-43
4. Вардомацкая Е.Ю. Экономико-математические методы и модели в логистике: конспект лекций. – Витебск: УО «ВГТУ», 2020. – 76 с.
5. Шориков А.Ф. Многокритериальная оптимизация формирования ассортимента продукции предприятия / А.Ф. Шориков, Е.С. Рассадина // Экономика региона, 2010. – № 2. – С. 189-196.
6. Мироседи С.А. Управление оптимальным объемом запасов / С.А. Мироседи, Т.Г. Мироседи, А.Ю. Севрюк // Инновационная наука, 2016. – № 12-1. – С. 150-152.

7. Иванова Л.Н. Методы оптимизации и алгоритм маршрутизации в транспортной логистике / Л.Н. Иванова, С.Е. Иванов // Экономика. Право. Инновации, 2024. – № 4. – С. 21–29. – DOI:10.17586/2713-1874-2024-4-21-29
8. Жук М.А. Моделирование системы поддержки оптимизации логистических цепочек поставок / М.А. Жук, И.А. Цыганова // Инновации и инвестиции, 2016. – № 12. – С. 104-108.
9. Иваненко А.Р. Модели динамического и персонализированного ценообразования в условиях развития цифровой экономики / А.Р. Иваненко, Е.Г. Калабина // e-FORUM, 2022. – Т. 6, № 1.
10. Мансурова Н.А. Разработка и применение модели динамического ценообразования для сглаживания конкуренции на маркетплейсах / Н.А. Мансурова, Е.С. Грачев // Экономические исследования, 2025. – № 2. – EDN: HJTSDH
11. Корабейников И.Н. Исследование изменения эластичности спроса на товары // Шаг в науку, 2023. – № 2. – С. 4-10.
12. Сергеев С.А. Управление запасами в условиях маркетплейсов: почему не работают стандартные методы? // Вестник Академии знаний, 2024. – № 5 (64). – С. 357-373.
13. Унижаев Н.В. Использование теории игр для моделирования последствий управленческих решений экономической безопасности организации // Экономика, предпринимательство и право, 2022. – Т. 12, № 1. – С. 149-164.
14. Кравченко С.М. Использование теории игр для моделирования манипулирования отзывами на маркетплейсах // ИВД, 2025. – № 9 (129).
15. Лизакова Р.А. Особенности продажи товаров на маркетплейсах / Р.А. Лизакова, В. Челябинца // Умная цифровая экономика, 2022. – № 3. – С. 12–16.
16. Истомина А.А. Управление товарными запасами на основе теории массового обслуживания / А.А. Истомина, В.Я. Бадеников, А.Л. Истомин // Вестник СибГУТИ, 2017. – № 3 (39). – С. 94-103.
17. Wang Y., Cai C. Strategic behavior and optimal inventory level in a make-to-stock queueing system with retrial customers. Axioms, 2024, vol. 13, no. 5, p. 319, DOI: 10.3390/axioms13050319
18. Истомина А.А. Поддержка принятия решений при управлении многономенклатурными запасами в условиях неопределенности / А.А. Истомина, В.Я. Бадеников, А.Л. Истомин // Информационные и математические технологии в науке и управлении, 2017. – № 2 (6). – С. 104-112.
19. Захорольная И.А. Исследование математических моделей потоков в системах с неограниченным числом линий методом предельной декомпозиции: дис. ... канд. физ.-мат. наук: 05.13.18 / Захорольная Ирина Алексеевна; Том. гос. ун-т. – Томск, 2012. – 171 с.
20. Zhang Y., Yue D., Wang J. A queueing-inventory system with lost sales and a general vacation policy. Journal of Industrial and Management Optimization, 2025, vol. 21, no. 6, pp. 4233-4258, DOI: 10.3934/jimo.2025051.
21. Rizki H., Yohandoko O. Restaurants using the M/M/c model: a case study during peak hours. International Journal of Global Operations Research, 2024, vol. 5, pp. 263-267.
22. Jacob J., Krishnamoorthy A. Product form solutions of some queueing inventory problems in random environment. Journal of Industrial and Management Optimization, 2023, vol. 19, pp. 1-16.
23. Моисеева С.П. Исследование потоков в системе $M|GI|\infty$ с повторными обращениями методом предельной декомпозиции / С.П. Моисеева, И.А. Ананина, А.А. Назаров // Вестник Томского государственного университета. Управление, вычислительная техника и информатика, 2009. – № 3(8). – С. 56-66.
24. Ананина И.А. Математическая модель процесса изменения дохода торговой компании, расширяющей свое присутствие на рынке // Вестник Томского государственного университета. Управление, вычислительная техника и информатика, 2011. – № 2(15). – С. 5-14.
25. Назаров А.А. Математическая модель процедуры пожизненной ренты / А.А. Назаров, И.А. Ананина // Известия Томского политехнического университета, 2011. – Т. 318, № 5. – С. 160-165.
26. Шкленник М.А. Исследование локального потока повторных обращений в многофазной системе массового обслуживания с неограниченным числом приборов / М.А. Шкленник, А.В. Подгайнов // Системы управления, информационные технологии и математическое моделирование: материалы V Всероссийской научно-практической конференции с международным участием. – С. 316-323.
27. Лавриненко М.Д. Исследование бесконечнолинейной СМО с проверкой качества обработки при нестационарном режиме работы / М.Д. Лавриненко, М.А. Шкленник // ИТММ, 2024. – С. 208-213.

Туренова Ирина Алексеевна. Кандидат физ.-мат. наук, доцент кафедры теории вероятностей и математической статистики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета. Область научных интересов – математическое моделирование экономических систем, теория массового обслуживания, математический маркетинг. Автор работ, в которых применяет метод предельной декомпозиции для анализа потоков обращений в системах с неограниченным числом приборов и обратной связью, и применения результатов для решения оптимизационных

задач маркетинговых услуг. AuthorID: 1120518, SPIN: 2771-6158, ORCID: 0000-0002-4373-9206, irenaturena@yandex.ru. 634050, Россия, Томск, пр. Ленина, 36.

Моисеева Светлана Петровна. Доктор физ.-мат. наук, профессор, заведующий кафедрой теории вероятностей и математической статистики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета. Область научных интересов – теория массового обслуживания и ее приложения к анализу информационных, экономических и социальных систем. Автор оригинальных методов исследования систем массового обслуживания немарковских систем массового обслуживания с неограниченным числом приборов, в том числе асимптотических. AuthorID: 9513, SPIN: 7526-2190, ORCID: 0000-0001-9285-1555, smoiseeva@mail.ru. 634050, Россия, Томск, пр. Ленина, 36.

Шкленник Мария Александровна. Кандидат физ.-мат. наук, доцент кафедры теории вероятностей и математической статистики Института прикладной математики и компьютерных наук Национального исследовательского Томского государственного университета. Область научных интересов – многофазные модели массового обслуживания с мгновенной обратной связью и методы их исследования. Автор модифицированного метод марковского суммирования для исследования потоков повторных обращений в системах с обратной связью. AuthorID: 937206, SPIN: 1790-5249, ORCID: 0000-0003-0993-8006, shklennikm@mail.ru 634050, Россия, Томск, пр. Ленина, 36.

UDC 519.872.4

DOI:10.25729/ESI.2026.43.3.006

Prospects of using queuing systems for modeling the operation of marketplaces

Irina A. Turenova, Svetlana P. Moiseeva, Maria A. Shklennik

National research Tomsk state university,

Russia, Tomsk, irenaturena@yandex.ru

Abstract. With the constantly growing number of online platforms and intensifying competition among marketplaces, optimizing their operations and improving management efficiency are becoming crucial tasks. Mathematical modelling of sales processes on marketplaces allows predicting changes in conditions and minimizing risks. Models must consider various factors such as demand seasonality, competition, changes in logistics, and marketplace policies. This helps suppliers optimize prices, inventory, and logistics processes, reducing financial losses and enhancing business stability. Marketplaces like Amazon, Wildberries, or Ozon process an enormous number of transactions and customer interactions daily, requiring the development of mathematical models for management systems and forecasting. This article explores mathematical models based on queuing systems (QS) describing their principles and potential applications in e-commerce. For predicting customer flows and optimizing the sales funnel on marketplaces, single- and two-phase queuing models with repeated requests are proposed, allowing forecasting parameters of planned marketing campaigns and assessing their impact on revenue and customer base. For modelling the sales funnel, multi-phase models are suggested, estimating the time a potential buyer spends at each funnel level, enabling evaluation of the effectiveness of different funnel stages and identifying bottlenecks that can reduce conversion; developing recommendations for sales funnel optimization, such as changing interface design, improving search systems, or offering more favourable payment terms; and predicting sales volume and optimizing product inventory based on analysing application flows at different funnel stages. The article lays the foundation for further research on adapting these models to real processes, including empirical validation, parameter determination, and algorithm development for optimizing seller sales strategies.

Keywords: E-commerce platform, sales funnel, econometric-mathematical modelling, stochastic methods, queuing systems, request flows, repeat inquiries, multi-phase QS

References

1. Buevich A.P. Marketpleysy kak klyuchevoye zveno sovremennykh kanalov sbyta [Marketplaces as a key link in modern distribution channels]. *Ekonomika, predprinimatel'stvo i pravo* [Economics, Entrepreneurship and Law], 2024, no. 6, pp. 2875-2884, DOI: 10.18334/epp.14.6.121249.

2. Panova E.A. Vozmozhnosti marketpleysov v razvitii i povyshenii konkurentosposobnosti malyykh i srednikh predpriyatiy [Opportunities for marketplaces in the development and competitiveness enhancement of small and medium enterprises]. Gosudarstvennoye upravleniye. Elektronnyy vestnik [Public Administration. Electronic Bulletin], 2021, no. 89, pp. 52–61, DOI: 10.24412/2070-1381-2021-89-52-61.
3. Vishnevsky V.M., Leokhin Yu.L., Fathulin T.J., Zanegin A.V. Metody mashinnogo obucheniya v reshenii zadachi prognozirovaniya sprosa na otdel'nyye vidy tovarov [Machine learning methods for solving the problem of demand forecasting for certain types of goods]. T-Comm, 2024, no. 10, pp. 34-43, DOI: 10.36724/2072-8735-2024-18-10-34-43.
4. Vardomatskaya E.Yu. Ekonomiko-matematicheskiye metody i modeli v logistike: konspekt lektsiy [Economic-mathematical methods and models in logistics: lecture notes]. Vitebsk, VSTU Publ., 2020, 76 p.
5. Shorikov A.F., Rassadina E.S. Mnogokriterial'naya optimizatsiya formirovaniya assortimenta produktsii predpriyatiya [Multi-criteria optimization of the formation of an enterprise's product assortment]. Ekonomika regiona [Economics of the Region], 2010, no. 2, pp. 189-196.
6. Mirosedi S.A., Mirosedi T.G., Sevryuk A.Yu. Upravleniye optimal'nym ob'yemom zapasov [Management of optimal inventory volume]. Innovatsionnaya nauka [Innovative Science], 2016, no. 12-1, pp. 150-152.
7. Ivanova L.N., Ivanov S.E. Metody optimizatsii i algoritm marshrutizatsii v transportnoy logistike [Optimization methods and routing algorithm in transport logistics]. Ekonomika. Pravo. Innovatsii [Economics. Law. Innovations], 2024, no. 4, pp. 21–29, DOI: 10.17586/2713-1874-2024-4-21-29.
8. Zhuk M.A., Tsyganova I.A. Modelirovaniye sistemy podderzhki optimizatsii logisticheskikh tsepochek postavok [Modeling a support system for optimizing logistics supply chains]. Innovatsii i investitsii [Innovations and Investments], 2016, no. 12, pp. 104-108.
9. Ivanenko A.R., Kalabina E.G. Modeli dinamicheskogo i personalizirovannogo tsenoobrazovaniya v usloviyakh razvitiya tsifrovoy ekonomiki [Models of dynamic and personalized pricing in the conditions of digital economy development]. e-FORUM, 2022, vol. 6, no. 1.
10. Mansurova N.A., Grachev E.S. Razrabotka i primeneniye modeli dinamicheskogo tsenoobrazovaniya dlya sglazhivaniya konkurentsii na marketpleysakh [Development and application of a dynamic pricing model for smoothing competition on marketplaces]. Ekonomicheskiye issledovaniya [Economic Studies], 2025, no. 2, EDN: HJTSDH.
11. Korabeinikov I.N. Issledovaniye izmeneniya elastichnosti sprosa na tovary [Study of changes in demand elasticity for goods]. Shag v nauku [Step into Science], 2023, no. 2, pp. 4-10.
12. Sergeev S.A. Upravleniye zapasami v usloviyakh marketpleysov: pochemu ne rabotayut standartnyye metody? [Inventory management in marketplace conditions: why do standard methods fail?]. Vestnik Akademii znaniy [Bulletin of the Academy of Knowledge], 2024, no. 5 (64), pp. 357-373.
13. Unizhaev N.V. Ispol'zovaniye teorii igr dlya modelirovaniya posledstviy upravlencheskikh resheniy ekonomicheskoy bezopasnosti organizatsii [Using game theory to model the consequences of management decisions on the economic security of an organization]. Ekonomika, predprinimatel'stvo i pravo [Economics, Entrepreneurship and Law], 2022, vol. 12, no. 1, pp. 149-164.
14. Kravchenko S.M. Ispol'zovaniye teorii igr dlya modelirovaniya manipulirovaniya otzyvami na marketpleysakh [Using game theory to model review manipulation on marketplaces]. IVD, 2025, no. 9 (129).
15. Lizakova R.A., Chelyapina V. Osobennosti prodazhi tovarov na marketpleysakh [Features of selling goods on marketplaces]. U mnaya tsifrovaya ekonomika [Smart Digital Economy], 2022, no. 3, pp. 12–16.
16. Istomina A.A., Badenikov V.Ya., Istomin A.L. Upravleniye tovarnymi zapasami na osnove teorii massovogo obsluzhivaniya [Inventory management based on queueing theory]. Vestnik SibGUTI [Bulletin of SibSUTI], 2017, no. 3 (39), pp. 94-103.
17. Wang Y., Cai C. Strategic behavior and optimal inventory level in a make-to-stock queueing system with retrial customers. Axioms, 2024, vol. 13, no. 5, p. 319, DOI: 10.3390/axioms13050319.
18. Istomina A.A., Badenikov V.Ya., Istomin A.L. Podderzhka prinyatiya resheniy pri upravlenii mnogonomenklaturnymi zapasami v usloviyakh neopredelennosti [Decision support in multi-item inventory management under uncertainty]. Informatsionnyye i matematicheskiye tekhnologii v nauke i upravlenii [Information and Mathematical Technologies in Science and Management], 2017, no. 2 (6), pp. 104-112.
19. Zakhrolnaya I.A. Issledovaniye matematicheskikh modeley potokov v sistemakh s neogranichennym chislom liniy metodom predel'noy dekompozitsii [Study of mathematical models of flows in systems with an unlimited number of lines by the method of limiting decomposition]. PhD dissertation. Tomsk State University, Tomsk, 2012, 171 p.
20. Zhang Y., Yue D., Wang J. A queueing-inventory system with lost sales and a general vacation policy. Journal of Industrial and Management Optimization, 2025, vol. 21, no. 6, pp. 4233-4258, DOI: 10.3934/jimo.2025051.

21. Rizki H., Yohandoko O. Restaurants using the M/M/c model: a case study during peak hours. *International Journal of Global Operations Research*, 2024, vol. 5, pp. 263-267.
22. Jacob J., Krishnamoorthy A. Product form solutions of some queueing inventory problems in random environment. *Journal of Industrial and Management Optimization*, 2023, vol. 19, pp. 1-16, DOI: 10.3934/jimo.2022226.
23. Moiseeva S.P., Ananina I.A., Nazarov A.A. Issledovaniye potokov v sisteme $M|GI|\infty$ s povtornymi obrashcheniyami metodom predel'noy dekompozitsii [Study of flows in the $M|GI|\infty$ system with retrials by the method of limiting decomposition]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravleniye, vychislitel'naya tekhnika i informatika* [Bulletin of Tomsk State University. Control, Computer Engineering and Informatics], 2009, no. 3(8), pp. 56-66.
24. Ananina I.A. Matematicheskaya model' protsessa izmeneniya dokhoda torgovoy kompanii, rasshiryayushchey svoye prisutstviye na rynke [Mathematical model of the income change process of a trading company expanding its market presence]. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravleniye, vychislitel'naya tekhnika i informatika* [Bulletin of Tomsk State University. Control, Computer Engineering and Informatics], 2011, no. 2(15), pp. 5-14.
25. Nazarov A.A., Ananina I.A. Matematicheskaya model' protsedury pozhiznennoy renty [Mathematical model of the annuity procedure]. *Izvestiya Tomskogo politekhnicheskogo universiteta* [Bulletin of Tomsk Polytechnic University], 2011, vol. 318, no. 5, pp. 160-165.
26. Shklenik M.A., Podgainov A.V. Issledovaniye lokal'nogo potoka povtornykh obrashcheniy v mnogofaznoy sisteme massovogo obsluzhivaniya s neogranichennym chislom priborov [Study of the local retrial flow in a multi-phase queueing system with an unlimited number of servers]. *Sistemy upravleniya, informatsionnyye tekhnologii i matematicheskoye modelirovaniye: materialy V Vserossiyskoy nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiyem* [Control Systems, Information Technologies and Mathematical Modeling: Proceedings of the V All-Russian Scientific and Practical Conference with International Participation], pp. 316-323.
27. Lavrinenko M.D., Shklenik M.A. Issledovaniye beskonechnolineynoy SMO s proverkoj kachestva obrabotki pri nestatsionarnom rezhime raboty [Study of an infinite-server queueing system with processing quality control under non-stationary operation mode]. *ITMM*, 2024, pp. 208-213.

Turenova Irina Alekseevna. *Candidate of Physical and Mathematical Sciences, Associate Professor at the Department of Probability Theory and Mathematical Statistics of the Institute of Applied Mathematics and Computer Science, National Research Tomsk State University. Research interests include mathematical modelling of economic systems, queueing theory, and mathematical marketing. The author has published works applying the method of limited decomposition to analyse request flows in systems with an unlimited number of servers and feedback and using the results to solve optimization problems in marketing services. AuthorID: 1120518, SPIN: 2771-6158, ORCID: 0000-0002-4373-9206, irenaturena@yandex.ru. 634050, Russia, Tomsk, Lenin Ave. 36.*

Moiseeva Svetlana Petrovna. *Doctor of Physical and Mathematical Sciences, Professor, Head of the Department of Probability Theory and Mathematical Statistics, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University. Research interests focus on queueing theory and its applications to the analysis of information, economic, and social systems. The author has developed original methods for studying queueing systems, including non-Markovian systems with an unlimited number of servers, with a particular emphasis on asymptotic methods. AuthorID: 9513, SPIN: 7526-2190, ORCID: 0000-0001-9285-1555, smoiseeva@mail.ru. 634050, Russia, Tomsk, Lenin Ave. 36.*

Shklennik Maria Alexandrovna. *Candidate of Physical and Mathematical Sciences, Associate Professor at the Department of Probability Theory and Mathematical Statistics, Institute of Applied Mathematics and Computer Science, National Research Tomsk State University. Research interests encompass multi-phase queueing models with instant feedback and methods for their analysis. The author has introduced a modified Markov summary method for studying repeated request flows in systems with feedback. AuthorID: 937206, SPIN: 1790-5249, ORCID: 0000-0003-0993-8006, shklennikm@mail.ru. 634050, Russia, Tomsk, Lenin Ave. 36.*

Статья поступила в редакцию 24.01.2026; одобрена после рецензирования 16.04.2026; принята к публикации 01.08.2026.

The article was submitted 01/24/2026; approved after reviewing 04/16/2026; accepted for publication 08/01/2026.

УДК: 636/639:004.942

DOI:10.25729/ESI.2026.43.3.007

Оптимизация численности сельскохозяйственных животных с учетом деградации пастбищ в Монголии

Иваньо Яровлав Михайлович, Энхбат Ариунжаргал

Иркутский государственный аграрный университет имени А.А. Ежевского, Россия,

Муниципальный район Иркутский, с. п. Молодежное, п. Молодежный, iymex@rambler.ru

Аннотация. Целью исследования является разработка оптимизационной модели для определения рациональной структуры поголовья различных видов сельскохозяйственных животных и их численности на определенной площади с учетом специализации хозяйства и экологической нагрузки на земельные ресурсы. Впервые построена модель, позволяющая определять структуру животных и их поголовье в условиях пастбищного животноводства при отсутствии и наличии деградации пастбищных земель, что актуально для сельского хозяйства Монголии. Целевая функция модели ориентирована на прибыль. Ограничения включают в себя объемы производимой продукции, обеспечение животных кормовыми и питательными веществами, трудозатраты на содержание животных, обеспечение водой, экологические требования к использованию животными пастбищ, численность разных групп. Поскольку обеспечение кормами сельскохозяйственных животных определяется урожайностью кормовых культур на пастбищах, проанализирована динамика объемов урожая кормовых культур на пастбищах разных регионов Монголии по данным 1960-2024 гг. с выделением трендов. Показано, что объемы урожая с 1992 года по настоящее время могут быть описаны с помощью экспоненциальной функции для Западного, Хангайского, Центрального и Восточного регионов Монголии. При реализации модели рассматривались домохозяйства с разной специализацией по отраслевому признаку. Проанализированы следующие виды животных: овцы, козы, лошади, крупный рогатый скот и верблюды. Выделено некоторое усредненное хозяйство по площади пастбища и трудозатратам на производство продукции. Определены производственно-экономические и экологические характеристики модели. Рассчитаны варианты оптимальных решений для разных по специализации домохозяйств с учетом и без учета деградации пастбищ. Проведен сравнительный анализ результатов, и получены наилучшие варианты по максимальной прибыли. Для реализации разработанной модели использовано приложение LP Solve. Разработанная модель полезна для планирования производства с разными специализациями по животноводческим отраслям и разным уровнем деградации пастбищ.

Ключевые слова: линейное программирование, оптимизация, пастбищное животноводство, деградация пастбищ, Монголия

Цитирование: Иваньо Я.М. Оптимизация численности сельскохозяйственных животных с учетом деградации пастбищ в Монголии / Я.М. Иваньо, А. Энхбат // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 83-92. – DOI:10.25729/ESI.2026.43.3.007.

Введение. Пастбищное животноводство, исторически сформировавшееся как кочевое, служит не только экономическим фундаментом, но и основой культурной идентичности, социальной структуры и продовольственной безопасности Монголии. Оно имеет свои особенности [1], обеспечивает существование значительной части населения и формирует экспортный потенциал страны. Однако в последние десятилетия отрасль сталкивается с серьезными вызовами, связанными с деградацией земель, снижением продуктивности пастбищ и опустыниванием за счет перевыпаса [2, 3]. Интенсификация животноводства в условиях рыночной экономики, климатические изменения, выражающиеся в учащении экстремальных погодных явлений, и социально-демографические сдвиги привели к стремительному росту поголовья скота. Рост поголовья скота, часто не согласованный с реальной кормовой емкостью угодий, усугубляет эти проблемы. Деградация экосистем, в свою очередь, ведет к снижению продуктивности животных, экономическим потерям хозяйств и росту рисков для средств к существованию населения, формируя порочный круг. В этих условиях критически важным становится применение научно обоснованных методов управления стадом и пастбищными ресурсами.

Насущной необходимостью является разработка и внедрение современных инструментов поддержки принятия решений, способных количественно оценивать последствия различных стратегий хозяйствования. Использование эколого-экономических математических моделей, интегрирующих биологические, производственные, экономические и экологические факторы в единую систему, позволяет перейти к адекватному планированию для повышения эффективности управления. Такой подход направлен на одновременное достижение двух стратегических целей: обеспечение стабильной экономической эффективности животноводческих хозяйств и сохранение экологической устойчивости пастбищных экосистем.

Целью исследования является разработка оптимизационной модели для определения рациональной структуры поголовья различных видов сельскохозяйственных животных и их численности на определенной площади с учетом специализации хозяйства и экологической нагрузки на земельные ресурсы.

Для достижения цели поставлены следующие задачи:

- 1) информационное обеспечение экстремальной модели, под которой понимается вариант оптимизационной модели, предназначенный для анализа структуры и численности поголовья сельскохозяйственных животных в условиях экстремально благоприятных и неблагоприятных природно-климатических событий с учетом вероятности их возникновения, заключающееся в сборе и систематизации данных по продуктивности животных, нормам потребления кормов и воды, трудозатратам, экономическим показателям и экологическим ограничениям для различных природно-климатических зон Монголии;
- 2) разработка модели оптимизации сельскохозяйственных животных на пастбище с целевой функцией максимизации прибыли и системой ограничений, включающей экологический показатель деградации земель;
- 3) реализация модели на основе данных хозяйств Монголии с учетом их специализации.

Материалы и методы. На основании справочных данных, в частности, «Нормы и рационы кормления сельскохозяйственных животных» [4], проанализированы материалы для определения потребностей животных в питательных веществах (кормовые единицы, сухое вещество, протеин, кальций). В исследовании использованы статистические данные, в том числе официальные отчеты и сведения по урожайности пастбищ в различных регионах Монголии [5], а также справочные данные по ценам на продукцию, затратам на корма и оплату труда, собранные из открытых источников и экспертных оценок, показатели деградации земель и коэффициенты негативного влияния выпаса [6, 7]. В табл. 1 приведены усредненные значения коэффициентов левых частей ограничений и целевой функции для моделирования численности сельскохозяйственных животных.

При построении модели использованы методы линейного программирования. В качестве коэффициентов при неизвестных целевой функции и ограничений использованы усредненные значения. Предложены разные варианты модели в зависимости от специализации хозяйства.

Основные результаты. Многие авторы для решения задач оптимизации производства продукции в разных отраслях сельского хозяйства используют задачи математического программирования разного вида: детерминированные, стохастические, интервальные, параметрические и другие [8-11] для решения управленческих задач [12, 13].

Из многих моделей выделим эколого-математическую модель, позволяющую оптимизировать производство сельскохозяйственной продукции сочетания отраслей с учетом загрязнения и эрозии почвы [14].

Модифицируем эту модель применительно к решению задачи оптимизации разных видов сельскохозяйственных животных при пастбищном животноводстве. Особенностью

такой модели является выделение видов животных по специализации хозяйства, учет деградации земель от чрезмерного выпаса, специфики климатических зон пастбищной территории.

Таблица 1. Данные для построения модели оптимизации численности сельскохозяйственных животных на пастбище применительно к усредненному домохозяйству Монголии

Виды животных	Овцы	Козы	Лошади	Крупный рогатый скот	Верблюды
Прибыль с 1 голов, руб.	2156,42	5207,91	58657,50	55581,20	77594,87
Цена мяса 1 кг, руб.	268	223	241	312	209
Цена шерсти 1 кг, руб.	23	3208			370
Цена молока 1 кг, руб.	100	148	300	119	500
Сено, ц	30	30	40	45	35
Сумма покупки корма (овес, ячмень, зеленый корм), ц	3,28	2,92	7,30	10,9	5,47
Расход воды за год, л	730	730	7300	10950	9130
Затраты труда, чел. ч	130	130	90	243,6	150
Площадь пастбища, га	1	0,9	7	5	6
Обеспечение потребности в питательных ц к.е.	4,38	4,38	16,9	20,4	25,1
Обеспечение потребности в сухом веществе, ц	5,48	5,48	24,1	24,8	31,8
Обеспечение потребности в протеине, ц	0,785	0,66	2,65	2,993	3,25
Обеспечение потребности в кальции, ц	0,026	0,022	0,072	0,14	0,22

Преобразованную модель можно записать следующим образом.

Максимизация общей прибыли от использования пастбищ для разных видов животных примет вид:

$$f = \sum_{k \in K} c_k y_k \rightarrow \max, \quad (1)$$

где c_k – прибыль от продажи продукции животных вида k (руб./гол), y_k – количество животных вида k , K – множество видов животных. Искомые переменными являются y_k – количество животных вида k .

Ограничение по обеспечению заданного объема производства животноводческой продукции запишем так:

$$r_k y_k \geq Y_k, \quad (2)$$

где r_k – продуктивность животных вида k (ц/гол), Y_k – заданный объем производства животноводческой продукции k (ц).

Для обеспечения потребности в кормовых и питательных веществах для животных можно использовать неравенство:

$$\sum_{j \in J} (1 - \alpha_j) \beta_{\theta j} \rho_j z \geq \sum_{k \in K} B_{\theta k} y_k \quad (\theta \in \Theta), \quad (3)$$

где $B_{\theta k}$ – минимальная потребность в питательном элементе θ для животного вида k (ц/гол.); $\beta_{\theta j}$ – содержание питательного элемента θ в единице кормовой продукции культуры j (ц. корм. ед.), Θ – множество питательных элементов, z – площадь пастбища, ρ_j – урожайность культуры j , α_j – коэффициент, характеризующий деградацию почвы на пастбище.

Необходимо соблюдать ограничение по трудозатратам:

$$\varepsilon_k y_k \leq E_k, \quad (4)$$

где ε_k – затраты трудовых ресурсов для ухода за животным (чел. ч/гол.); E_k – имеющиеся трудовые ресурсы для производства продукции вида k (чел. ч).

Ограничение по обеспечению водой животных имеет вид:

$$\sum_{k \in K} l_k y_k \leq TQ, \quad (5)$$

где l_k – объем воды, необходимый для одного животного вида k за год (м^3), T – продолжительность года (в днях), Q – расход воды от источника за год (м^3), y_k – количество животных вида k .

Для эффективного использования пастбищ:

$$\sum_{i \in j} \sum_{k \in K} (1 - \alpha_j) v_{jk} y_k \leq z,$$

где v_{jk} – норма использования площади пастбища с культурой j и животным вида k (га/гол).

Искомые переменные должны быть неотрицательными:

$$y_k \geq 0. \quad (7)$$

Модель, благодаря введению ограничений по количеству животных определенного вида, может быть использована для разных групп хозяйств по специализации ведения пастбищного животноводства, что позволяет задавать различные ограничения на количество животных определенных видов. Это способствует моделированию разных стратегий хозяйств (например, преимущественное разведение мелкого рогатого скота, комбинация содержания мелкого и крупного рогатого скота и др.).

$$y_k \leq U_k, \quad (8)$$

где U_k – заданная численность животных вида k .

Коэффициент деградации α_j является системообразующим параметром эколого-экономической модели. Его вариация позволяет количественно оценить последствия для поголовья животных, рынка продукции, логистики кормов и использования трудовых ресурсов. При значении $\alpha_j=0$ система находится в равновесном состоянии, и дефицит кормовой базы отсутствует. Последовательное увеличение коэффициента, который, согласно статистическим данным [2], может достигать максимального значения $\alpha_j=0,33$, позволяет моделировать возможности использования деградирующих пастбищ. Очевидно, что для разных культур этот коэффициент принимает разные значения. Поэтому приведенное значение является усредненной величиной.

Согласно принципам модели линейного программирования (1) – (7), рост α_j напрямую сокращает доступный ресурс в ограничении по кормам (3) и эффективному использованию площади (6). Это приводит к снижению оптимальной численности поголовья (y_k) для всех видов животных k . Причем наиболее значительное сокращение ожидается для видов с высокой нормой потребления кормов и площади (крупный рогатый скот, верблюды, лошади).

Снижение предложения животноводческой продукции на внутреннем рынке, обусловленное сокращением поголовья, при стабильном или растущем спросе создает условия для роста цен. Наиболее значительная ценовая динамика прогнозируется для мяса, как продукта первой необходимости. Цены на шерсть и кашемир, помимо внутреннего баланса спроса и предложения, демонстрируют высокую корреляцию с конъюнктурой мирового рынка, в частности, со спросом со стороны Китайской Народной Республики. Колебания цен могут достигать 50–100% в периоды после экстремальных погодных явлений, которые усугубляются деградированным состоянием пастбищ.

Анализ данных общего урожая кормовых культур на пастбищах разных регионов Монголии за 1960-2024 годы позволяет выявить несколько ключевых тенденций характеристики (рис. 1). Урожай снижался с 1960 по 1967 год по линейному тренду, затем увеличивался с 1968 по 1985 год, по логарифмической функции, после чего линейно убывал с 1986 по 1991 год (табл. 2). Однако, достигнув минимума в 1991 году, урожай с 1992 года по настоящее время возрастает по экспоненте. На рис. 1 показан точечный прогноз урожая до 2029 года по экспоненциальной функции при допущении инерционности процесса роста.

Таблица 2. Тренды общий объем урожая кормовых культур на пастбищах в Монголии по данным 1960-2024 гг.

Характеристика	Уравнение	R^2	F-критерий Фишер	Уровень значимости F-критерия, α	t-критерий Стьюдента
Урожай кормовых культур на пастбищах, т	Линейная (2)				
	$\rho_t = -109.0t + 1480$	0,80	24,05	0.0027	-4.9
	Логарифмическая (3)				
	$\rho_t = 546e^{0.0533t}$	0,92	172.4	5.5×10^{-10}	13.1г
	Линейная (4)				
	$\rho_t = -87.0t + 3670$	0,78	14,49	0,019	-3,8
	Экспоненциальная (5)				
	$\rho_t = 598e^{0.0287t}$	0.90	294,0	$2,3 \times 10^{-17}$	17,2

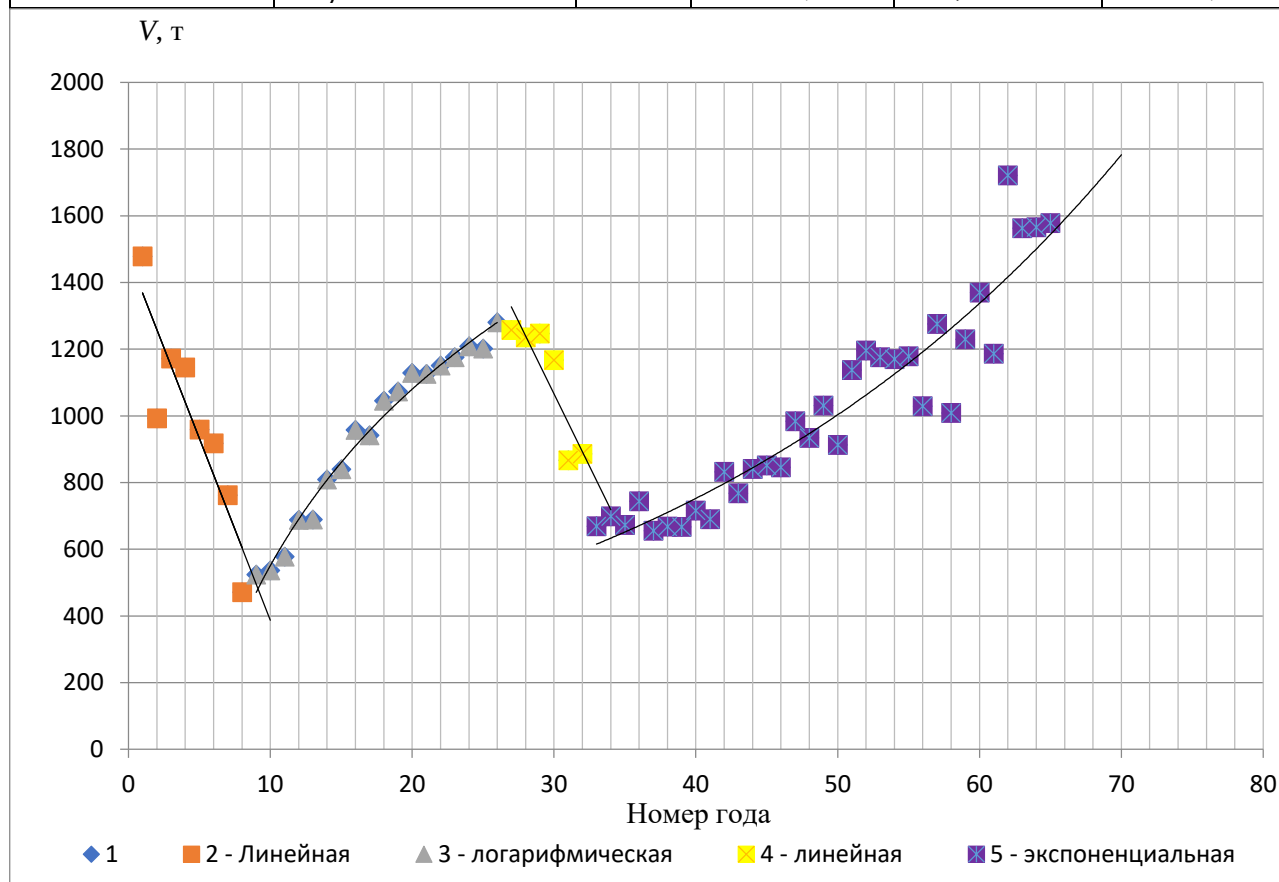


Рис. 1. Тренды общий объем урожая кормовых культур на пастбищах разных регионов Монголии по данным 1960-2024 гг. с использованием линейной (2) и (4), логарифмической (3) и экспоненциальной функции (5)

Полученные колебания урожая обобщены по четырем административным регионам Монголии: Западном, Хангайском, Центральном и Восточном. Динамика урожая каждого региона может быть описана экспоненциальным трендом с разной интенсивностью роста (табл. 3).

Обеспечение водопоя животных отражает ограничение (5). Пространственное распределение водных ресурсов неравномерно по территории Монголии. Регионы с использованием речных ресурсов (Хангайский, некоторые части Западного региона, некоторые части Восточного региона) характеризуются относительно развитой гидрографической сетью. Поверхностные водотоки (реки, ручьи) здесь выступают основным

источником для водопоя [15, 16]. В этих регионах также потенциально возможно применение простейших методов полива сенокосных участков.

Таблица 3. Объем урожая кормовых культур на пастбищах общий, Западном, Хангайском, Центральном и Восточном регионах с 1992 по 2024 год

Характеристика	Уровнение	R^2	F-критерий Фишер	Уровень значимости F-критерия, α	t-статистика Стьюдента
Урожай кормовых культур на пастбищах, т	Западный				
	$\rho_t = 122e^{0,0232t}$	0,80	133.6	9×10^{-13}	11.6
	Хангайский				
	$\rho_t = 178e^{0,0319t}$	0,85	174.8	2.8×10^{-14}	13.2
	Центральный				
	$\rho_t = 168e^{0,0291t}$	0.84	160.2	8.7×10^{-14}	12.7
	Восточный				
	$\rho_t = 97,0e^{0,0346t}$	0.84	158.7	9.9×10^{-14}	12.6

Регионы с дефицитом поверхностных вод (некоторые части Западного региона, некоторые части Восточного региона, Центральный) отличаются аридным и семиаридным климатом, где реки носят эпизодический или сезонный характер. Основным источником водоснабжения являются подземные воды, доступные через традиционные колодцы большой глубины и артезианские скважины. В критических ситуациях практикуется подвоз воды автоцистернами.

На основе модели (1) – (8) с определением коэффициентов целевой функции и левых частей ограничений, а также правых частей определены оптимальные решения для разных хозяйств по специализации с учетом деградации пастбищ (табл. 4).

Модель (1) – (8) реализована для некоторого усредненного хозяйства по площади пастбища.

Исходные данные для модели разделены на две группы. Первая группа включает в себя производственные показатели, такие, как выход мяса, молока, шерсти на 1 голову; нормы потребления сена, концентратов, воды; потребность в питательных веществах; затраты труда в чел./ч. Используются данные из нормативных справочников, адаптированные к условиям Монголии. Для урожайности пастбищ использованы данные за 1960-2024 гг. Вторая группа состоит из производственно-экономических показателей: оплата труда, цена реализации продукции, себестоимость кормов, прибыль, дополнительные корма.

Для решения задачи использован симплекс-метод. Для реализации модели использовано приложение LP Solve.

Данные табл. 4 показывают, как различные управленческие сценарии влияют на экономическую эффективность животноводства при деградации пастбищ и ее отсутствии.

Первый вариант без ограничения (8), позволил получить максимальную прибыль 13,97 млн руб. при сбалансированном количестве стада с преобладанием КРС (178 гол.) и значительном количестве мелкого рогатого скота (МРС).

Введение ограничений на виды животных, требующих больших пастбищных площадей или воды (верблюды, КРС, лошади), изменяют структуру стада и поголовье животных. Например, третий вариант приводит к резкому увеличению поголовья овец и коз (184 и 533 голов соответственно) и к снижению прибыли на 25% по сравнению с базовым сценарием. Это характеризует особенности ведения хозяйства с малым рогатым скотом и требуемым для этого ресурсам.

Таблица 4. Оптимальные планы производства животноводческой продукции (головы) и значения целевой функции (тыс. руб.) при отсутствии деградированных пастбищ и их наличии – $\alpha_j=0,30$ (отмечены *)

Специализация	Овцы	Козы	Лошади	КРС	Верблюды	Целевая функция, тыс. руб.
1. Скотоводство, козоводство, овцеводство	105	158	12	178	30	13965,458
	541	0	0	73	12	6177,815*
2. Козоводство, овцеводство, коневодство	176	195	103	53	10	11183,217
	531	36	11	56	10	5863,890*
3. Козоводство, овцеводство, скотоводство	184	533	50	70	6	10463,928
	250	299	14	70	2	6962,923*
4. Скотоводство, коневодство, верблюдоводство	34	16	45	167	34	14684,578
	0	95	25	104	33	10318,428*
5. Козоводство, скотоводство, овцеводство	80	218	30	175	6	13287,2250
	110	396	0	96	6	8074,677*

Четвертый вариант не только не снижает, но и превышает прибыль первого варианта (14,68 млн руб.). Модель, которая лишена возможности содержать большое количество лошадей, перераспределяет ресурсы в пользу более рентабельных в данных условиях видов, увеличивая долю высокопродуктивного КРС или МРС.

Пятый вариант имеет значения для хозяйств с небольшим числом верблюдов. Здесь упор делается на козоводство и скотоводство с небольшим поголовьем овец.

Согласно данным табл. 3, деградация пастбищ способствует значительным потерям прибыли и влияет на структуру видов животных и их количество. Сравнительные результаты показывают, что расхождения между прибылью от пастбищного животноводства для деградированных и недеградированных пастбищ с разными вариантами специализации изменяются от 30 до 56 %. При этом наилучшие результаты получены для четвертого и пятого вариантов, в том числе с учетом деградации пастбищ.

Следует отметить, что приведенные результаты получены для некоторого усредненного домохозяйства с площадью пастбища 1400 га и общими трудозатратами, не превышающими 90000 чел. ч. Обычно при расчетах число ограничений составляет от 14 до 17, а количество искомых переменных – 5. Информация для модели готовилась в течение шести месяцев. Поскольку задача небольшой размерности, она решается за считанные минуты. При корректном вводе ограничений оптимальное решение всегда определялось.

Практически во всех сценариях поголовье овец и коз остается высоким или даже увеличивается. Это подтверждает их ключевую роль, как наиболее адаптивного и устойчивого вида в условиях нестабильной ресурсной базы. Они выступают «буфером», который позволяет хозяйству оставаться рентабельным при ухудшении внешних условий.

Заключение. Разработана и реализована оптимизационная модель линейного программирования для управления пастбищным животноводством в Монголии с учетом особенностей характеристик модели. В качестве сельскохозяйственных животных рассмотрены овцы, козы, лошади, крупный рогатый скот и верблюды. Ключевым преимуществом разработанной модели является комплексный учет не только экономических

и производственных, но и экологических ограничений за счет нормированных характеристик и коэффициента деградации пастбища.

Разработанная модель позволяет получать разные варианты планирования с учетом специализации хозяйств и выявлять оптимальные структуры поголовья для каждого вида животных в усредненных условиях деятельности.

Полученные результаты могут быть использованы органами государственного управления и отдельными хозяйствами для разработки стратегий устойчивого развития животноводства, направленных на сохранение пастбищных экосистем Монголии при обеспечении экономической эффективности.

Список источников

1. Речкин С.В. Математическое моделирование процесса пастбищного животноводства при использовании цифровых технологий / С.В. Речкин, А.П. Сырбаков, В.А. Ротова // Вестник Алтайского государственного аграрного университета, 2024. – № 3 (233). – С. 98-103.
2. Министерство продовольствия, сельского хозяйства и легкой промышленности, Отчет об использовании пастбищ для животноводческого сектора – Улан-Батор: Монголии, 2025. – URL: <https://mofa.gov.mn/branch/maa/616f8fba73bc4a5fc70f21f3>
3. Монголия: Экологические вызовы и пути их решения / Под ред. Л. Энхбаяр. – Улан-Батор: Академия наук Монголии, 2022. – 210 с.
4. Нормы и рационы кормления сельскохозяйственных животных: Справочное пособие / А.П. Калашников, Н.И. Клейменов, В.Н. Баканов и др. – Москва: России, 2003. – 359 с.
5. Статистические данные Монголии. – URL: <https://www.1212.mn/mn> (дата обращения 22.11.2025)
6. Сводные результаты сельскохозяйственного сектора за 2022 год / Под ред. Б. Батдаваа. Улан-Батор, 2022. – 184 с.
7. Решения проблем пастбищ в Монголии / Под ред. А. Бакей. – Улан-Батор: ПРООН, 2019. – URL: https://www.undp.org/sites/g/files/zskgke326/files/migration/mn/Pastureland_Mongolia_Report.pdf (дата обращения 23.10.2025).
8. Математические и цифровые технологии оптимизации производства продовольственной продукции / Я.М. Иваньо, П.Г. Асалханов, М.Н. Барсукова и др.; под ред. Я.М. Иваньо. – Молодежный, 2021. – 219 с.
9. Юдин, Д.Б. Задачи и методы стохастического программирования/ Д.Б. Юдин. – Москва: Советское радио, 1979. – 392 с.:
10. Задачи линейной оптимизации с неточными данными /М. Фидлер., Й.Недома, Я. Рамик, И. Рон, К. Циммерманн. – М.: Институт компьютерных исследований 2008. – 288 с.
11. Барсукова М.Н. Параметрическая оптимизация получения продовольственной продукции с учетом особенностей характеристик временных рядов / М.Н. Барсукова, Я.М. Иваньо, В.В. Цыренжапова, и др // Лесной вестник. Forestry Bulletin, 2025. – Т. 29, № 3. – С. 181-193.
12. Гусейнов А.М. Автоматизированные системы управления пастбищами / А.М. Гусейнов, В.В. Климов // Вопросы агроэкологии, 2023. – № 2. – С. 67–74.
13. Ivanyo Ya. M., Petrova S. A. Comprehensive assessment of the development potential of plant production in the municipal district //Siberian Journal of Life Sciences and Agriculture, 2025, vol. 17, no. 2, pp. 535-552.
14. Иваньо Я. М. Эколого-математическое моделирование производства аграрной продукции: монография /Я.М. Иваньо, Е.А. Ковалева. – Молодежный, 2024. – 118 с.
15. Вода - Национальная программа – URL: <https://eghpp.mn/%D1%83%D1%81-%D2%AF%D0%BD%D0%B4%D1%8D%D1%81%D0%BD%D0%B8%D0%B9-%D1%85%D3%A9%D1%82%D3%A9%D0%BB%D0%B1%D3%A9%D1%80/> (дата обращения 25.11.2025).
16. Экологическо-экономические отчеты: отчет водного потока 2023 года / Под ред. Б. Батдаваа. – Улан-Батор: Монголии, 2024. – 22 с.

Иваньо Ярослав Михайлович. Доктор технических наук, профессор кафедры информатики и математического моделирования института экономики, управления и прикладной информатики. Иркутский государственный аграрный университет имени А.А. Ежевского. AuthorID: 106394, SPIN: 9654-8057, ORCID: 0000-0003-4118-7185, iymex@rambler.ru. 664038, Россия, Иркутская область, муниципальный район Иркутский, сельское поселение Молодежное, поселок Молодежный.

Энхбат Ариунжаргал. Аспирантка ФГБОУ ВО Иркутский государственный аграрный университет.
AuthorID: 1331322, SPIN: 2234-7964, 664038, Россия, Иркутская область, муниципальный район Иркутский, сельское поселение Молодежное, поселок Молодежный.

UDC: 636/639:004.942

DOI:10.25729/ESI.2026.43.3.007

Optimization of the number of farms livestock taking into account the degradation of pastelands in Mongolia

Yaroslav M. Ivanyo, Enkhbat Ariunjargal

Irkutsk State Agricultural University named after A.A. Ezhevsky,
Russia, Irkutsk region, Irkutsk municipal district, Molodezhnoye rural settlement, Molodezhny,
iymex@rambler.ru

Abstract. The aim of this study is to develop an optimization model for determining the rational livestock structure and population size for various livestock species in a given area, taking into account farm specialization and the environmental load on land resources. This is the first model developed that allows for determining the livestock structure and population size under conditions of pasture livestock farming, both with and without pasture land degradation, which is relevant for Mongolian agriculture. The model's objective function is profit-oriented. Constraints include production volumes, provision of animals with feed and nutrients, labor costs for animal maintenance, water supply, environmental requirements for pasture use by animals, and the population size of different groups. Since the supply of feed for livestock is determined by the yield of forage crops on pastures, the dynamics of forage crop yields on pastures in different regions of Mongolia are analyzed based on data from 1960 to 2024. It is shown that harvest volumes from 1992 to the present can be described using an exponential function for the Western, Khangai, Central, and Eastern regions of Mongolia. The model was implemented using households with different industry specializations. The following livestock species were analyzed: sheep, goats, horses, cattle, and camels. A certain average farm size was identified based on pasture area and labor costs for production. The production, economic, and environmental characteristics of the model were determined. Optimal solutions were calculated for households with different specializations, with and without pasture degradation. A comparative analysis of the results was conducted, and the best options for maximum profit were obtained. The LP Solve application was used to implement the developed model. The developed model is useful for production planning with different livestock specializations and varying levels of pasture degradation.

Keywords: linear programming, optimization, pasture livestock farming, pasture degradation, Mongolia

References

1. Rechkin S.V., Syrbakov A.P., Rotova V.A. Matematicheskoye modelirovaniye protsessa pastbishchnogo zhivotnovodstva pri ispol'zovanii tsifrovyykh tekhnologiy [Mathematical modeling of the process of pasture livestock farming using digital technologies]. Vestnik Altayskogo gosudarstvennogo agrarnogo universiteta [Bulletin of the Altai State Agrarian University], 2024, no. 3 (233), pp. 98-103.
2. Ministerstvo prodovol'stviya, sel'skogo khozyaystva i legkoy promyshlennosti, Otchet ob ispol'zovanii pastbishch dlya zhivotnovodcheskogo sektora [Ministry of Food, Agriculture and Light Industry, Report on Pasture Use for the Livestock Sector], Ulanbator, 2025. Available at: <https://mofa.gov.mn/branch/maa/616f8fba73bc4a5fc70f21f3> (accessed: 11/22/2025).
3. Mongoliya: Ekologicheskiye vyzovy i puti ikh resheniya [Mongolia: Environmental Challenges and Solutions]. Ed. by L. Enkhbayar. Ulaanbaatar, Mongolian Academy of Sciences Publ., 2022, 210 p.
4. Kalashnikov A.P., Kleymenov N.I., Bakanov V.N. et al. Normy i ratsiony kormleniya sel'skokhozyaystvennykh zhivotnykh: Spravochnoye posobiye [Feeding Standards and Rations for Farm Animals: A Reference Guide]. Moscow, Rossiya Publ., 2003, 359 p.
5. Statisticheskiye dannyye Mongolii [Statistical data of Mongolia]. Available at: <https://www.1212.mn/mn> (accessed: 11/23/2025).
6. Svodnyye rezul'taty sel'skokhozyaystvennogo sektora za 2022 god [Consolidated results of the agricultural sector for 2022]. Ed. by B. Batdavaa. Ulanbator, 2022, 184 p.

7. Resheniya problem pastbishch v Mongolii [Solutions to Pasture Problems in Mongolia]. Ed. by A. Bakey. Ulan-Bator: PROON, 2019. Available at: https://www.undp.org/sites/g/files/zskgke326/files/migration/mn/Pastureland_Mongolia_Report.pdf (accessed: 11/23/2025).
8. Ivanyo Ya. M., Asalkhanov P.G., Barsukova M.N. et al. Matematicheskiye i tsifrovyye tekhnologii optimizatsii proizvodstva prodovol'stvennoy produktsii: monografiya [Mathematical and Digital Technologies for Optimizing Food Production: Monograph]. Molodezhnyy, 2021, 219 p.
9. Yudin D.B. Zadachi i metody stokhasticheskogo programmirovaniya [Problems and methods of stochastic programming]. Moscow, Sovetskoye radio [Soviet Radio] Publ., 1979, 392 p.
10. Fiedler M., Nedoma J., Ramik J. et al. Zadachi lineynoy optimizatsii s netochnymi dannymi [Problems of linear optimization with imprecise data]. Moscow, Institute of Computer Science Publ., 2008, 288 p.
11. Barsukova M.N., Ivanyo Ya.M., Tsyrenzhapova V.V. et al. Parametricheskaya optimizatsiya polucheniya prodovol'stvennoy produktsii s uchetom osobennostey kharakteristik vremennykh ryadov [Parametric optimization of food production taking into account the features of time series characteristics]. Lesnoy vestnik. Forestry Bulletin, 2025, vol. 29, no. 3, pp. 181-193.
12. Guseynov A.M., Klimov V.V. Avtomatizirovannyye sistemy upravleniya pastbishchami [Automated pasture management systems]. Voprosy agroekologii [Issues of Agroecology], 2023, no. 2, pp. 67–74.
13. Ivanyo Ya. M., Petrova S. A. Comprehensive assessment of the development potential of plant production in the municipal district //Siberian Journal of Life Sciences and Agriculture, 2025, vol. 17, no. 2, pp. 535-552.
14. Ivanyo Ya. M., Kovaleva Ye.A. Ekologo-matematicheskoye modelirovaniye proizvodstva agrarnoy produktsii: monografiya [Ecological and mathematical modeling of agricultural production: monograph]. Molodezhny, 2024, 118 p.
15. Voda – Natsional'naya programma [Water – National Program]. Available at: <https://eghpp.mn/%D1%83%D1%81-%D2%AF%D0%BD%D0%B4%D1%8D%D1%81%D0%BD%D0%B8%D0%B9-%D1%85%D3%A9%D1%82%D3%A9%D0%BB%D0%B1%D3%A9%D1%80/> (accessed: 11/25/2025).
16. Ekologicheskoye-ekonomicheskoye otchety: otchet vodnogo potoka 2023 goda [Environmental and economic reports: water flow report 2023]. Ed. by B. Batdavaa.. Ulanbator, 2024, 22 p.

Yaroslav M. Ivanyo. Doctor of technical sciences, professor of the department of informatics and mathematical modeling of the institute of economics, management and applied informatics. Irkutsk state agriculture university named after A.A. Ezhevsky. AuthorID: 106394, SPIN: 9654-8057, ORCID: 0000-0003-4118-7185, iymex@rambler.ru. 664038, Russia, Irkutsk region, Irkutsk municipal district, Molodezhnoye rural settlement, Molodezhny.

Enkhbat Ariunjargal. PhD student, FSBEI HE Irkutsk SAU. eariuka329@gmail.com. AuthorID: 1331322, SPIN: 2234-7964, 664038, Russia, Irkutsk region, Irkutsk municipal district, Molodezhnoye rural settlement, Molodezhny.

Статья поступила в редакцию 23.12.2025; одобрена после рецензирования 18.01.2026; принята к публикации 02.08.2026.

The article was submitted 12/23/2025; approved after reviewing 01/18/2026; accepted for publication 08/02/2026.

Математические, информационные и интеллектуальные технологии в энергетике

УДК 621.396

DOI:10.25729/ESI.2026.43.3.008

Метод идентификации дефектов в технических системах, описываемых линейными моделями

Жирабок Алексей Нилович¹, Зуев Александр Валерьевич²

¹Дальневосточный федеральный университет, Россия, Владивосток, zhirabok@mail.ru

²Институт проблем морских технологий ДВО РАН, Россия, Владивосток

Аннотация. Рассматривается задача идентификации (оценивания) величин дефектов в динамике и датчиках технических систем, описываемых линейными динамическими моделями с постоянными параметрами, представленными системой обыкновенных дифференциальных уравнений, с учетом действия неконтролируемых внешних возмущений и шумов измерений в имеющихся датчиках. Цель работы состоит в построении средств идентификации, использующих данные об управляющих воздействиях и показания датчиков системы, для определения величин дефектов как в динамике, так и датчиках. Полученная информация о величинах дефектов необходима для компенсации влияния этих дефектов системой отказоустойчивого управления за счет коррекции управляющих воздействий. Решение задачи основано на так называемых интервальных наблюдателях минимальной размерности, нечувствительных к внешним возмущениям. Такие наблюдатели строятся на базе редуцированной (имеющую меньшую размерность) модели исходной системы, также нечувствительной к внешним возмущениям. Для построения редуцированной модели получены соответствующие соотношения с использованием жордановой канонической формы матрицы, описывающей динамику модели и интервального наблюдателя. Жорданова форма необходима для обеспечения соответствующих свойств наблюдателя. Далее на основе интервального наблюдателя и дифференциатора, формирующего точную оценку производной за конечное время, выводятся соотношения, позволяющие получить интервальную оценку величин дефектов в динамике и датчиках системы. Новизна работы состоит в том, что, в отличие от классического способа использования интервальных наблюдателей средства идентификации в настоящей работе строятся на основе редуцированных моделей, нечувствительных к внешним возмущениям. Это позволяет уменьшить вычислительную сложность процедуры идентификации и за счет нечувствительности к внешним возмущениям повысить точность оценивания величин дефектов. Теоретические положения проиллюстрированы практическим примером оценки величины дефектов в электроприводе манипуляционного робота. Проведенное моделирование на основе пакета Matlab продемонстрировало корректность теоретических результатов. Показано, что использование редуцированной модели электропривода позволило уменьшить вычислительную сложность процедуры идентификации. Полученные соотношения могут быть использованы для построения отказоустойчивых систем.

Ключевые слова: линейные системы, модели, дефекты, интервальные наблюдатели

Цитирование: Жирабок А.Н. Метод идентификации дефектов в технических системах, описываемых линейными моделями / А.Н. Жирабок, А.В. Зуев // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 93-102. – DOI:10.25729/ESI.2026.43.3.008.

Введение. Известно, что в технических системах в процессе их эксплуатации могут возникать дефекты, для компенсации их влияния применяются различные методы, многие из которых используют информацию о величинах этих дефектов. Непосредственное измерение таких величин обычно наталкивается на значительные трудности. В этих случаях задача может быть решена на основе известных математических моделей этих систем и штатных датчиков, показания которых после соответствующей обработки поставляют искомую информацию.

Существуют различные способы обработки этой информации, среди которых наиболее известными служат подходы на основе скользящих наблюдателей [1-7], разработанные для различных классов систем и дефектов как в динамике, так и датчиках. При этом для

построения скользящих наблюдателей на исходную систему накладываются различные ограничения, в частности, требуется, чтобы она была минимально фазовой или детектируемой, кроме того, должно выполняться условие согласования [1]. Недостатком таких наблюдателей являются их высокая чувствительность к различным помехам, сопровождающим работу системы.

Интересный метод для решения рассматриваемой задачи был предложен в [8] на основе так называемых интервальных наблюдателей, которые формируют интервальные оценки рассматриваемых функций. За последние 20 лет было предложено много методов построения таких наблюдателей для различных классов систем [9-21]. Следует отметить, что в [8] решение задачи оценивания получено за счет наложения требования минимальной фазовости исходной системы, кроме того, там не учитываются шумы измерений, которые всегда присутствуют в реальных системах.

В настоящей работе решается задача оценивания величин дефектов в динамике и датчиках на основе интервальных наблюдателей в технических системах, описываемых линейными динамическими моделями, без требования минимальной фазовости при наличии внешних возмущений и шумов измерений. Как и в [15-19], решение опирается на редуцированную (имеющую меньшую размерность) модель исходной системы; это повышает точность оценивания, уменьшает вычислительную сложность и дает возможность учесть шумы измерений.

1. Основные положения. Рассматриваются системы, описываемые уравнениями

$$\begin{aligned}\dot{x}(t) &= Fx(t) + Gu(t) + Dd(t) + Lp(t), \\ y(t) &= Hx(t) + D_s d_s(t) + w(t),\end{aligned}\tag{1}$$

где $x(t) \in R^n$, $y(t) \in R^l$ и $u(t) \in R^m$ – векторы состояния, выхода и управления, F , G , H , D , L и D_s – постоянные матрицы соответствующих размеров; скалярные функции $d(t)$ и $d_s(t)$ описывают дефекты в динамике и датчиках соответственно и полагаются неизвестными; неизвестная функция $w(t) \in R^l$ описывает шумы измерений, предполагается, что $\underline{w} \leq w(t) \leq \bar{w}$ для известных $\underline{w}$, $\bar{w}$, где отношение « $\leq$ » понимается поэлементно; $p(t) \in R^s$ – неизвестная функция времени, описывающая действующее на систему возмущение. Предполагается также, что начальное состояние $x(0)$ удовлетворяет условию $\underline{x}(0) \leq x(0) \leq \bar{x}(0)$ для известных $\underline{x}(0)$ и $\bar{x}(0)$.

Отметим, что матрица F описывает динамику системы, G – характер управляющих воздействий, H – размещение датчиков; матрицы D и L характеризуют места в динамике, подверженные дефектам и внешним возмущениям, D_s определяет датчики, в которых возможны дефекты.

Рассмотрим вначале дефекты в динамике, предполагая, что $D_s = 0$. По аналогии с [5] будем предполагать, что выполняется условие согласования $\text{rank}(HD) = \text{rank}(D)$. Решение задачи базируется на редуцированной (имеющей меньшую размерность) модели исходной системы (1), нечувствительной к возмущениям и описываемой уравнениями

$$\begin{aligned}\dot{x}_*(t) &= F_*x_*(t) + J_*Hx(t) + G_*u(t) + D_*d(t), \\ y_*(t) &= H_*x_*(t),\end{aligned}\tag{2}$$

где $x_* \in R^k$ – вектор состояния модели, k – ее размерность, F_* , J_* , G_* , D_* и H_* – матрицы соответствующих размеров, подлежащие определению.

Для решения задачи будем полагать, что существуют матрицы Φ и R , которые связывают векторы состояния и выхода системы (1) и модели (2) и удовлетворяют уравнениям [15-17]

$$RH = H_*\Phi, \quad \Phi F = F_*\Phi + J_*H, \quad G_* = \Phi G, \quad D_* = \Phi D, \quad \Phi L = 0. \quad (3)$$

2. Построение редуцированной модели. Известно [9-11], что для корректной работы интервального наблюдателя матрица F_* должна быть устойчивой и метцлеровой, т.е. иметь неотрицательные внедиагональные элементы. Этим условиям соответствует диагональная жорданова форма матрицы, когда

$$F_* = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & \lambda_k \end{pmatrix}, \quad (4)$$

где $\lambda_1, \dots, \lambda_k$ – собственные числа, которые должны быть отрицательными. Матрицы H_* и R определяются из первого уравнения в (3), записанного в виде

$$(R \quad -H_*) \begin{pmatrix} H \\ \Phi \end{pmatrix} = 0, \quad (5)$$

условием разрешимости которого является неравенство

$$\text{rank} \begin{pmatrix} \Phi \\ H \end{pmatrix} < \text{rank}(\Phi) + \text{rank}(H). \quad (6)$$

Известно [15-17], что для матрицы (4) второе уравнение в (3) для каждой строки матрицы Φ принимает вид

$$(Q \quad -J_*) \begin{pmatrix} L_0(F - \lambda I_n) \\ H \end{pmatrix} = 0, \quad (7)$$

где L_0 – матрица максимального ранга, удовлетворяющая условию $L_0 L = 0$, I_n – единичная матрица. Задавая конкретное значение $\lambda < 0$, из уравнения (7) следует найти матрицы Q и J_* такие, что $\Phi = QL_0$ удовлетворяет условию (6). Если оно выполняется, определяются матрицы H_* и R из алгебраического уравнения (5), G_* и D_* – из (3). В противном случае задача рассмотренным методом решения не имеет.

3. Точная оценка. Для выбранного значения $\lambda < 0$ модель (2) является одномерной и принимает вид

$$\begin{aligned} \dot{x}_*(t) &= \lambda x_*(t) + J_* H x(t) + G_* u(t) + D_* d(t), \\ y_*(t) &= x_*(t), \end{aligned} \quad (8)$$

D_* представляет собой скаляр. Если шумы измерения отсутствуют или ими можно пренебречь, то с учетом равенств $y_*(t) = x_*(t) = Ry(t)$ и $J_* H y(t) = J_* y(t)$ величина дефекта $d(t)$ может быть оценена точно в виде

$$d(t) = D_*^{-1} (\dot{y}_*(t) - \lambda y_*(t) - J_* y(t) - G_* u(t)). \quad (9)$$

Для вычисления производной $\dot{y}_*(t)$ предлагается использовать дифференциатор, предложенный в [22]:

$$\begin{aligned} \dot{\vartheta}_1(t) &= \vartheta_0(t), \\ \vartheta_0(t) &= \vartheta_2(t) - \beta_1 |\vartheta_1(t) - y_*(t)|^{1/2} \text{sign}(\vartheta_1(t) - y_*(t)), \\ \dot{\vartheta}_2(t) &= -\beta_2 \text{sign}(\vartheta_1(t) - y_*(t)), \end{aligned} \quad (10)$$

где β_1, β_2 – положительные константы, выбираемые согласно рекомендациям [22]. Как показано в [22], при отсутствии возмущений в $y_*(t)$ переменная $\vartheta_2(t)$ дает точную оценку производной $\dot{y}_*(t)$ за конечное время.

4. Интервальная оценка. Наличие шумов измерений делает оценку (9) приближенной. В этом случае использование интервального наблюдателя позволяет найти гарантированные пределы (границы), в которых будет находиться оценка по аналогии с доверительным интервалом в математической статистике.

Интервальный наблюдатель строится на основе модели (2) и по аналогии с [9-11] принимает вид

$$\begin{aligned}\dot{\underline{x}}_*(t) &= \lambda \underline{x}_*(t) + J_* y(t) + G_* u(t) + J_*^- \underline{w} - J_*^+ \bar{w}, \\ \dot{\bar{x}}_*(t) &= \lambda \bar{x}_*(t) + J_* y(t) + G_* u(t) + J_*^- \bar{w} - J_*^+ \underline{w}, \\ \underline{y}_*(t) &= \underline{x}_*(t), \\ \bar{y}_*(t) &= \bar{x}_*(t).\end{aligned}\tag{11}$$

Здесь $J_*^+ = \max(0, J_*)$, $J_*^- = J_*^+ - J_*$, нетрудно видеть, что $J_*^+ \geq 0$, $J_*^- \geq 0$; $\underline{x}_*(t)$ и $\bar{x}_*(t)$ – компоненты состояния интервального наблюдателя.

Известно [9-11], что если матрица F_* является метцлеровой, то из условия $\underline{x}(0) \leq x(0) \leq \bar{x}(0)$ следует $\underline{x}_*(t) \leq x_*(t) \leq \bar{x}_*(t)$ при всех $t \geq 0$. Так как $\underline{y}_*(t) = \underline{x}_*(t)$ и $\bar{y}_*(t) = \bar{x}_*(t)$, то по аналогии с [8] нетрудно заключить, что из $\underline{y}_*(t) \leq y_*(t) \leq \bar{y}_*(t)$ следует существование переменной $0 \leq \gamma(t) \leq 1$, такой, что

$$y_*(t) = \underline{y}_*(t) + \gamma(t)(\bar{y}_*(t) - \underline{y}_*(t)) = \bar{y}_*(t) + (\gamma(t) - 1)(\bar{y}_*(t) - \underline{y}_*(t))\tag{12}$$

для всех $t \geq 0$. Заменим $y_*(t)$ на $Ry(t)$, что дает

$$\gamma(t) = \frac{Ry(t) - \underline{y}_*(t)}{\bar{y}_*(t) - \underline{y}_*(t)}.$$

Для получения верхней границы $\bar{d}(t)$ вычислим $\dot{y}_*(t)$ на основе первой части равенства (12) и уравнения (11):

$$\begin{aligned}\dot{y}_*(t) &= \dot{\underline{y}}_*(t) + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \gamma(t) \frac{d(\bar{y}_*(t) - \underline{y}_*(t))}{dt} = \\ &= \lambda \underline{y}_*(t) + J_* y(t) + G_* u(t) + J_*^- \underline{w} - J_*^+ \bar{w} + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &+ \gamma(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \gamma(t) |J_*| (\bar{w} - \underline{w}),\end{aligned}$$

где $|J_*|$ – матрица, составленная из абсолютных значений элементов матрицы J_* . Заменим в полученном выражении переменную $\dot{y}_*(t)$ правой частью равенства (8) с $x_*(t) = y_*(t)$:

$$\begin{aligned}\lambda y_*(t) + J_* Hx(t) + G_* u(t) + D_* d(t) &= \\ &= \lambda \underline{y}_*(t) + J_* y(t) + G_* u(t) + J_*^- \underline{w} - J_*^+ \bar{w} + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &+ \gamma(t)(\bar{y}_*(t) - \underline{y}_*(t)) + |J_*| (\bar{w} - \underline{w})\end{aligned}$$

и преобразуем его с учетом $y_*(t) = Ry(t)$ с учетом того, что $J_* y(t) - J_* Hx(t) = J_* w(t)$:

$$\begin{aligned}D_* d(t) &= \lambda(\underline{y}_*(t) - Ry(t)) + J_* w(t) + J_*^- \underline{w} - J_*^+ \bar{w} + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &+ \gamma(t)(\bar{y}_*(t) - \underline{y}_*(t)) + |J_*| (\bar{w} - \underline{w}).\end{aligned}$$

Так как

$$\begin{aligned} J_* w(t) + J_*^- \underline{w} - J_*^+ \bar{w} &= J_*^+ w(t) - J_*^- w(t) + J_*^- \underline{w} - J_*^+ \bar{w} = \\ &= J_*^+ (w(t) - \bar{w}) + J_*^- (\underline{w} - w(t)) \leq 0, \end{aligned}$$

то, удаляя эти слагаемые из предыдущего выражения, получаем искомую формулу для $\bar{d}(t)$:

$$\begin{aligned} \bar{d}(t) &= D_*^{-1}(\lambda(\underline{y}_*(t) - Ry(t)) + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &+ \gamma(t)(\bar{y}_*(t) - \underline{y}_*(t) + |J_*|(\bar{w} - \underline{w}))). \end{aligned}$$

Выполняя аналогичные операции, получаем выражение для нижней границы:

$$\begin{aligned} \underline{d}(t) &= D_*^{-1}(\lambda(\underline{y}_*(t) - Ry(t)) + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &+ (\gamma(t) - 1)(\bar{y}_*(t) - \underline{y}_*(t) + |J_*|(\bar{w} - \underline{w}))). \end{aligned}$$

Нетрудно видеть, что $\underline{d}(t) \leq d(t) \leq \bar{d}(t)$. Для вычисления производной $\dot{\gamma}(t)$ может быть использован дифференциатор вида (10).

5. Оценка дефекта в датчике. При рассмотрении дефекта в j -м датчике уравнение (7) принимает вид

$$(Q - J_*) \begin{pmatrix} L_0(F - \lambda I_n) \\ H^{(j)} \end{pmatrix} = 0, \quad (13)$$

где $H^{(j)}$ – матрица H без j -й строки, поскольку дефектный датчик не должен участвовать в построении редуцированной модели, которая в предположении $D = 0$ принимает вид:

$$\begin{aligned} \dot{x}_*(t) &= F_* x_*(t) + J_* H^{(j)} x(t) + G_* u(t), \\ y_*(t) &= H_* x_*(t) + RD_s d_s(t). \end{aligned}$$

На ее основе строится диагностический наблюдатель с вектором состояния $x_o(t)$:

$$\begin{aligned} \dot{x}_o(t) &= F_* x_o(t) + J_* y(t) + G_* u(t), \\ y_o(t) &= H_* x_o(t). \end{aligned} \quad (14)$$

Введем ошибку оценивания $e(t) = \Phi x(t) - x_o(t)$ и рассмотрим переменную $y_o(t)$:

$$\begin{aligned} y_o(t) &= H_* x_o(t) = H_*(\Phi x(t) - e(t)) = H_* \Phi x(t) - H_* e(t) = \\ &= RHx(t) - H_* e(t) = R(y(t) - w(t) - D_s d_s(t)) - H_* e(t), \end{aligned}$$

откуда при условии $RD_s \neq 0$ следует

$$d_s(t) = (RD_s)^{-1}(Ry(t) - y_o(t) - Rw(t) - H_* e(t)). \quad (15)$$

Так как $d_s(t)$ – скаляр, то $RD_s \neq 0$ также скаляр и $(RD_s)^{-1}$ всегда существует.

Если шумы измерений отсутствуют или ими можно пренебречь, из (15) может быть получена точная оценка функции $d_s(t)$. Действительно, из устойчивости наблюдателя следует $e(t) \rightarrow 0$, а тогда из (15) при $w(t) = 0$ получаем

$$d_s(t) = (RD_s)^{-1}(Ry(t) - y_o(t)).$$

При наличии шумов измерений отношение $e(t) \rightarrow 0$ не выполняется, и приведенная оценка становится неточной. В этом случае можно получить интервальную оценку. Для этого введем нижнюю и верхнюю границы ошибки $e(t)$:

$$\underline{e}(t) = \underline{x}_*(t) - x_o(t), \quad \bar{e}(t) = \bar{x}_*(t) - x_o(t). \quad (16)$$

где переменные $\underline{x}_*(t)$ и $\bar{x}_*(t)$ формируются интервальным наблюдателем (11).

Известно [9-11], что из условия $\underline{x}(0) \leq x(0) \leq \bar{x}(0)$ следует $\underline{e}(t) \geq 0$ и $\bar{e}(t) \geq 0$ при всех $t \geq 0$. С учетом (15) и (16) выражения для нижней и верхней границ оценки величины дефекта, полученные на основе (15), принимают вид

$$\begin{aligned} \underline{d}_s(t) &= (RD_s)^{-1}(Ry(t) - y_o(t) - ((RD_s)^{-1}R)^+ \bar{w} + ((RD_s)^{-1}R)^- \underline{w} - \\ &\quad - ((RD_s)^{-1}H_*)^+ \bar{e}(t) + ((RD_s)^{-1}H_*)^- \underline{e}(t)), \\ \bar{d}_s(t) &= (RD_s)^{-1}(Ry(t) - y_o(t) - ((RD_s)^{-1}R)^+ \underline{w} + ((RD_s)^{-1}R)^- \bar{w} - \\ &\quad - ((RD_s)^{-1}H_*)^+ \underline{e}(t) + ((RD_s)^{-1}H_*)^- \bar{e}(t)). \end{aligned}$$

Из сказанного ясно, что для получения интервальной оценки должны использоваться наблюдатели (11) и (14). По аналогии с [15-17] можно показать, что $\underline{d}_s(t) \leq d_s(t) \leq \bar{d}_s(t)$.

Пример. Рассмотрим электропривод робота, описанный уравнениями

$$\begin{aligned} \dot{x}_1(t) &= k_1 x_2(t), \\ \dot{x}_2(t) &= k_2 x_2(t) + k_3 x_3(t) + \rho(t), \\ \dot{x}_3(t) &= k_4 x_2(t) + k_5 x_3(t) + k_6 u(t) + d(t), \\ y_1(t) &= x_1(t) + w_1(t), \\ y_2(t) &= x_3(t) + d_s(t) + w_2(t), \end{aligned}$$

где переменные $x_1(t)$, $x_2(t)$ и $x_3(t)$ – это скорость и угол вращения вала двигателя и ток якоря соответственно; коэффициенты $k_1 - k_6$ представляют параметры электропривода; возмущение $\rho(t)$ обусловлено наличием внешнего нагрузочного момента, приведенного к валу двигателя; дефект $d(t)$ вызван изменением активного сопротивления обмотки двигателя из-за его перегрева. Рассматриваемый электропривод описывается матрицами:

$$F = \begin{pmatrix} 0 & k_1 & 0 \\ 0 & k_2 & k_3 \\ 0 & k_4 & k_5 \end{pmatrix}, \quad G = \begin{pmatrix} 0 \\ 0 \\ k_5 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad L = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad D_s = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Нетрудно видеть, что $L_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, тогда уравнение (7) принимает вид

$$(Q \quad -J_*) \begin{pmatrix} -\lambda & k_1 & 0 \\ 0 & k_4 & k_5 - \lambda \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} = 0.$$

При $\lambda = k_5 < 0$ оно имеет решение $Q = (k_4 \quad -k_1)$, $J_* = (-k_4 k_5 \quad 0)$, откуда $\Phi = (k_4 \quad 0 \quad -k_1)$; условие (6) выполняется и тогда $R = (k_4 \quad -k_1)$, $H_* = 1$ и $G_* = -k_1 k_6$, $D_* = -k_1$. Модель (2) принимает вид

$$\dot{x}_*(t) = k_5 x_*(t) - k_4 k_5 H_1 x(t) - k_1 k_6 u(t) - k_1 d(t).$$

Так как $k_4 < 0$, то $J_*^+ = 0$ и $J_*^- = (|k_4 k_5| \quad 0)$; интервальный наблюдатель описывается уравнениями

$$\begin{aligned} \dot{\underline{y}}_*(t) &= k_5 \underline{y}_*(t) - k_4 k_5 y_1(t) - k_1 k_6 u(t) + |k_4 k_5| \underline{w}_1, \\ \dot{\bar{y}}_*(t) &= k_5 \bar{y}_*(t) - k_4 k_5 y_1(t) - k_1 k_6 u(t) + |k_4 k_5| \bar{w}_1. \end{aligned} \tag{17}$$

Выражения для верхней и нижней границ принимают вид

$$\begin{aligned} \bar{d}(t) &= -(1/k_1)(k_5(\bar{y}_*(t) - Ry(t)) + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &\quad + k_5 \gamma(t)(\bar{y}_*(t) - \underline{y}_*(t) + |k_4 k_5|(\bar{w}_1 - \underline{w}_1))), \\ \underline{d}(t) &= -(1/k_1)(k_5(\underline{y}_*(t) - Ry(t)) + \dot{\gamma}(t)(\bar{y}_*(t) - \underline{y}_*(t)) + \\ &\quad + k_5(\gamma(t) - 1)(\bar{y}_*(t) - \underline{y}_*(t) + |k_4 k_5|(\bar{w}_1 - \underline{w}_1))), \end{aligned}$$

где $Ry(t) = k_4 y_1(t) - k_1 y_2(t)$, $\gamma(t) = \frac{k_4 y_1(t) - k_1 y_2(t) - \underline{y}_*(t)}{\bar{y}_*(t) - \underline{y}_*(t)}$.

Для дефекта во втором датчике $H^{(2)} = (1 \ 0 \ 0)$, уравнение (13) принимает вид

$$(Q - J_*) \begin{pmatrix} -\lambda & k_1 & 0 \\ 0 & k_4 & k_5 - \lambda \\ 1 & 0 & 0 \end{pmatrix} = 0.$$

При $\lambda = k_5 < 0$ оно имеет решение и модель, совпадающие с полученными выше, на основе чего строится диагностический наблюдатель (14):

$$\begin{aligned} \dot{x}_o(t) &= k_5 x_o(t) - k_4 k_5 y_1(t) - k_1 k_6 u(t), \\ y_o(t) &= x_o(t). \end{aligned}$$

Так как $RD_s = -k_1$, то $(RD_s)^{-1}R = (k_4/k_1 \ -1)$ и $((RD_s)^{-1}H_*) = -1/k_1$; поскольку $k_1 > 0$ и $k_4 < 0$, то $((RD_s)^{-1}R)^+ = 0$, $((RD_s)^{-1}R)^- = (|k_4|/k_1 \ 1)$, $((RD_s)^{-1}H_*)^+ = 0$ и $((RD_s)^{-1}H_*)^- = 1/k_1$. Приведем выражения для нижней и верхней границ оценки величины дефекта:

$$\begin{aligned} \underline{d}_s(t) &= -(1/k_1)(k_4 y_1 - k_1 y_2 - y_o(t) + (|k_4|/k_1)\underline{w}_1 + \underline{w}_2) + \underline{e}(t), \\ \bar{d}_s(t) &= -(1/k_1)(k_4 y_1 - k_1 y_2 - y_o(t) + (|k_4|/k_1)\bar{w}_1 + \bar{w}_2) + \bar{e}(t), \end{aligned}$$

где $\underline{e}(t) = \underline{y}_*(t) - y_o(t)$, $\bar{e}(t) = \bar{y}_*(t) - y_o(t)$.

Отметим, что, так как исходная система трехмерна, то наблюдатель, построенный по классической схеме, имел бы размерность 6, наблюдатель (17) – размерность 2, что предопределяет меньшую вычислительную сложность процедуры идентификации.

Для моделирования приняты дискретизованные уравнения электропривода с $k_1 = k_4 = k_6 = 1$, $k_2 = -0,1$, $k_3 = -0,4$, $k_5 = -0,5$ и $u(t) = 0,5 \sin(t/2)$. Шумы измерений $w_1(t)$ и $w_2(t)$ представлены независимыми случайными величинами, равномерно распределенными на интервале $(-0,01; 0,01)$; дефект в датчике имитируется следующим образом: $d_s(t) = 0$ при $t < 30$ и $t > 60$, $d_s(t) = 0,5$ при $60 \geq t \geq 30$. Результаты моделирования приведены на рисунке 1, где показано поведение функции $d_s(t)$, а также ее верхних и нижних границ.

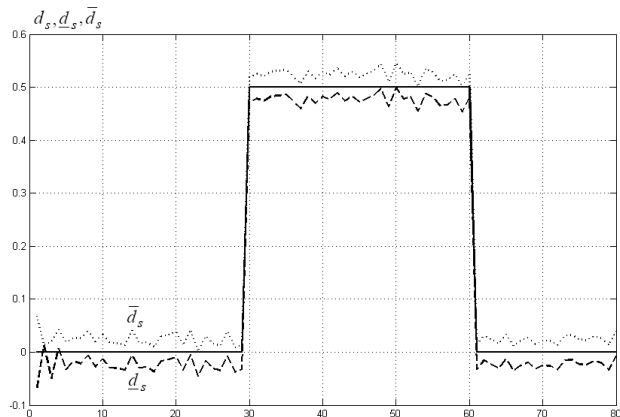


Рис.1. Графики функции $d_s(t)$, верхней $\bar{d}_s(t)$ и нижней $\underline{d}_s(t)$ границ

Заключение. В работе решена задача оценивания (идентификации) величин дефектов в системах, описываемых линейными моделями в присутствии внешних возмущений и шумов измерений. В отличие от известных методов решения этой задачи с использованием

скользящих наблюдателей, разработанный в статье подход позволяет расширить класс систем, для которых идентификация может быть осуществлена. Это объясняется тем, что метод построения скользящих наблюдателей накладывает ограничения на системы, для которых требуется решить задачу идентификации дефектов. Кроме того, по сравнению с классическими методами построения интервальных наблюдателей предложенный подход позволяет уменьшить вычислительную сложность процедуры идентификации за счет использования редуцированной модели исходной системы. Так, в примере интервальный наблюдатель имеет размерность 2, а построенный по классической схеме – 6. Предложенный метод может быть использован для построения отказоустойчивых систем путем компенсации оцененной величины дефекта системой управления.

Благодарности. Работа выполнена при финансовой поддержке Российского научного фонда (проект № 25-19-00590), <https://rscf.ru/project/25-19-00590/>.

Список источников

1. Edwards C., Spurgeon S., Patton R. Sliding mode observers for fault detection and isolation. *Automatica*, 2000, vol. 36, pp. 541–553, DOI: 10.1016/S0005-1098(99)00177-6.
2. Chen L., Edwards C., Alwi H. et al. Flight evaluation of a sliding mode online control allocation scheme for fault tolerant control. *Automatica*, 2020, vol. 114, p. 108829, DOI: 10.1016/j.automatica.2020.108829.
3. Kalsi K., Hui S., Zak S. Unknown input and sensor fault estimation using sliding-mode observers. *Proceedings of 2011 ACC, San Francisco*, 2011, pp. 1364-1369.
4. Yan X., Edwards C. Sensor fault detection and isolation for nonlinear systems based on a sliding mode observer. *International Journal of Adaptive Control and Signal Processing*, 2007, vol. 21, pp. 657-673.
5. Rajesh R., Shtessela Y., Edwards C. Accuracy improvement of dynamic sensors using sliding mode observers with dynamic extension. *Sensors and Actuators A: Physical*, 2020, vol. 316, p. 112396, DOI: 10.1016/j.sna.2020.112396.
6. Hamdi H., Rodrigues M., Rabaoui B. et al. A fault estimation and fault-tolerant control based sliding mode observer for LPV descriptor systems with time delay. *International Journal of Applied Mathematics and Computer Science*, 2021, vol. 31, no. 2, pp. 247-258, DOI: 10.34768/amcs-2021-001.
7. Sergiyenko O., Tyrsa V., Zhirabok A. et al. Sensor fault identification in linear and nonlinear dynamic systems via sliding mode observers. *IEEE Sensors Journal*, 2022, vol. 22, pp. 10173-10182, DOI: 10.1109/JSEN.2021.3080118.
8. Ефимов Д.В. Построение интервальных наблюдателей для динамических систем с неопределенностями / Д.В. Ефимов, Т. Раисси // *Автоматика и телемеханика*, 2016. – № 2. – С. 5-49.
9. Efimov D., Raissi T., Chebotarev S. et al. Interval state observer for nonlinear time varying systems. *Automatica*, 2013, vol. 49, pp. 200-206, DOI: 10.1016/j.automatica.2012.07.004.
10. Chebotarev S., Efimov D., Raissi T. et al. Interval observers for continuous-time LPV systems with L1/L2 performance. *Automatica*, 2015, vol. 58, pp. 82-89, DOI: 10.1016/j.automatica.2015.05.009.
11. Zhu F., Fu Y., Dinh T. Asymptotic convergence unknown input observer design via interval observer. *Automatica*, 2023, no. 1, DOI: 10.1016/j.automatica.2022.110744.
12. Khan A., Xie W., Zhang L. et al. Design and applications of interval observers for uncertain dynamical systems. *IET Circuits, Devices & Systems*, 2020, vol. 14, pp. 721-740, DOI: 10.1049/iet-cds.2020.0004.
13. Xu F. Generalized set-theoretic interval observer using element-wise nonnegative transformation. *Automatica*, 2020, vol. 114, pp. 1-14, DOI: 10.1016/j.automatica.2020.109452.
14. Nguyen D., Dinh T., Wang Z. et al. Robust interval estimation of state and unknown inputs for linear continuous-time systems: an L1-gain characterization. *European Journal of Control*, 2024, vol. 82, p. 11172, DOI: 10.1016/j.ejcon.2024.101172.
15. Жирабок А.Н. Метод построения интервальных наблюдателей для стационарных линейных систем / А.Н. Жирабок, А.В. Зуев, Ким Чхун Ир // *Известия РАН. Теория и системы управления*, 2022. – № 5. – С. 3-13.
16. Жирабок А.Н. Интервальные наблюдатели для непрерывных систем с параметрическими неопределенностями / А.Н. Жирабок, А.В. Зуев, В.Ф. Филаретов и др. // *Автоматика и телемеханика*, 2023. – № 11. – С. 3-16.
17. Жирабок А.Н. Интервальные наблюдатели для оценивания неизвестных входов в дискретных стационарных динамических системах / А.Н. Жирабок, А.В. Зуев // *Автоматика и телемеханика*, 2024. – № 8. – С. 3-19.

18. Sergiyenko O., Zhirabok A., Ibraheem I. et al. Interval observers for discrete-time linear systems with uncertainties. *Symmetry*, 2022, vol. 14, pp. 21-31.
19. Zhirabok A., Zuev A., Filaretov V. Interval parity relations design for fault diagnosis. *International Journal of Adaptive Control and Signal Processing*, 2024, vol. 38, no. 7, pp. 2520-2531.
20. Dinh T., Tran G. Systematic interval observer design for linear systems. *Automatica*, 2025, vol. 61, pp. 1-6.
21. Wang T., Li Y., Xiang W. Design of interval observer for continuous linear large-scale systems with disturbance attenuation. *Journal of the Franklin Institute*, 2022, vol. 359, pp. 3910-3929.
22. Levant A. Higher-order sliding modes, differentiation and output-feedback control. *International Journal of Control*, 2003, vol. 76, pp. 924-941.

Жирабок Алексей Нилович. Профессор, д.т.н., Дальневосточный федеральный университет, профессор. Основные направления исследований – теория сложных систем, техническая диагностика. AuthorID: 14905, SPIN: 8492-4471, ORCID: 0000-0001-5927-7117. zhirabok@mail.ru. 690922, Россия, Владивосток, Русский остров, п. Аякс, 10.

Зуев Александр Валерьевич. Доцент, д.т.н., Институт проблем морских технологий ДВО РАН, старший научный сотрудник. Основные направления исследований – адаптивное и робастное управление, робототехника. AuthorID: 239412, SPIN: 3691-4666, ORCID: 0000-0002-0934-6222, alvzuev@yandex.ru. 690990, Россия, Владивосток, ул. Суханова, 5.

UDC 621.396

DOI:10.25729/ESI.2026.43.3.008

Method of fault estimation in technical systems described by linear models

Alexey N. Zhirabok¹, Alexander V. Zuev²

¹Far Eastern Federal University, Russia, Vladivostok, zhirabok@mail.ru

²Institute of Marine Technology Problems FEB RAS, Russia, Vladivostok

Abstract. The problems of fault estimation in dynamics and sensors of technical systems described by linear dynamic models with time invariant parameters described by differential equations subject to the unknown external disturbances and measurement noise in sensors are studied. The purpose is to design a device using information about inputs and readings to estimate values of faults both in dynamics and sensors. Such information is necessary to compensate the influence of these faults by fault tolerant control systems based on control correction. A solution is based on so-called interval observers of minimal dimension insensitive to the external disturbances. Such observers are constructed based on the reduced-order model of the original system insensitive to the disturbances as well. To construct the reduced-order model, the appropriate relations are used based on the Jordan canonical form describing dynamics of the model and observer. The Jordan canonical form is necessary to use the appropriate properties of the observer. Based on such observer and differentiator which produces exact estimate in finite time, the interval estimate relations are derived producing the relations to estimate faults both in dynamics and sensors. Novelty of the paper is that unlike typical way of using the interval observer, a device for estimation is constructed based on the reduced-order model insensitive to the external disturbances. This allows reducing the computational complexity of the estimation procedure and increasing the accuracy of estimation due to insensitivity to the disturbances. Theoretical results are illustrated by an example of fault estimation in the electrical actuator of manipulation robot. Simulation based on the package Matlab confirms a correctness of theoretical results. It is shown that using the reduced-order model allows reducing the computational complexity of the estimation procedure. The obtained relations can be used to design fault tolerant systems.

Keywords: linear systems, models, faults, interval observers

Acknowledgements: The work was supported by the Russian Science Foundation, project no. 25-19-00590, <https://rscf.ru/project/25-19-00590/>.

References

1. Edwards C., Spurgeon S., Patton R. Sliding mode observers for fault detection and isolation. *Automatica*, 2000, vol. 36, pp. 541-553, DOI: 10.1016/S0005-1098(99)00177-6.
2. Chen L., Edwards C., Alwi H. et al. Flight evaluation of a sliding mode online control allocation scheme for fault tolerant control. *Automatica*, 2020, vol. 114, p. 108829, DOI: 10.1016/j.automatica.2020.108829.

3. Kalsi K., Hui S., Zak S. Unknown input and sensor fault estimation using sliding-mode observers. Proceedings of 2011 ACC, San Francisco, 2011, pp. 1364-1369.
4. Yan X., Edwards C. Sensor fault detection and isolation for nonlinear systems based on a sliding mode observer. International Journal of Adaptive Control and Signal Processing, 2007, vol. 21, pp. 657-673.
5. Rajesh R., Shtessela Y., Edwards C. Accuracy improvement of dynamic sensors using sliding mode observers with dynamic extension. Sensors and Actuators A: Physical, 2020, vol. 316, p. 112396, DOI: 10.1016/j.sna.2020.112396.
6. Hamdi H., Rodrigues M., Rabaoui B. et al. A fault estimation and fault-tolerant control based sliding mode observer for LPV descriptor systems with time delay. International Journal of Applied Mathematics and Computer Science, 2021, vol. 31, no. 2, pp. 247-258, DOI: 10.34768/amcs-2021-001.
7. Sergiyenko O., Tyrsa V., Zhirabok A. et al. Sensor fault identification in linear and nonlinear dynamic systems via sliding mode observers. IEEE Sensors Journal, 2022, vol. 22, pp. 10173-10182, DOI: 10.1109/JSEN.2021.3080118.
8. Efimov D.V., Raissi T. Postroyeniye interval'nykh nablyudateley dlya dinamicheskikh sistem s neopredelennostyami [Design of interval observers for dynamic systems with uncertainties]. Avtomatika i telemekhanika [Automation and Remote Control], 2016, no. 2, pp. 5-49.
9. Efimov D., Raissi T., Chebotarev S. et al. Interval state observer for nonlinear time varying systems. Automatica, 2013, vol. 49, pp. 200-206, DOI: 10.1016/j.automatica.2012.07.004.
10. Chebotarev S., Efimov D., Raissi T. et al. Interval observers for continuous-time LPV systems with L1/L2 performance. Automatica, 2015, vol. 58, pp. 82-89, DOI: 10.1016/j.automatica.2015.05.009.
11. Zhu F., Fu Y., Dinh T. Asymptotic convergence unknown input observer design via interval observer. Automatica, 2023, no. 1, DOI: 10.1016/j.automatica.2022.110744.
12. Khan A., Xie W., Zhang L. et al. Design and applications of interval observers for uncertain dynamical systems. IET Circuits, Devices & Systems, 2020, vol. 14, pp. 721-740, DOI: 10.1049/iet-cds.2020.0004.
13. Xu F. Generalized set-theoretic interval observer using element-wise nonnegative transformation. Automatica, 2020, vol. 114, pp. 1-14, DOI: 10.1016/j.automatica.2020.109452.
14. Nguyen D., Dinh T., Wang Z. et al. Robust interval estimation of state and unknown inputs for linear continuous-time systems: an L1-gain characterization. European Journal of Control, 2024, vol. 82, p. 11172, DOI: 10.1016/j.ejcon.2024.101172.
15. Zhirabok A.N., Zuev A.V., Kim Chun Ir. Metod postroyeniya interval'nykh nablyudateley dlya statsionarnykh lineynykh sistem [Method for designing interval observers for stationary linear systems]. Izvestiya RAN. Teoriya i sistemy upravleniya [Proceedings of RAS. Theory and Control Systems], 2022, no. 5, pp. 3-13.
16. Zhirabok A.N., Zuev A.V., Filaretov V.F. et al. Interval'nyye nablyudateli dlya nepreryvnykh sistem s parametricheskimi neopredelennostyami [Interval observers for continuous systems with parametric uncertainties]. Avtomatika i telemekhanika [Automation and Remote Control], 2023, no. 11, pp. 3-16.
17. Zhirabok A.N., Zuev A.V. Interval'nyye nablyudateli dlya otsenivaniya neizvestnykh vkhodov v diskretnykh statsionarnykh dinamicheskikh sistemakh [Interval observers for unknown input estimation in discrete stationary dynamic systems]. Avtomatika i telemekhanika [Automation and Remote Control], 2024, no. 8, pp. 3-19.
18. Sergiyenko O., Zhirabok A., Ibraheem I. et al. Interval observers for discrete-time linear systems with uncertainties. Symmetry, 2022, vol. 14, pp. 21-31.
19. Zhirabok A., Zuev A., Filaretov V. Interval parity relations design for fault diagnosis. International Journal of Adaptive Control and Signal Processing, 2024, vol. 38, no. 7, pp. 2520-2531.
20. Dinh T., Tran G. Systematic interval observer design for linear systems. Automatica, 2025, vol. 61, pp. 1-6.
21. Wang T., Li Y., Xiang W. Design of interval observer for continuous linear large-scale systems with disturbance attenuation. Journal of the Franklin Institute, 2022, vol. 359, pp. 3910-3929.
22. Levant A. Higher-order sliding modes, differentiation and output-feedback control. International Journal of Control, 2003, vol. 76, pp. 924-941.

Zhirabok Alexey Nilovich. Professor, Dr. of Sciences, Far Eastern Federal University, professor. The main direction of researches – complex system theory, fault diagnosis. AuthorID: 14905, SPIN: 8492-4471, ORCID: 0000-0001-5927-7117. zhirabok@mail.ru. 690922, Russia, Vladivostok, Russian Island, Ajax village, 10.

Zuev Alexander Valerievich. Institute of Marine Technology Problems of the Far Eastern Branch of the Russian Academy of Sciences, Senior Researcher. The main direction of researches – adaptive and robust control and robotics. AuthorID: 239412, SPIN: 3691-4666, ORCID: 0000-0002-0934-6222, alvzuev@yandex.ru. 690990, Russia, Vladivostok, st. Sukhanova, 5.

Статья поступила в редакцию 08.01.2026; одобрена после рецензирования 17.02.2026; принята к публикации 30.07.2026.

The article was submitted 01/08/2026; approved after reviewing 02/17/2026; accepted for publication 07/30/2026.

УДК 519.816: 519.863: 621.311.16

DOI:10.25729/ESI.2026.43.3.009

Разработка гибридного метода оптимального распределения электрической нагрузки с учетом трудноформализуемых факторов

Панкратьев Павел Сергеевич

Братский государственный университет, Россия, Братск, scud33@inbox.ru

Аннотация. Оптимальное распределение электрических нагрузок между энергоисточниками является важным процессом, позволяющим достичь минимальных финансовых издержек у поставщика энергии при текущем уровне спроса. Данный процесс, в мировой практике также известный как экономическое диспетчирование, традиционно строится на принципе равенства относительных приростов стоимости топлива, что позволяет определить оптимальное распределение мощностей с точки зрения достижения экономической эффективности. Однако современное состояние энергетики, характеризующееся глобальной ESG-трансформацией и реализацией принципов корпоративной социальной ответственности, требует учета не только прямых экономических показателей, но и социально-экологических. Последние, как правило, зачастую бывают трудноформализуемыми или вовсе неформализуемыми, что вносит значительную сложность в процесс оптимизации. Данные условия определяют необходимость появления в обозначенном процессе лица, принимающего решение (ЛПР), с присущими ему внутренними ценностными установками. В описанных обстоятельствах традиционный подход к оптимизации по критерию минимума затрат становится недостаточным, что диктует необходимость разработки новых гибридных методов для её реализации. В рассматриваемой статье предложен гибридный метод распределения мощностей, основанный на синтезе классического метода множителей Лагранжа, обобщенного условиями Каруша–Куна–Таккера, и теории многокритериальной полезности (MAUT). Научная новизна подхода заключается в прямой интеграции многокритериальной функции ценности, отражающей субъективные предпочтения ЛПР, непосредственно в структуру оптимизационной задачи. Для апробации разработанного гибридного метода смоделирован изолированный энергорайон с суммарной нагрузкой 500 МВт. Оптимизационная задача в статической постановке решена для двух сценариев: базового – с минимизацией топливных издержек, и социально ориентированного – с учетом участия ЛПР и соответствующего активного ограничения по уровню функции ценности ($V \geq 0,60$). Результаты показали, что удовлетворение требований ЛПР на уровне ценности 0,6 приводит к существенной перестройке энергобаланса – замещению 43% базовой угольной генерации мощностями газовой ТЭС. Показано, что возникший при этом рост затрат (на 22%) компенсируется достижением требуемого уровня социальной ценности ($V = 0,60$). Вычисленные значения двойственных переменных (множителей Лагранжа) интерпретированы, как «теневые цены» неформализуемых ограничений, отражающие предельную стоимость улучшения социально-экологической обстановки.

Ключевые слова: оптимальное распределение нагрузки, экономическое диспетчирование, системный анализ, теория многокритериальной полезности (MAUT), метод множителей Лагранжа, условия Каруша–Куна–Таккера, теневая цена, принятие решений

Цитирование: Панкратьев П.С. Разработка гибридного метода оптимального распределения электрической нагрузки с учетом трудноформализуемых факторов / П.С. Панкратьев // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 103-118. – DOI:10.25729/ESI.2026.43.3.009.

Введение. Оптимальное распределение активной мощности между генерирующими источниками энергосистемы представляет собой одну из главных задач оперативно-диспетчерского управления. Основным смысл такого процесса традиционно сводится к минимизации издержек на генерацию всей мощности в системе при условии равенства последней и питаемой ею нагрузки потребителей [1–6]. Однако в настоящее время в мировой практике и в том числе в России существует устойчивая тенденция к реализации принципов корпоративной социальной ответственности (КСО, CSR – Corporate Social Responsibility) [7–9] и учёту ESG-факторов (Environmental, Social, Governance) – то есть факторов неэкономических: социальных, экологических и управленческих [3, 5, 7, 8, 10–18], – что также связано с необходимостью поиска баланса в рамках концепции энергетической трилеммы (безопасность, доступность, экология) [10, 14]. В этой связи традиционный подход к

оптимизации по критерию минимума затрат и соответствующему ему балансному распределению мощности становится недостаточным [3–5, 10, 11, 15, 18–20].

В российской практике, особенно в изолированных энергорайонах и зонах децентрализованного энергоснабжения, до недавнего времени решение о распределении мощностей часто принималось в условиях «ручного управления», когда приоритет надежности или социального спокойствия мог доминировать над экономической эффективностью [5, 11, 12, 14, 16, 18, 21–23]. Однако с 1 января 2024 года ситуация изменилась: в соответствии с Федеральным законом № 174-ФЗ функции оперативно-диспетчерского управления в этих территориальных энергосистемах перешли к единому системному оператору (АО «СО ЕЭС»), который, согласно указанному закону, в своей деятельности прямо и безальтернативно руководствуется принципом достижения экономической эффективности при осуществлении оперативно-диспетчерского управления. Несмотря на централизацию управления, острые социальные и экологические вопросы на местах в изолированных районах сохраняют свою актуальность. Смена модели управления лишь обостряет необходимость поиска инструментария, который позволил бы учесть эти традиционные «местные» неэкономические факторы в рамках централизованного диспетчирования.

Методология включения экологических ограничений в рассматриваемую задачу оптимизации нагрузок была разработана еще в начале 1970-х годов [24] и активно развивалась [4, 25]. Однако современные вызовы, связанные с необходимостью учёта неэкономических факторов, согласованных с принципами CSR и ESG, усложняют данную задачу, так как многие учитываемые ограничения могут иметь трудноформализуемую, либо вовсе неформализуемую природу, а оценки по таким факторам могут носить качественный характер [5, 8, 13, 17, 19–21, 26]. К таким факторам относятся, например, социальная приемлемость используемых технологий генерации, экологические показатели, репутационные последствия для компании [4, 5, 7, 9, 10, 13, 17, 19, 27], а также, что критично для изолированных территорий: риски нарушения топливной безопасности (зависимость от «северного завоза»), приоритет надежности теплоснабжения в суровых климатических условиях и необходимость сохранения рабочих мест на градообразующих объектах энергетики.

Данное обстоятельство предопределяет появление в процессе диспетчеризации лица, принимающего решение (ЛПР), с внесением в данный процесс определённой доли субъективности при назначении оценок. Сама же задача диспетчеризации приобретает, таким образом, многокритериальный слабоструктурированный характер.

В описанной ситуации ЛПР необходимо прибегать к использованию особых подходов и методов, которые учитывали бы многокритериальность и позволили бы ему формализовать оценки по упомянутым неэкономическим трудноформализуемым факторам.

С точки зрения конечного результата работы методов многокритериального анализа последние можно подразделить на две основные группы. К первой группе относятся методы, результатом работы которых является получение Парето-оптимального множества альтернатив (фронт Парето). Ко второй группе – методы, позволяющие получить ранжировку альтернатив (от худшей к лучшей), вершиной которой является единственная и самая наилучшая альтернатива.

Методы многокритериального анализа, предлагающие построение множества Парето-оптимальных решений, могут быть сложны для оперативного использования в разрезе задачи диспетчирования: высокая когнитивная нагрузка, ложащаяся на диспетчера ввиду необходимости выбора из десятков альтернатив, а также вынужденность принимать решение «здесь и сейчас» могут спровоцировать появление ошибок [3, 15, 18, 20, 28, 29].

Масштабный аналитический обзор современного состояния проблемы диспетчеризации [30] и многокритериальной оптимизации в целом [31] свидетельствует о том, что в современной энергетике имеет место переход к гибридным интеллектуальным стратегиям, способным функционировать в условиях высокой неопределенности и многокритериальности. Тем не менее, проблема обоснованного выбора единственного варианта из множества компромиссных Парето-оптимальных решений всё еще остаётся актуальной [29, 30].

Наилучшей стратегией с точки зрения оперативного использования диспетчером видится предоставление ему единственного наилучшего варианта перераспределения нагрузки. Это позволило бы мгновенно реагировать на изменения спроса, исключая задержки и ошибки, неизбежные при необходимости выбора из множества равнозначных альтернатив. В связи с этим алгоритм диспетчеризации перспективно основывать на многокритериальных методах, обеспечивающих получение именно единственного наилучшего решения.

В статье автором предложен гибридный метод на основе синтеза двух подходов – теории многокритериальной полезности (MAUT) и метода Лагранжа, обобщенного условиями Каруша-Куна-Таккера, – позволяющий учесть разнородные неэкономические требования в едином оптимизационном алгоритме и решить задачу распределения мощности в энергосистеме без необходимости «ручного» перебора множества Парето-оптимальных вариантов в поисках из них наилучшего. Основная идея предложенного гибридного метода сводится к интеграции многокритериальной функции полезности непосредственно в оптимизационную задачу в виде жесткого ограничения, что позволяет найти единственное, лучшее решение, удовлетворяющее заданному уровню притязаний ЛПП. Кроме того, данный подход весьма полезен с точки зрения получения экономических оценок («теневых цен») неформализуемых социально-экологических требований, анализ которых позволяет оценить прямую денежную стоимость социально ориентированного («этического») выбора ЛПП.

Математическая модель объекта. Предлагаемый подход представляет собой синтез методов теории полезности и нелинейного программирования. Вся процедура решения разделяется на два этапа: первый – этап формализации предпочтений ЛПП («оффлайн», то есть до распределения мощностей), и второй – этап оперативной оптимизации («онлайн», то есть непосредственно во время распределения).

Общая постановка задачи распределения мощностей следующая. Рассматривается энергосистема, состоящая из N генерирующих установок (источников энергии). Целевая функция минимизации издержек C_T имеет вид:

$$\min C_T = \min \sum_{i=1}^N C_i(P_i),$$

где $C_i(P_i) = a_i P_i^2 + b_i P_i + c_i$ – характеристика издержек i -го источника (на примере тепловой электростанции), которую можно аппроксимировать квадратичным полиномом; P_i – мощность i -го источника.

Задача решается при выполнении следующих ограничений:

1. Баланс активной мощности: сумма генерации должна быть равна нагрузке потребителей D :

$$\sum_{i=1}^N P_i - D = 0.$$

2. Технические ограничения: мощность каждого энергоисточника должна находиться в допустимом диапазоне.

$$P_i^{\min} \leq P_i \leq P_i^{\max}, \quad i = 1 \dots N.$$

Интеграция предпочтений ЛПР на основе MAUT. Для формализации субъективных предпочтений ЛПР применяется теория многокритериальной полезности (Multi-Attribute Utility Theory, MAUT) [21, 32–34], используемая на первом этапе процедуры решения (до распределения мощностей). Поскольку настоящая задача решается в детерминированной постановке (без учета вероятностных рисков), в соответствии с понятийным аппаратом MAUT далее вместо понятия «*полезность*» (*utility*) будет использовано понятие «*ценность*» (*value*).

Следует отметить еще один терминологический нюанс: переменные, описывающие состояние системы, в классической терминологии MAUT именуются *атрибутами* (*attributes*). В контексте решаемой задачи в качестве атрибутов выступают активные мощности источников P_i . Эти атрибуты используются для оценки достижения локальных целей – минимизации негативного воздействия соответствующего источника энергии на окружающую среду и социум.

Таким образом, формируются однокритериальные функции ценности (ОФЦ) $v_i(P_i)$, отражающие степень удовлетворенности ЛПР состоянием конкретного атрибута. Для упрощения принята линейная зависимость функции ценности, убывающая с ростом мощности:

$$v_i(P_i) = 1 - \frac{P_i}{P_i^{base}},$$

где $v_i=1$ соответствует наилучшему состоянию (источник отключен), а $v_i=0$ – наихудшему (работа на максимальной мощности).

Для агрегирования локальных оценок по полученным ОФЦ используется аддитивная многокритериальная функция ценности (МФЦ) $V(P)$. Ее использование правомерно при допущении о независимости атрибутов по предпочтениям. МФЦ $V(P)$ в аддитивной форме имеет вид:

$$V(P) = \sum_{i=1}^N w_i \cdot v_i(P_i) = \sum_{i=1}^N w_i \left(1 - \frac{P_i}{P_i^{base}} \right),$$

где: w_i – весовой коэффициент значимости i -го источника, принимающий значения от 0 до 1 (при этом $\sum w_i=1$); P_i^{base} – базовая мощность для нормировки (обычно $P_i^{base} = P_i^{\max}$).

Таким образом, w_i можно интерпретировать как весовой коэффициент важности минимизации мощности i -го источника, где более высокие веса присваиваются источникам с более высоким негативным социально-экологическим воздействием (например, для угольной ТЭС).

В условиях новой модели управления изолированными энергосистемами к процедуре назначения весов w_i на первом этапе предполагается привлекать коллективное ЛПР – представителей региональных органов власти и системного оператора. Это позволяет сформировать своего рода «профиль социальной ответственности» энергорайона, обязательный для исполнения в оперативных алгоритмах. При этом для получения численных значений весовых коэффициентов w_i могут быть использованы различные подходы. В частности, классический метод MAUT, основанный на процедуре лотерей [32, 33], обладает высокой аксиоматической строгостью, однако эта процедура может быть весьма сложна для практического применения коллективным ЛПР. На практике широкое применение находят более адаптивные подходы: например, метод SWING [35], SMARTER [36] или метод парных сравнений АНР [37]. В каждом из этих подходов концепция получения предпочтений существенно отличается, однако «на выходе» веса в результате нормировки обретают требуемые значения в диапазоне от 0 до 1, пригодные для использования в аддитивной многокритериальной функции ценности.

После алгебраических преобразований функция приводится к компактному линейному виду, удобному для дальнейших аналитических преобразований:

$$V(P) = 1 - \sum_{i=1}^N k_i P_i, \quad (1)$$

где $k_i = \frac{w_i}{P_i^{base}}$ – коэффициент, который имеет физический смысл «удельного снижения социальной ценности» на 1 МВт мощности для i -го энергоисточника.

Метод Лагранжа и условия Каруша-Куна-Таккера. Для работы на втором этапе алгоритма (непосредственно во время распределения нагрузки) в качестве оптимизирующего инструмента использован метод множителей Лагранжа, обобщенный условиями Каруша-Куна-Таккера (ККТ), позволяющими включать в оптимизационную задачу ограничения типа неравенств через математический аппарат дополняющей нежесткости [38–40]. Ключевой особенностью предложенного в настоящей работе гибридного метода является интеграция требования по минимальному уровню ценности $V(P) \geq V_{min}$ непосредственно в структуру функции Лагранжа в виде активного ограничения. Параметр V_{min} при этом также избирается коллективным ЛПП в ходе определенных диалоговых процедур.

На данном втором этапе методы Лагранжа и МАУТ работают в связке, формируя модифицированную функцию (лагранжиан):

$$L(P, \lambda, \mu) = \sum_{i=1}^N C_i(P_i) + \lambda \left(D - \sum_{i=1}^N P_i \right) + \mu (V_{min} - V(P)). \quad (2)$$

Стоит отметить, что технические ограничения по источникам P_{min}/P_{max} также учитываются через соответствующие множители ККТ, но для упрощения записи здесь опущены.

Для поиска оптимального решения далее лагранжиан (2) необходимо продифференцировать по всем варьируемым переменным и, в соответствии с условиями ККТ, приравнять к нулю полученные выражения, что будет соответствовать условиям требуемой стационарности. В результате этих алгебраических операций образуется выражение:

$$\frac{\partial L}{\partial P_i} = \frac{dC_i}{dP_i} - \lambda + \mu \frac{\partial (V_{min} - V(P))}{\partial P_i} = 0. \quad (3)$$

Подставляя выражение (1) для $V(P)$ в (3), получим модифицированное условие равенства относительных приростов предельных издержек (Marginal Cost, MC):

$$MC_i(P_i) + \mu \cdot k_i = \lambda, \quad (4)$$

где: $MC_i(P_i) = 2a_i P_i + b_i$ – классические предельные издержки по топливу; μ – двойственная переменная («теневая цена») социального ограничения; $\mu \cdot k_i$ – «социально-экологическая надбавка» для i -го источника.

Экономическая интерпретация выражения (4) заключается в следующем: в точке оптимума система воспринимает источники с низким уровнем социальной приемлемости (с высоким k_i) как более дорогие, что приводит к их «автоматической» разгрузке в пользу источников с более высоким уровнем социальной приемлемости, даже если топливные издержки последних оказались выше. Слагаемое $\mu \cdot k_i$ в данном случае играет роль своеобразного «штрафа» за низкую социальную ценность i -го источника, что влечёт за собой «искусственное» завышение предельных издержек MC_i .

Общая схема работы предложенного гибридного метода. Общая схема реализации предложенного гибридного метода распределения нагрузок представлена на рисунке 1. На первом этапе (этап стратегического планирования) решения принимаются заблаговременно (до распределения мощностей) при участии представителей региональных органов власти и

руководства системного оператора, которые на основе внешних факторов в ходе обсуждения принимают решение о настроечных параметрах модели (w_i , $V_{\min}$). Эти данные, а также прочие оперативные данные поступают на вход модели на втором этапе (этап оперативного управления), где автоматизированный алгоритм выполняет оптимизацию режима без оперативного вмешательства диспетчера. Главной особенностью схемы является однократная передача утвержденных неэкономических приоритетов в математическое ядро алгоритма, что позволяет получить единственное оптимальное решение, содержащее как информацию о распределении мощностей для диспетчера, так и соответствующий этому распределению экономический индикатор («теневую цену») для обоснования тарифов регулятором. При этом следует отметить, что настроечные параметры модели (w_i , $V_{\min}$) не являются статичными величинами: процедура стратегического планирования предполагает их периодическую актуализацию для адаптации работы алгоритма к изменению внешних условий.

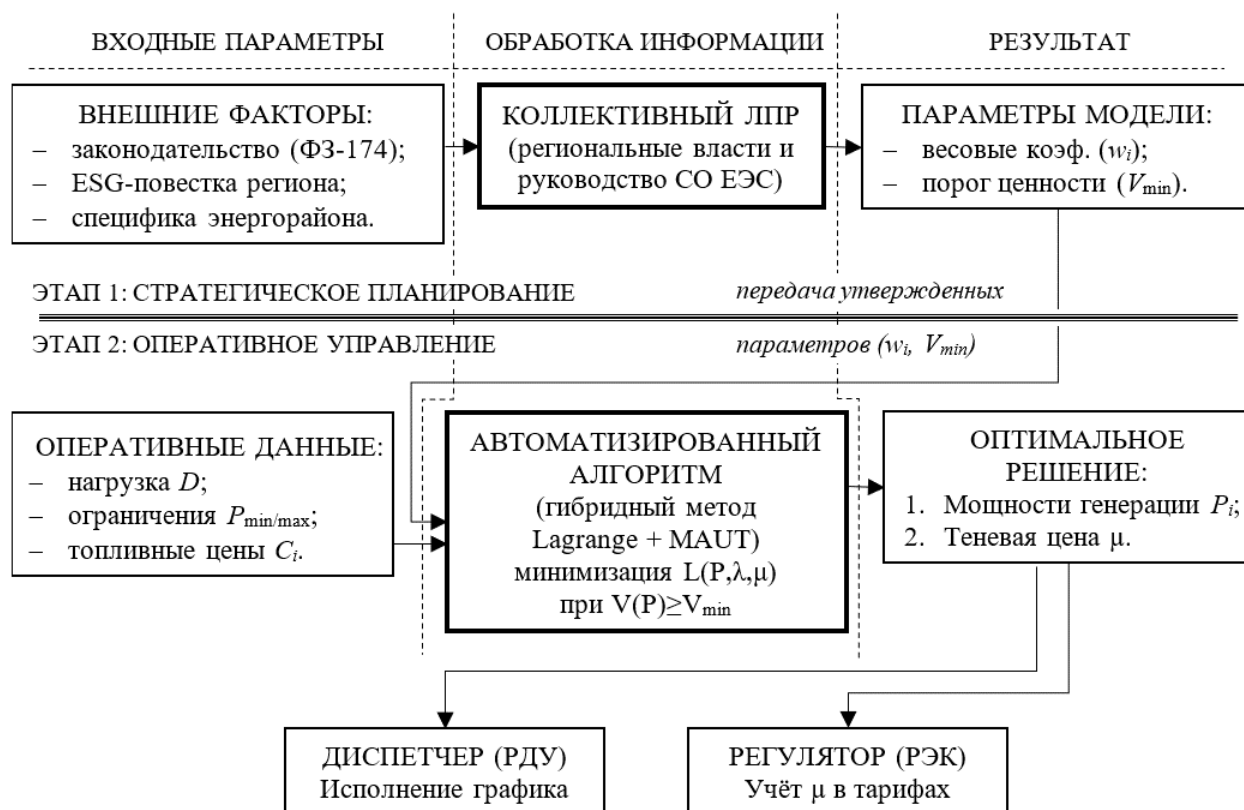


Рис. 1. Блок-схема предлагаемого гибридного метода диспетчеризации

Отличием предложенного подхода от существующих многокритериальных методов оптимизации является то, что последние предполагают построение множества Парето-оптимальных решений. Выбор наилучшего решения из данного множества создает высокую когнитивную нагрузку на диспетчера и может повлечь за собой появление ошибок. Предложенный синтез MAUT и метода Лагранжа позволяет избежать этого: в результате оптимизации отыскивается единственное решение, строго соответствующее заданному уровню притязаний ЛПР. При этом ключевой особенностью является то, что двойственная переменная μ наделяется физическим смыслом «стоимости социальной ответственности»: она количественно, в денежном эквиваленте, определяет, насколько возрастут суммарные издержки в системе при повышении ценности решения на малую величину. Таким образом, качественные ESG-требования приобретают конкретный экономический вид. Для системного оператора в изолированных системах этот показатель служит математическим обоснованием операционных решений, отклоняющихся от минимума топливных затрат, а для собственников

генерирующего оборудования – является доказательной базой для защиты соответствующих расходов в региональных органах тарифного регулирования.

Апробация предложенного метода. Характеристика энергосистемы. Ввиду методологического характера настоящей работы, а также объективной невозможности доступа к реальным данным, составляющим коммерческую тайну (характеристики генерирующего оборудования, структура топливных издержек и т.п.), апробация метода далее представлена на репрезентативном модельном примере.

Предлагаемый метод является универсальным и инвариантным относительно размерности задачи, количества источников и величин нагрузок. Однако для корректной демонстрации его работоспособности выбор количества источников в моделируемой системе целесообразно ограничить условием $N=3$. Это обусловлено тем, что для системы меньшей размерности ($N=2$) задача балансового распределения мощности сводится к тривиальному аналитическому решению, которое может быть найдено и без использования предлагаемого метода – путём выражения одной переменной через другую из уравнения баланса ($P_2=D-P_1$) и последующего решения полученного алгебраического уравнения. В то же время, увеличение размерности задачи ($N>3$) в рамках данного исследования будет избыточным: оно не привнесёт новой методологической ценности и не изменит концептуальную суть алгоритма, но существенно затруднит визуализацию и интерпретацию результатов. Таким образом, моделирование энергосистемы с количеством источников $N=3$ в настоящем исследовании представляется необходимым и достаточным условием, обеспечивающим баланс между методологической честностью и наглядностью изложения.

Используемая далее модель энергосистемы характеризуется контрастными, ярко выраженными параметрами, что позволяет детально проследить механизм действия алгоритма и интерпретировать полученные результаты. Для апробации предложенного гибридного метода смоделирован изолированный энергорайон с суммарной нагрузкой 500 МВт и тремя источниками электроснабжения, контрастно отличающимися своими основными технико-экономическими характеристиками:

- угольная ТЭС (G1) – обеспечивает базовую генерацию, имеет наименьшие затраты на выработку 1 МВт мощности ввиду дешевизны топлива, однако оказывает наибольшее негативное социально-экологическое воздействие;
- газовая ТЭС (G2) – манёвренная мощность со средними по величине затратами на топливо; характеризуется наилучшими показателями социальной приемлемости и экологической безопасности;
- дизельная электростанция (ДЭС) (G3) – пиковая мощность; отличается самыми высокими затратами на топливо и значительным уровнем негативных внешних эффектов (выбросы, шум, низкая социальная оценка).

Основные характеристики энергоисточников рассматриваемой системы представлены в таблице 1.

Таблица 1. Параметры источников электроэнергии и весовые коэффициенты MAUT

Источник	$P_{\min}$, МВт	$P_{\max}$, МВт	Функция издержек $C_i(P)$, \$/ч	Вес w_i	Удельный вред k_i , ед. ценности/МВт
G_1 (уголь)	50	400	$0,02P_1^2 + 20P_1 + 1000$	0,6	$1,5 \times 10^{-3}$
G_2 (газ)	30	450	$0,04P_2^2 + 40P_2 + 800$	0,1	$0,22 \times 10^{-3}$
G_3 (дизель)	10	200	$0,08P_3^2 + 80P_3 + 500$	0,3	$1,5 \times 10^{-3}$

Сценарии моделирования. Сравнительный анализ эффективности разработанного гибридного метода проводится на примере двух стратегий.

Сценарий А (базовый экономический): решается классическая задача оптимизации по критерию минимума топливных затрат. Ограничений по МФЦ нет ($\mu=0$) – таким образом, социальные факторы не влияют на распределение нагрузки.

Сценарий Б (социально-ориентированный): производится поиск компромисса между экономическими и социально-экологическими факторами. Коллективный ЛПР устанавливает жесткое требование к качеству решения: уровень ценности $V(P)$ не должен опускаться ниже порогового значения 0,60. Данная планка ($V \geq 0,60$) выбрана экспертно и, по соображениям коллективного ЛПР, соответствует высокому стандарту социально-экологической ответственности.

Анализ сценария А. В данном сценарии в отсутствие социальных требований оптимизация проводится только с учётом критерия стоимости. Опуская все вычисления, приведём лишь итоговые результаты расчётов по данному сценарию. Предельные издержки по ДЭС (источник G_3) составляют $MC_3 \geq 80$ \$/МВт·ч, что значительно превышает издержки по угольной (G_1) и газовой (G_2) ТЭС ($MC_1 = 36$ и $MC_2 = 48$ \$/МВт·ч соответственно), поэтому ДЭС – отключена ($P_3=0$). Угольная ТЭС, имеющая наиболее низкие издержки на генерацию мощности, загружается до своего технического максимума $P_1=400$ МВт, а оставшиеся 100 МВт покрываются газовой станцией, $P_2=100$ МВт. Итоговые минимально возможные затраты для системы, состоящей из двух работающих источников G_1 и G_2 с мощностями $P_1=400$ МВт и $P_2=100$ МВт соответственно, составляют 17400 \$/ч. Поскольку в данном сценарии отсутствует требование повышения социально-экологического благополучия, при текущем распределении мощностей наличествует низкий уровень ценности ($V \approx 0,38$), так как угольная ТЭС, имеющая наиболее низкий уровень социальной приемлемости, работает на своём пределе ввиду дешевизны угля.

Анализ сценария Б. В этом сценарии, как уже отмечалось, учитываются социально-экологические требования ЛПР. Нагрузка между источниками теперь распределяется с учётом налагаемого данным требованием ограничения по уровню ценности $V \geq 0,60$. Поскольку в базовом сценарии А ценность составляла $V=0,38$, что ниже требуемой в данном сценарии Б ($V \geq 0,60$), то ограничение-неравенство теперь становится активным. Так как любое дальнейшее повышение ценности сверх минимально требуемой ($V=0,60$) сопряжено с ростом издержек, алгоритм минимизации затрат удерживает решение строго на этой границе ($V=0,60$). В силу этого, согласно условию дополняющей нежесткости ККТ, неравенство математически трансформируется в строгое равенство $V(P) = 0,60$. При этом дизельная станция (G_3) в оптимальном решении остается отключенной ($P_3=0$). Этот факт даёт право найти оптимальное распределение мощностей между источниками G_1 и G_2 из решения системы уравнений, включающей баланс мощности и уравнение активного ограничения по ценности:

$$\begin{cases} P_1 + P_2 = 500 \\ 1 - 0,0015P_1 - 0,000222P_2 = 0,60 \end{cases}.$$

Решая систему через подстановку $P_2=500 - P_1$, получаем: $P_1 \approx 226,7$ МВт, $P_2 \approx 273,3$ МВт. Результаты также верифицированы численно в программной среде Python с использованием библиотеки SciPy.

В качестве примечаний к полученным результатам по данному сценарию необходимо отметить следующее.

Во-первых, дизельная станция (G_3) по-прежнему остается отключенной ($P_3=0$). Хотя показатели социальной приемлемости ДЭС превосходят таковые у угольной ТЭС, и априорно можно было бы ожидать включения ДЭС в баланс мощностей, оптимизационный алгоритм, тем не менее, сопоставив экономические издержки и социальные выгоды, определил включение ДЭС в баланс нецелесообразным, отдавая приоритет более дешевой, хотя и менее

социально одобряемой, угольной мощности (G_1) в сочетании с увеличением загрузки газовой станции (G_2). Данный результат подтверждает способность метода селективно определять оптимальный состав работающего оборудования в условиях многомерной оптимизации ($N \geq 3$), корректно обрабатывая граничные решения с полным исключением из баланса неэффективных источников.

Во-вторых, полученное упрощение задачи до системы линейных алгебраических уравнений является частным случаем. Данная ситуация будет иметь место всегда, когда количество активных источников N_{act} в системе будет равняться двум – вне зависимости от их общего количества N ; иначе говоря, при условии $N_{act} = N - N_0 = 2$, где N_0 – количество отключенных источников. Во всех остальных случаях (при $N_{act} \geq 3$) поиск решения потребует полноценного процесса минимизации функции Лагранжа.

Из сказанного также следует, что переход к упрощенному решению возможен только после того, как оптимизационный алгоритм подтвердит активность ограничения по ценности ($\mu > 0$) и определит состав включенного оборудования с $N_{act} = 2$.

В-третьих, полученное значение мощности для газовой генерации $P_2 = 273,3$ МВт находится достаточно далеко от верхней границы интервала разрешенной мощности данного источника G_2 ($P_{max} = 450$ МВт). Это показывает, что ограничение выработки обусловлено именно требованием учёта социального-экологического фактора, а не достижением технического предела этой станции.

Стоит отметить, что в настоящем основополагающем исследовании рассматриваются только статические условия – без учёта динамики нагрузки в суточном (и годовом) цикле. Учёт суточной (годовой) динамики и отыскание соответствующего баланса мощностей при помощи программной среды Python является целью дальнейших исследований.

Итоговые показатели по двум сценариям приведены в таблице 2.

Таблица 2. Сравнительный анализ сценариев оптимизации при $D = 500$ МВт

Параметр	Сценарий А – базовый экономический	Сценарий Б – социально ориентированный	Отклонение
Мощность источника G_1 (угольная ТЭС), МВт	400,0	226,7	–173,3
Мощность источника G_2 (газовая ТЭС), МВт	100,0	273,3	+173,3
Мощность источника G_3 (ДЭС), МВт	0,0	0,0	0,0
Суммарные издержки (C_T), \$/ч	17 400	21 305	+3905
Ценность по МФЦ $V(P)$	0,38	0,60	+0,22
Системная цена (λ), \$/МВт·ч	48,0	67,5	+19,5
Теневая цена (μ), \$/ед. цен.	0	25 657	–

Обсуждение результатов. Экономическая интерпретация двойственных переменных. Полученные значения множителей Лагранжа позволяют выйти за рамки простого распределения мегаватт и дают ЛПР мощный аналитический инструмент.

Во-первых, *теневая цена социального ограничения* (μ) в сценарии Б (порядка 25 657 \$/ед. цен.) по сути является индикатором «цены притязаний ЛПР»: чем выше μ , тем дороже обходится каждый следующий шаг в сторону улучшения социально-экологической обстановки. Это означает, что если ЛПР решит ужесточить требование с $V = 0,60$ до $V = 0,61$, это приведет к мгновенному росту часовых издержек системы на 256 \$/ч. Это и есть прямой финансовый эквивалент «этики» или социально-экологической ответственности. Зная это

число, ЛПР может не просто декларировать приверженность CSR- и ESG-принципам, но и четко осознавать, какую цену платит предприятие или регион за каждое управленческое решение, принятое в парадигме таких принципов.

Во-вторых, введение социально-экологических требований трансформирует *системную предельную цену* (λ). Рост с 48,0 до 67,5 \$/МВт·ч наглядно демонстрирует, как социальные-экологические ограничения фактически превращаются в «виртуальный налог» на генерацию. Разница в 19,5 \$/МВт·ч является экономическим обоснованием для формирования «зеленых тарифов» или субсидий. Таким образом, предложенный гибридный метод переводит предпочтения ЛПР по качественным трудноформализуемым социально-экологическим показателям на язык цифр, понятный финансовым директорам и регуляторам.

Визуализация результатов и анализ компромисса. Визуализация полученных результатов (рисунок 2) позволяет наглядно проследить механизм достижения компромисса в решении, учитывающем и экономические, и социально-экологические факторы, – в точке Б. Здесь достигается компромисс между экономическим распределением мощностей и этическими притязаниями ЛПР. Вектор, обозначенный как «Цена этики», фактически отражает вынужденное смещение системы из зоны абсолютного экономического минимума на изокосту с более высоким уровнем издержек (21 305 \$/ч).

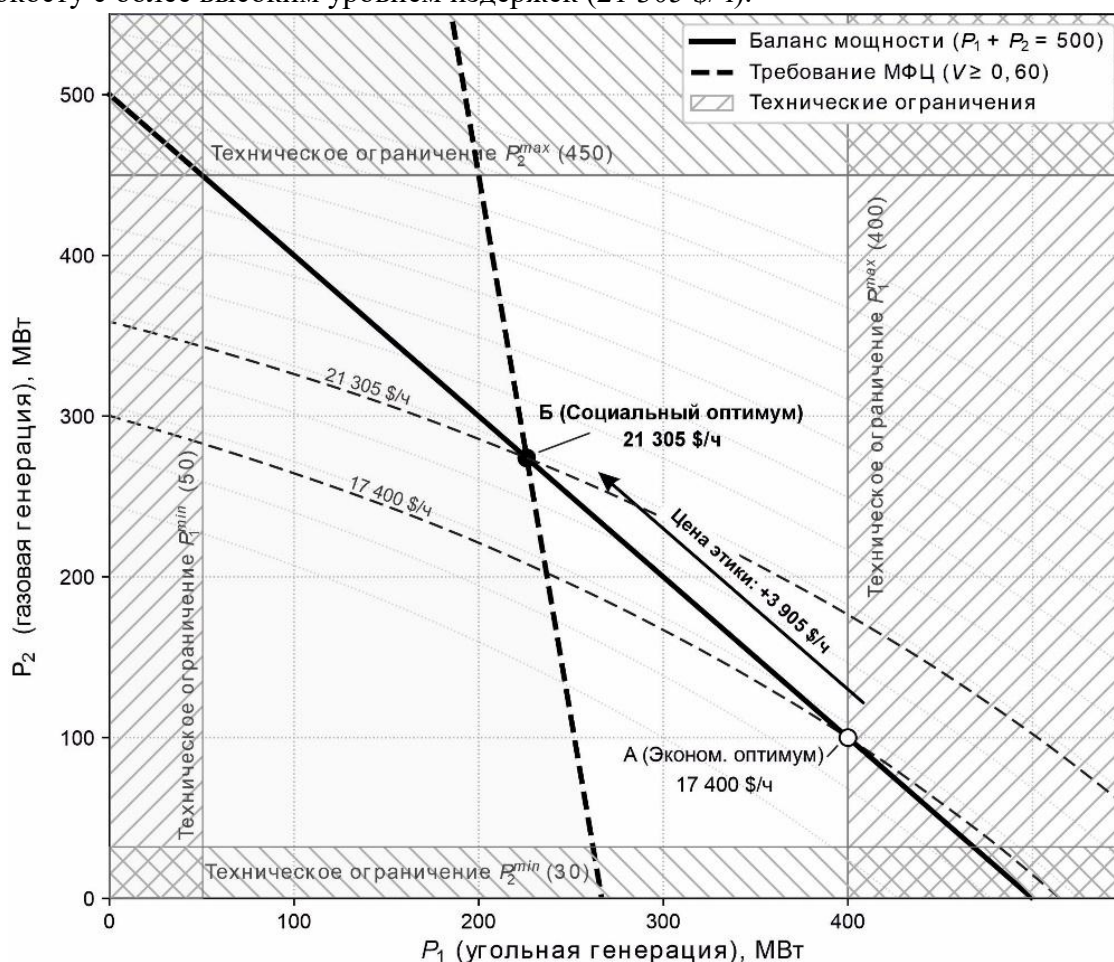


Рис. 2. Визуализация решения задачи оптимизации по двум сценариям

Показательным является и график зависимости полных затрат от заданного уровня ценности (рисунок 3). На нем отчетливо прослеживается переломный момент в поведении системы. До отметки $V=0,38$ система находится в так называемой зоне неактивных ограничений: здесь социально-экологические требования ЛПР настолько мягкие, что они удовлетворяются «сами собой» в ходе обычной минимизации затрат. Однако после преодоления этого порога ситуация меняется и далее наблюдается почти экспоненциальный

рост стоимости решения, лежащий полностью в зоне уже активного ограничения. Это сигнализирует ЛПР о том, что повышение планки его притязаний сверх $V=0,38$ с каждым шагом будет требовать всё больших и больших вложений, так как в данном случае в системе приходится замещать дешёвую угольную генерацию, имеющую низкий уровень социальной приемлемости, более дорогими, но в то же время более «социально одобряемыми» источниками.

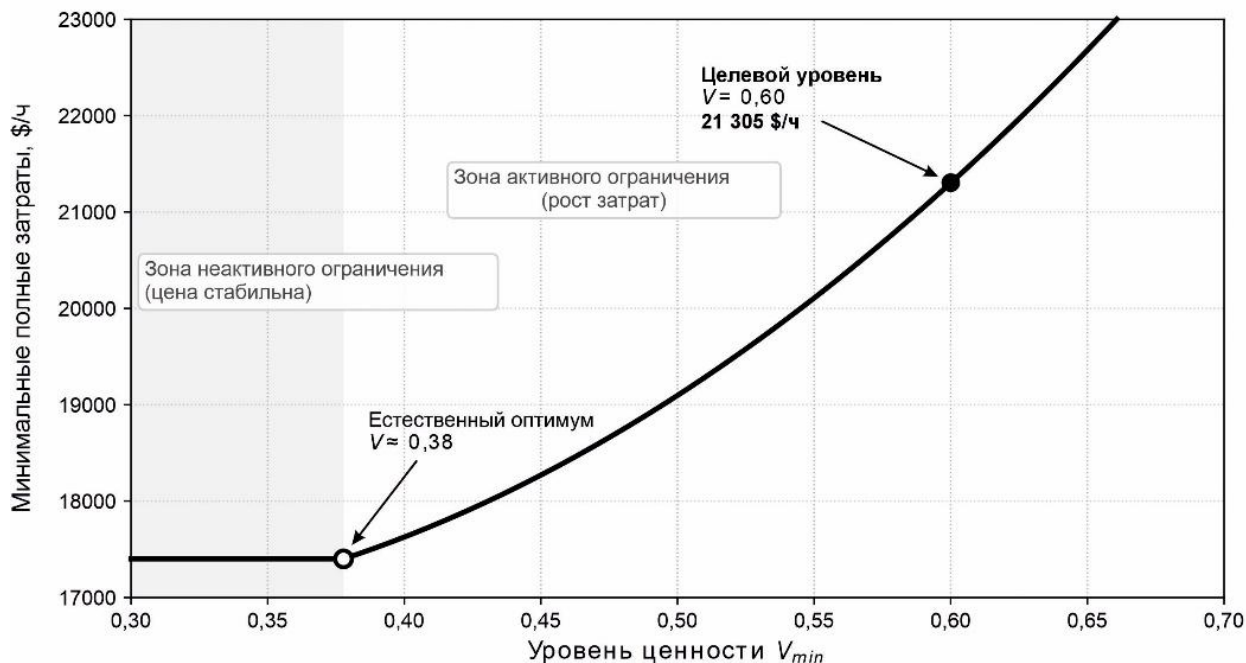


Рис. 3. Динамика роста минимальных полных затрат в зависимости от заданного уровня ценности

Практическая значимость. Если сравнивать предложенный подход с традиционным, когда управление распределением генерации происходит в режиме онлайн на основании лишь условия минимума совокупных затрат на выработку мощности в системе, можно отметить следующие моменты. Прежде всего, разработанный гибридный метод гарантирует, что найденный единственный вариант распределения нагрузки будет самым экономичным при строго заданном уровне социальной приемлемости. Здесь нет места субъективным ошибкам диспетчера.

Кроме того, алгоритм обладает высокой адаптивностью: при любых колебаниях цен на топливо или при пересмотре ЛПР значимости социально-экологических факторов система мгновенно пересчитывает оптимум, не требуя ручной корректировки.

Не менее важным аспектом является и прозрачность принятия решений: наличие четких количественных индикаторов, таких, как λ и μ , делает процесс выбора аргументированным, что превращает «интуитивное» управление в аудируемую процедуру, результаты которой можно обосновать как перед руководством, так и перед внешними регуляторами.

Отдельно стоит внести еще одну важную ремарку. Если в простейших случаях – где количество источников N невелико, как в приведённом примере, – распределение мощностей иногда можно найти геометрически, то в системах реальной размерности, где количество источников N может составлять десятки или сотни, без аппарата множителей Лагранжа невозможно будет найти ни сам оптимум, ни тем более определить экономическую стоимость (теневую цену) социально-экологических требований. Предложенный метод в данном смысле универсален и применим к энергосистемам любой размерности.

Заключение. В работе представлен новый подход к решению задачи оптимального распределения нагрузки, позволяющий интегрировать неэкономические

трудноформализуемые предпочтения ЛПП в строгую математическую модель. На примере простейшего случая с тремя электрическими станциями была продемонстрирована работа предложенного гибридного метода.

Основные результаты исследования:

1. Сформулирована задача оптимального распределения нагрузки с учётом субъективных предпочтений ЛПП. Средством формализации субъективных предпочтений выступает многокритериальная функция ценности, внедрённая непосредственно в математический аппарат метода множителей Лагранжа (в частности, в условия Каруша-Куна-Таккера) в виде ограничения-неравенства.
2. Численное моделирование показало, что для удовлетворения притязаний ЛПП в отношении социально-экологических факторов на уровне $V=0,60$ необходимо перераспределить нагрузку от источников с более низкими топливными издержками и одновременно более низким уровнем социальной приемлемости (угольная ТЭС) к источникам с более высокими топливными издержками и одновременно более высоким уровнем социальной приемлемости (газовая ТЭС). В рассмотренном примере такое замещение мощности в пользу газовой генерации составило 43% от исходного сценария.
3. Продemonстрирована возможность количественного измерения «цены этики» в энергетике. В рамках рассмотренного примера суммарные издержки возросли на 22% (3 905 \$/ч). При этом была определена предельная стоимость дальнейшего улучшения социально-экологической ситуации – 256 \$/ч за каждый процентный пункт (0,01) прироста функции ценности.

Научная новизна подхода заключается в прямой интеграции многокритериальной функции ценности, отражающей субъективные предпочтения ЛПП, непосредственно в структуру оптимизационной задачи. В отличие от существующих методов, ориентированных на построение множества Парето-оптимальных решений, предложенный гибридный метод позволяет определить единственное решение, соответствующее заданному уровню притязаний ЛПП.

Разработанный гибридный метод предлагается к использованию в программно-аппаратных комплексах системного оператора (АО «СО ЕЭС») для управления режимами технологически изолированных энергосистем. Он позволяет разрешить противоречие между законодательным требованием достижения экономической эффективности и необходимостью учета региональной специфики (социальных и экологических факторов) в разрезе актуальных сегодня ESG- и CSR-принципов, обеспечивая прозрачность и обоснованность диспетчерских решений.

Дальнейшее развитие идеи, представленной в настоящей работе, предполагает учёт колебаний нагрузки во времени и интеграцию стохастической генерации ВИЭ, где функция ценности (или полезности) может быть дополнена критериями надежности энергоснабжения.

Список источников

1. Wood A.J., Wollenberg B.F., Sheblé G.B. Power generation, operation and control. Hoboken: IEEE Wiley, 2014, 658 p.
2. Happ H.H. Optimal power dispatch – A comprehensive survey. IEEE Transactions on Power Apparatus and Systems, 1977, vol. 96(3), pp. 841-854, DOI: 10.1109/T-PAS.1977.32397.
3. Abido M.A. Environmental/economic power dispatch using multiobjective evolutionary algorithms. IEEE Transactions on Power Systems, 2003, vol. 18(4), pp. 1529-1537, DOI: 10.1109/TPWRS.2003.818693.
4. Talaq J.H., El-Hawary F., El-Hawary M.E. A summary of environmental/economic dispatch algorithms. IEEE Transactions on Power Systems, 1994, vol. 9(3), pp. 1508-1516, DOI: 10.1109/59.336110.
5. Voropai N.I. Multi-Criteria Problems in Electric Power System Expansion Planning. Energy Systems Research, 2018, vol. 1(2), pp. 27-34, DOI: 10.25729/esr.2018.02.0003.
6. Горнштейн В.М. Методы оптимизации режимов энергосистем / В.М. Горнштейн, Б.П. Мирошниченко, А.В. Пономарев. – Москва: Энергия, 1981. – 336 с.

7. Капанова Л.Д. ESG-трансформация как парадигма устойчивого развития экономики России / Л.Д. Капанова, О.А. Полищук // Вестник Евразийской науки, 2023. – Т. 15. – № 3. – С. 1-17.
8. Коданева С.И. Энергетический переход: перспективы и механизмы реализации / С.И. Коданева // Россия и современный мир, 2022. – № 4. – С. 162-183. – DOI: 10.31249/rsm/2022.04.09.
9. Zhang L., Xu S., Chen J., Li H., et al. The impact of ESG ratings on corporate overseas performance: A quasi-experiment based on a staggered DID model. *Economic Analysis and Policy*, 2025, vol. 88, pp. 1453-1466, DOI: 10.1016/j.eap.2025.10.023.
10. Pollitt M.G. The European Single Market in Electricity: An Economic Assessment. *Review of Industrial Organization*, 2019, vol. 55(1), pp. 63-87, DOI: 10.1007/s11151-019-09682-w.
11. Гительман Л.Д. Инструменты устойчивого развития региональной энергетики / Л.Д. Гительман, В.В. Добродей, М.В. Кожевников // Экономика региона, 2020. – Т. 16. – № 4. – С. 1208-1223. – DOI: 10.17059/ekon.reg.2020-4-14.
12. Сухарева Е.В. Декарбонизационная политика России в условиях ESG-трансформации мировой энергетики / Е.В. Сухарева, М.В. Кахальников // Общество: политика, экономика, право, 2025. – № 3. – С. 25-33. – DOI: 10.24158/per.2025.3.2.
13. Nosko P. Analysis of ESG Ratings' Methodologies in the Russian Market: Transparency and Convergence Issues. *Contemporary World Economy*, 2024, vol. 2(2), pp. 60-75, DOI: 10.17323/2949-5776-2024-2-2-60-75.
14. Довбий И.П. Финансовые и экономические условия энергоперехода для национальной экономики / И.П. Довбий // Financial Journal, 2022. – Т. 14. – № 5. – С. 25-42. – DOI: 10.31107/2075-1990-2022-5-25-42.
15. Abido M.A. Multiobjective particle swarm optimization for environmental/economic dispatch problem. *Electric Power Systems Research*, 2009, vol. 79(7), pp. 1105-1113, DOI: 10.1016/j.epsr.2009.02.005.
16. Lombardi P., Sokolnikova T., Suslov K., Voropai N., et al. Isolated power system in Russia: A chance for renewable energies? *Renewable Energy*, 2016, vol. 90, pp. 532-541, DOI: 10.1016/j.renene.2016.01.016.
17. Тихомиров А.А. Влияние ESG факторов на эффективность и инвестиционную привлекательность компаний / А.А. Тихомиров, Д.Х. Харчилава // Вестник Алтайской академии экономики и права, 2024. – Т. 2. – № 9. – С. 312-318. – DOI: 10.17513/vaael.3738.
18. Воропай Н.И. Направления и проблемы трансформации электроэнергетических систем / Н.И. Воропай // Электричество, 2020. – Т. 7. – № 7. – С. 12-21. – DOI: 10.24160/0013-5380-2020-7-12-21.
19. Browne D., O'Regan B., Moles R. Use of multi-criteria decision analysis to explore alternative domestic energy and electricity policy scenarios in an Irish city-region☆. *Energy*, 2010, vol. 35(2), pp. 518-528, DOI: 10.1016/j.energy.2009.10.020.
20. Munda G. Social multi-criteria evaluation for urban sustainability policies. *Land Use Policy*, 2006, vol. 23(1), pp. 86-94, DOI: 10.1016/j.landusepol.2004.08.012.
21. Шакиров В.А. Методика многокритериального двухуровневого анализа пунктов размещения электростанций / В.А. Шакиров, П.С. Панкратьев // Искусственный интеллект и принятие решений, 2017. – № 1. – С. 69-83.
22. Шутов И.В. Раскрытие сущности категории «Энергетическая безопасность» на примере обеспечения устойчивого и доступного энергообеспечения изолированных районов Дальнего Востока и Арктики / И.В. Шутов // Индустриальная экономика, 2022. – Т. 3. – № 4. – С. 259-263. – DOI: 10.47576/2712-7559_2022_4_3_259.
23. Шакиров В.А. Многокритериальный выбор состава оборудования гибридного энергокомплекса с использованием генетического алгоритма / В.А. Шакиров, И.Н. Погодаева // Системы. Методы. Технологии, 2024. – Т. 62. – № 2. – С. 61-67. – DOI: 10.18324/2077-5415-2024-2-61-67.
24. Gent M., Lamont J. Minimum-Emission Dispatch. *IEEE Transactions on Power Apparatus and Systems*, 1971, vol. PAS-90(6), pp. 2650-2660, DOI: 10.1109/TPAS.1971.292918.
25. El-Keib A.A., Ma H., Hart J.L. Environmentally constrained economic dispatch using the LaGrangian relaxation method. *IEEE Transactions on Power Systems*, 1994, vol. 9(4), pp. 1723-1729, DOI: 10.1109/59.331423.
26. Belton V., Stewart T.J. Multiple Criteria Decision Analysis. Boston, MA: Springer US, 2002, DOI: 10.1007/978-1-4615-1495-4
27. Wüstenhagen R., Wolsink M., Bürer M.J. Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy*, 2007, vol. 35(5), pp. 2683-2691, DOI: 10.1016/j.enpol.2006.12.001.
28. Deb K., Pratap A., Agarwal S., Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 2002, vol. 6(2), pp. 182-197, DOI: 10.1109/4235.996017.
29. Zio E., Baraldi P., Pedroni N. Optimal power system generation scheduling by multi-objective genetic algorithms with preferences. *Reliability Engineering & System Safety*, 2009, vol. 94(2), pp. 432-444, DOI: 10.1016/j.ress.2008.04.004.

30. Marzbani F., Abdelfatah A. Economic Dispatch Optimization Strategies and Problem Formulation: A Comprehensive Review. *Energies*, 2024, vol. 17(3), pp. 550, DOI: 10.3390/en17030550.
31. Nyingu B.T., Masike L., Mbukani M.W.K. Multi-Objective Optimization of Load Flow in Power Systems: An Overview. *Energies*, 2025, vol. 18(22), pp. 6056, DOI: 10.3390/en18226056.
32. Keeney R.L. Siting energy facilities. New York, N.Y: Academic Press, 1980, 413 p.
33. Keeney R.L., Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. Cambridge New York Melbourne: Cambridge University Press, 1993, 569 p.
34. Isihak S., Akpan U., Bhattacharyya S. Prioritizing projects in rural electrification planning: Integrating MAUT with OnSSET. *Energy Strategy Reviews*, 2025, vol. 61, pp. 101807, DOI: 10.1016/j.esr.2025.101807.
35. Winterfeldt D. von, Edwards W. Decision analysis and behavioral research. Cambridge: Cambridge Univ. Press, 1986, 624 p.
36. Edwards W., Barron F.H. SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement. *Organizational Behavior and Human Decision Processes*, 1994, vol. 60(3), pp. 306-325, DOI: 10.1006/obhd.1994.1087.
37. Saaty T.L. Multicriteria decision making: the analytic hierarchy process; planning, priority setting, resource allocation. New York: McGraw-Hill, 1980, 287 p.
38. Bertsekas D.P. Constrained optimization and Lagrange multiplier methods. Belmont, Mass: Athena Scientific, 1996, 395 p.
39. Kuhn H.W., Tucker A.W. Nonlinear programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 481–492.
40. Васильев Ф.П. Методы оптимизации / Ф.П. Васильев. – Москва: Издательство «Факториал Пресс», 2002. – 824 с.

Панкратьев Павел Сергеевич. Кандидат технических наук, доцент, доцент кафедры энергетики Братского государственного университета, AuthorID: 721258, SPIN: 9514-3134, ORCID: 0000-0002-0921-7842, scud33@inbox.ru. 665709, г. Братск, ул. Макаренко, д. 40.

UDC 519.816: 519.863: 621.311.16

DOI:10.25729/ESI.2026.43.3.009

Development of a hybrid method for optimal electric load distribution considering hard-to-formalize factors

Pavel S. Pankratev

Bratsk state university, Russia, Bratsk, scud33@inbox.ru

Abstract. Optimal distribution of electric loads between energy sources is a crucial process enabling the minimization of financial costs for the energy supplier at the current demand level. This process, internationally known as economic dispatch, is traditionally built on the principle of equality of incremental fuel costs, which allows for determining the optimal power allocation in terms of achieving economic efficiency. However, the modern state of the energy sector, characterized by global ESG transformation and the implementation of Corporate Social Responsibility (CSR) principles, requires considering not only direct economic indicators but also socio-environmental ones. The latter are often hard to formalize or entirely non-formalizable, introducing significant complexity to the optimization process. These conditions necessitate the involvement of a Decision Maker (DM) with their inherent internal value system. Under these circumstances, the traditional cost-minimization approach becomes insufficient, dictating the need for new hybrid methods. This article proposes a hybrid method for power allocation based on the synthesis of the classical Lagrange multiplier method, generalized by Karush–Kuhn–Tucker (KKT) conditions, and Multi-Attribute Utility Theory (MAUT). The scientific novelty of the approach lies in the direct integration of a multi-attribute value function, reflecting the DM's subjective preferences, into the structure of the optimization problem. To validate the developed hybrid method, an isolated power system with a total load of 500 MW was modeled. The optimization problem in a static formulation was solved for two scenarios: a baseline scenario (minimizing fuel costs) and a socially-oriented scenario (considering DM participation with an active constraint on the value function level $V \geq 0.60$). The results showed that meeting the DM requirements at a value level of 0.60 leads to a substantial reconfiguration of the energy balance,

substituting 43% of base-load coal generation with gas power capacity. It is demonstrated that the resulting cost increase (by 22%) is compensated by achieving the required level of social value. The calculated values of dual variables (Lagrange multipliers) are interpreted as "shadow prices" of non-formalizable constraints, reflecting the marginal cost of improving the socio-environmental situation.

Keywords: optimal load distribution, economic dispatch, system analysis, Multi-Attribute Utility Theory (MAUT), Lagrange multiplier method, Karush-Kuhn-Tucker conditions, shadow price, decision making

References

1. Wood A.J., Wollenberg B.F., Sheblé G.B. Power generation, operation and control. Hoboken: IEEE Wiley, 2014, 658 p.
2. Happ H.H. Optimal power dispatch – A comprehensive survey. IEEE Transactions on Power Apparatus and Systems, 1977, vol. 96(3), pp. 841-854, DOI: 10.1109/T-PAS.1977.32397.
3. Abido M.A. Environmental/economic power dispatch using multiobjective evolutionary algorithms. IEEE Transactions on Power Systems, 2003, vol. 18(4), pp. 1529-1537, DOI: 10.1109/TPWRS.2003.818693.
4. Talaq J.H., El-Hawary F., El-Hawary M.E. A summary of environmental/economic dispatch algorithms. IEEE Transactions on Power Systems, 1994, vol. 9(3), pp. 1508-1516, DOI: 10.1109/59.336110.
5. Voropai N.I. Multi-Criteria Problems in Electric Power System Expansion Planning. Energy Systems Research, 2018, vol. 1(2), pp. 27-34, DOI: 10.25729/esr.2018.02.0003.
6. Gornshtein V.M., Miroshnichenko B.P., Ponomarev A.V. Metody optimizatsii rezhimov energosistem [Methods of optimization of power system operating modes]. Moscow: Energiya, 1981, 336 p.
7. Kapranova L.D., Polishchuk O.A. ESG-transformatsiya kak paradigma ustoychivogo razvitiya ekonomiki Rossii [ESG transformation as a paradigm of sustainable development of the Russian economy]. Vestnik Evraziyskoy nauki [Bulletin of Eurasian Science], 2023, vol. 15, no. 3, pp. 1–17.
8. Kodaneva S.I. Energeticheskiy perekhod: perspektivy i mekhanizmy realizatsii [Energy transition: prospects and implementation mechanisms]. Rossiya i sovremennyy mir [Russia and the modern world], 2022, no. 4, pp. 162–183, DOI: 10.31249/rsm/2022.04.09.
9. Zhang L., Xu S., Chen J., Li H., et al. The impact of ESG ratings on corporate overseas performance: A quasi-experiment based on a staggered DID model. Economic Analysis and Policy, 2025, vol. 88, pp. 1453-1466, DOI: 10.1016/j.eap.2025.10.023.
10. Pollitt M.G. The European Single Market in Electricity: An Economic Assessment. Review of Industrial Organization, 2019, vol. 55(1), pp. 63-87, DOI: 10.1007/s11151-019-09682-w.
11. Gitelman L.D., Dobrodey V.V., Kozhevnikov M.V. Instrumenty ustoychivogo razvitiya regional'noy energetiki [Tools for sustainable development of regional energy]. Ekonomika regiona [Economy of region], 2020, vol. 16, no. 4, pp. 1208-1223, DOI: 10.17059/ekon.reg.2020-4-14.
12. Sukhareva E.V., Kakhal'nikov M.V. Dekarbonizatsionnaya politika Rossii v usloviyakh ESG-transformatsii mirovoy energetiki [Russia's decarbonization policy in the context of ESG transformation of global energy]. Obshchestvo: politika, ekonomika, pravo [Society: politics, economics, law], 2025, no. 3, pp. 25-33, DOI: 10.24158/pep.2025.3.2.
13. Nosko P. Analysis of ESG Ratings' Methodologies in the Russian Market: Transparency and Convergence Issues. Contemporary World Economy, 2024, vol. 2(2), pp. 60-75, DOI: 10.17323/2949-5776-2024-2-2-60-75
14. Dovbiy I.P. Finansovyye i ekonomicheskiye usloviya energoperekhoda dlya natsional'noy ekonomiki [Financial and economic conditions of energy transition for the national economy]. Financial Journal, 2022, vol. 14, no. 5, pp. 25–42, DOI: 10.31107/2075-1990-2022-5-25-42.
15. Abido M.A. Multiobjective particle swarm optimization for environmental/economic dispatch problem. Electric Power Systems Research, 2009, vol. 79(7), pp. 1105-1113, DOI: 10.1016/j.epsr.2009.02.005.
16. Lombardi P., Sokolnikova T., Suslov K., Voropai N., et al. Isolated power system in Russia: A chance for renewable energies? Renewable Energy, 2016, vol. 90, pp. 532-541, DOI: 10.1016/j.renene.2016.01.016.
17. Tikhomirov A.A., Kharchilava D.Kh. Vliyaniye ESG faktorov na effektivnost' i investitsionnyuyu privlekatel'nost' kompaniy [The impact of ESG factors on the efficiency and investment attractiveness of companies]. Vestnik Altayskoy akademii ekonomiki i prava [Bulletin of the Altai Academy of Economics and Law], 2024, vol. 2, no. 9, pp. 312–318, DOI: 10.17513/vaael.3738.
18. Voropai N.I. Napravleniya i problemy transformatsii elektroenergeticheskikh sistem [Directions and problems of transformation of electric power systems]. Elektrichestvo [Electricity], 2020, vol. 7, no. 7, pp. 12–21, DOI: 10.24160/0013-5380-2020-7-12-21.
19. Browne D., O'Regan B., Moles R. Use of multi-criteria decision analysis to explore alternative domestic energy and electricity policy scenarios in an Irish city-region☆. Energy, 2010, vol. 35(2), pp. 518-528, DOI: 10.1016/j.energy.2009.10.020.

20. Munda G. Social multi-criteria evaluation for urban sustainability policies. *Land Use Policy*, 2006, vol. 23(1), pp. 86–94, DOI: 10.1016/j.landusepol.2004.08.012.
21. Shakirov V.A., Pankrat'yev P.S. Metodika mnogokriterial'nogo dvukhurovneвого analiza punktov razmeshcheniya elektrostantsiy [Methodology of multi-criteria two-level analysis of power plant locations]. *Iskusstvennyy intellekt i prinyatiye resheniy* [Artificial intelligence and decision making], 2017, no. 1, pp. 69–83.
22. Shutov I.V. Raskrytiye sushchnosti kategorii "Energeticheskaya bezopasnost'" na primere obespecheniya ustoychivogo i dostupnogo energoobespecheniya izolirovannykh rayonov Dal'nego Vostoka i Arktiki [Disclosure of the essence of the category "Energy security" on the example of ensuring sustainable and affordable energy supply to isolated areas of the Far East and the Arctic]. *Industrial'naya ekonomika* [Industrial economics], 2022, vol. 3, no. 4, pp. 259–263, DOI:10.47576/2712-7559_2022_4_3_259.
23. Shakirov V.A., Pogodayeva I.N. Mnogokriterial'nyy vybor sostava oborudovaniya gibridnogo energokompleksa s ispol'zovaniyem geneticheskogo algoritma [Multi-criteria selection of equipment composition for a hybrid energy complex using a genetic algorithm]. *Sistemy. Metody. Tekhnologii* [Systems. Methods. Technologies], 2024, vol. 62, no. 2, pp. 61–67, DOI:10.18324/2077-5415-2024-2-61-67.
24. Gent M., Lamont J. Minimum-Emission Dispatch. *IEEE Transactions on Power Apparatus and Systems*, 1971, vol. PAS-90(6), pp. 2650–2660, DOI:10.1109/TPAS.1971.292918.
25. El-Keib A.A., Ma H., Hart J.L. Environmentally constrained economic dispatch using the LaGrangian relaxation method. *IEEE Transactions on Power Systems*, 1994, vol. 9(4), pp. 1723–1729, DOI:10.1109/59.331423.
26. Belton V., Stewart T.J. Multiple Criteria Decision Analysis. Boston, MA: Springer US, 2002, DOI:10.1007/978-1-4615-1495-4.
27. Wüstenhagen R., Wolsink M., Bürer M.J. Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy*, 2007, vol. 35(5), pp. 2683–2691, DOI:10.1016/j.enpol.2006.12.001.
28. Deb K., Pratap A., Agarwal S., Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 2002, vol. 6(2), pp. 182–197, DOI:10.1109/4235.996017.
29. Zio E., Baraldi P., Pedroni N. Optimal power system generation scheduling by multi-objective genetic algorithms with preferences. *Reliability Engineering & System Safety*, 2009, vol. 94(2), pp. 432–444, DOI:10.1016/j.res.2008.04.004.
30. Marzbani F., Abdelfatah A. Economic Dispatch Optimization Strategies and Problem Formulation: A Comprehensive Review. *Energies*, 2024, vol. 17(3), pp. 550, DOI:10.3390/en17030550.
31. Nyingu B.T., Masike L., Mbukani M.W.K. Multi-Objective Optimization of Load Flow in Power Systems: An Overview. *Energies*, 2025, vol. 18(22), pp. 6056, DOI:10.3390/en18226056.
32. Keeney R.L. Siting energy facilities. New York, N.Y.: Academic Press, 1980, 413 p.
33. Keeney R.L., Raiffa H. Decisions with multiple objectives: preferences and value tradeoffs. Cambridge New York Melbourne: Cambridge University Press, 1993, 569 p.
34. Isihak S., Akpan U., Bhattarcharya S. Prioritizing projects in rural electrification planning: Integrating MAUT with OnSSET. *Energy Strategy Reviews*, 2025, vol. 61, pp. 101807, DOI:10.1016/j.esr.2025.101807.
35. Winterfeldt D. von, Edwards W. Decision analysis and behavioral research. Cambridge: Cambridge Univ. Press, 1986, 624 p.
36. Edwards W., Barron F.H. SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement. *Organizational Behavior and Human Decision Processes*, 1994, vol. 60(3), pp. 306–325, DOI:10.1006/obhd.1994.1087.
37. Saaty T.L. Multicriteria decision making: the analytic hierarchy process ; planning, priority setting, resource allocation. New York: McGraw-Hill, 1980, 287 p.
38. Bertsekas D.P. Constrained optimization and Lagrange multiplier methods. Belmont, Mass: Athena Scientific, 1996, 395 p.
39. Kuhn H.W., Tucker A.W. Nonlinear programming. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, 1951, pp. 481–492.
40. Vasil'yev F.P. Metody optimizatsii [Optimization methods]. Moscow: Izdatel'stvo "Faktorial Press", 2002, 824 p.

Pankratev Pavel Sergeevich. Ph.D., Associate Professor, Associate Professor of the Department of Power Engineering, Bratsk State University. AuthorID: 721258, SPIN: 9514-3134, ORCID: 0000-0002-0921-7842, scud33@inbox.ru. 665709, Bratsk, 40 Makarenko St.

Статья поступила в редакцию 02.03.2026; одобрена после рецензирования 17.03.2026; принята к публикации 01.08.2026.

The article was submitted 03/02/2026; approved after reviewing 03/17/2026; accepted for publication 08/01/2026.

УДК 004.94:621.311.245:004.82

DOI:10.25729/ESI.2026.43.3.010

Разработка цифровой модели цифрового двойника ветровой электростанции с использованием онтологического моделирования

Шукин Никита Игоревич

Институт систем энергетики им. Л.А. Мелентьева СО РАН,

Россия, Иркутск, niksha14@mail.ru

Аннотация. В статье рассматривается задача интеграции разнородных математических моделей, применяемых при формировании цифровой модели цифрового двойника ветровой электростанции. Актуальность исследования определяется тем, что ветровая электростанция представляет собой не один объект, а сложную техническую систему, включающую ряд взаимосвязанных подсистем: ветровую установку, аэродинамический контур, механическую передачу, силовую электронику, систему передачи энергии, дизельную электростанцию, аккумуляторные батареи и потребителей. Для каждой из указанных подсистем в научной литературе используются различные классы моделей, отличающиеся составом исходных данных, уровнем детализации, расчетным шагом, принятыми допущениями, ограничениями и областью применимости. При разработке цифрового двойника такие модели должны быть не только реализованы в одном программном комплексе, но и согласованы между собой по смыслу параметров, единицам измерения, временным характеристикам, граничным условиям и правилам передачи результатов расчетов. Цель работы состоит в обосновании целесообразности применения онтологического подхода к интеграции моделей в цифровой модели цифрового двойника ветровой электростанции. Показано, что при отсутствии явной формализации предметной области усложнение цифрового двойника сопровождается дублированием данных, появлением неявных зависимостей между расчетными модулями, затруднением сопровождения программной системы и ростом затрат на развитие новой функциональности. Предлагается использовать онтологическую модель, как средство ограничения этих эффектов, задающее общий слой терминов, связей, ограничений и требований к данным. Целесообразно применять такую модель на этапе проектирования новой функциональности цифрового двойника, включая внедрение новых расчетных моделей, изменение компонентов визуализации, расширение базы данных и уточнение связей между подсистемами. Новизна подхода заключается в рассмотрении онтологии не только как средства представления знаний о предметной области, но и как инструмента снижения организационных и программно-инженерных рисков при интеграции моделей цифрового двойника. Отдельное внимание уделено тому, что онтологическая модель позволяет фиксировать знания специалистов в формализованном виде и тем самым снижать зависимость проекта от ограниченного числа сотрудников, одновременно владеющих программной инженерией, математическим моделированием и спецификой энергетических объектов. В работе обосновывается, что использование онтологического подхода может способствовать повышению согласованности компонентов цифровой модели, снижению вероятности ошибок при обмене данными между расчетными модулями и упрощению последующего развития цифрового двойника ветровой электростанции.

Ключевые слова: цифровой двойник, ветровая электростанция, математическое моделирование, онтологическое моделирование, интеграция моделей

Цитирование: Шукин Н.И. Разработка цифровой модели цифрового двойника ветровой электростанции с использованием онтологического моделирования / Н.И. Шукин // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 119-133. – DOI:10.25729/ESI.2026.43.3.010.

Введение. Развитие цифровых двойников энергетических объектов требует перехода от отдельных библиотек расчетных функций к комплексным цифровым моделям, которые способны отражать состав объекта, его режимы, состояние оборудования, внешние воздействия и управленческое воздействие. Для ветровой электростанции производство электрической энергии зависит от стохастического ветрового потока, состояния механических узлов ветровой установки, характеристик генератора и преобразователей, режима аккумуляторных батарей, участия резервной дизельной электростанции и графика потребления. Поэтому цифровой двойник ветровой электростанции не может быть построен как одна универсальная модель. Он должен объединять набор моделей разного назначения и разной природы.

Научная литература демонстрирует значительное количество публикаций по моделированию аэродинамики ветровых установок, кривых мощности, механики привода, электрических машин, силовой электроники, систем накопления энергии, дизель-генераторных установок и нагрузок [1–10]. Однако при переходе от отдельных моделей к цифровому двойнику возникает иной уровень задачи. Разработчику требуется согласовать модели между собой: определить, какие параметры являются общими, какие данные передаются между модулями, какие величины должны храниться в информационной модели, а какие вычисляются по требованию. При отсутствии формального описания предметной области эти связи часто фиксируются в коде, технических комментариях или в неформальных договоренностях между специалистами.

Такой способ разработки допустим на этапе раннего прототипирования, но становится рискованным при расширении цифрового двойника. В него добавляются новые модели, новые источники данных, новые сценарии расчета, новые средства визуализации, новые возможности по управлению объектом и новые требования к мониторингу. По мере роста функциональности увеличивается вероятность появления дублирующихся сущностей, противоречивых трактовок параметров, несовместимых единиц измерения и повторной реализации одинаковых расчетов в разных подсистемах. Для программных систем, постоянно взаимодействующих с изменяющейся предметной областью, эта проблема хорошо согласуется с законами эволюции программного обеспечения: без специальных мер сложность системы возрастает, а ее качество со временем снижается [11, 12].

В статье предлагается рассматривать онтологический подход, как средство интеграции математических моделей в цифровую модель цифрового двойника ветровой электростанции. В таком подходе онтология задает не только словарь предметной области, но и структуру согласования расчетных моделей, данных, технических требований и программных компонентов. Это позволяет переносить часть неявных знаний экспертов из неформального общения и документации в формализованную модель, доступную для анализа, повторного использования и контроля согласованности.

1. Модели подсистем ветровой электростанции. Ветровая электростанция включает несколько групп объектов, которые различаются по физической природе, временным масштабам и целям моделирования. При проектировании цифрового двойника важно не сводить эти модели к единой детализации. Наоборот, цифровая модель должна поддерживать разные уровни описания: от расчетов годовой выработки по кривой мощности до динамического моделирования переходных процессов в электрической части.

Наиболее общая расчетная зависимость для ветровой установки:

$$P = \xi \cdot \pi \cdot R^2 \cdot \frac{\rho \cdot V^3}{2} \cdot \eta_{\text{ред}} \cdot \eta_{\text{ген}} \quad (1)$$

где ξ – коэффициент использования энергии ветра, R – радиус ротора, ρ – плотность воздуха, V – скорость ветра, $\eta_{\text{ред}}$ – КПД редуктора, $\eta_{\text{ген}}$ – КПД генератора. Эта формула не использует аэродинамическую модель, но задает связь между внешними условиями и механическим входом ветровой установки. В прикладных расчетах она дополняется моделями ветрового профиля, турбулентности, сдвига и поворота ветра, влияния плотности воздуха, ограничений по пусковой, номинальной и предельной скорости [1, 2].

Внешние части ветровой установки – ротор, лопасти, ступица, гондола, система ориентации и башня – обычно моделируются через аэродинамические, прочностные и кинематические зависимости. Для задач энергетического расчета часто достаточно паспортной или параметрической кривой мощности, отражающей связь скорости ветра и электрической мощности. Для задач диагностики, управления и ресурса оборудования требуются более детальные модели: распределение нагрузок по лопасти, изгибные и

крутильные колебания, воздействие турбулентности, состояние системы регулирования угла лопасти и ориентации ротора по направлению ветра. Исследование [2] показывает, что кривые мощности зависят не только от конструктивных параметров ветровой установки, но и от внешних условий: турбулентности, плотности воздуха, сдвига и поворота ветра. Следовательно, цифровой двойник должен хранить не только паспортные параметры агрегата, но и описание условий, при которых они применимы.

Механическая часть ветровой установки связывает аэродинамический момент ротора с электромагнитным моментом генератора. В простейшем случае она описывается уравнением движения

$$J \frac{d\omega}{dt} = M - M_{\text{мех}}. \quad (2)$$

где J – осевой момент инерции вращающихся масс роторов генератора и турбины, ω – угловая скорость, M – момент, действующий на ротор генератора, $M_{\text{мех}}$ – момент сопротивления механизма, присоединенного к ротору двигателя.

Для ветровых установок с редуктором требуется учет передаточного отношения, упругости вала, демпфирования и динамики редуктора. Для прямоприводных установок акцент переносится на генератор и силовой преобразователь, поэтому механическая модель является связующим звеном между аэродинамикой и электрической частью, а не отдельным расчетом, который можно изолировать от остальных подсистем.

Ветрогенератор и электрическая часть могут быть представлены моделями асинхронного генератора, генератора двойного питания, синхронного генератора с постоянными магнитами, выпрямителя, инвертора, звена постоянного тока, трансформаторов и устройств защиты. В работах по динамическому моделированию ветровых установок показано, что даже обобщенная модель переменнo скоростной установки должна включать блоки аэродинамики, привода, генератора, электрического управления и сетевого интерфейса [3]. Для генераторов с постоянными магнитами важны модели преобразователя, управления активной и реактивной мощностью, поддержания напряжения звена постоянного тока и режимов максимального отбора мощности [4, 5].

Передача энергии от ветровой установки к потребителю или внешней сети описывается моделями кабельных линий, трансформаторов, коммутационной аппаратуры, потерь, ограничений по напряжению, частоте и реактивной мощности. Для изолированных систем и гибридных микросетей этот блок нельзя опускать, поскольку малые изменения нагрузки или генерации могут приводить к отклонению частоты и напряжения. В цифровом двойнике модель передачи энергии должна быть связана с информацией о топологии схемы, состояниях оборудования в схеме, допустимых токах, настройках защит и составе работающего оборудования.

Дизельная электростанция обычно рассматривается как резервный или балансирующий источник в составе автономной ветродизельной системы. Ее модель включает дизельный двигатель, синхронный генератор, регулятор скорости, регулятор возбуждения, ограничения по минимальной нагрузке, расход топлива, время пуска и режимы совместной работы с ветровой генерацией. В классических и современных работах по ветродизельным и ветродизель-аккумуляторным системам дизельный генератор рассматривается не только как источник мощности, но и как элемент поддержания баланса между генерацией и потреблением [6, 7]. Поэтому в цифровой модели необходимо фиксировать не только установленную мощность дизельной электростанции, но и логику включения, ограничения экономичной работы и связь с аккумуляторами.

Аккумуляторные батареи выполняют функции сглаживания ветровой генерации, покрытия кратковременного дефицита, накопления избытков энергии и уменьшения числа

пусков дизельной электростанции. Их модели различаются по уровню детализации: от простого энергетического баланса до эквивалентных электрических схем с учетом внутреннего сопротивления, поляризации, напряжения холостого хода, температуры, состояния заряда и деградации. В динамических расчетах часто используется переменное состояние заряда аккумуляторной батареи, которое связывает ток батареи, номинальную емкость, эффективность заряда и разряда. Обобщенная запись может быть представлена как:

$$SOC(t) = SOC(t_0) + \frac{1}{Q_{max}} \int_{t_0}^t I(t) dt. \quad (3)$$

где $SOC(t_0)$ – начальный уровень заряда в момент времени t_0 , Q_{max} – номинальная емкость аккумуляторной батареи, $I(t)$ – ток батареи в момент времени t .

Работы по валидации динамических моделей аккумуляторов показывают, что даже относительно простые модели могут быть пригодны для имитационного моделирования при корректном определении параметров по данным производителя и испытаний [8].

Потребители в составе автономной или локальной ветровой электростанции моделируются не только через суммарный график нагрузки. Для оценки статических режимов используют модели постоянной мощности, постоянного тока, постоянного сопротивления и их комбинации. Для динамических расчетов учитывают асинхронные двигатели, электронные преобразователи, нагревательные нагрузки, пусковые токи и восстановление нагрузки после провалов напряжения. Стандартные подходы к моделированию нагрузки в расчетах потокораспределения и динамической устойчивости предлагают представлять нагрузку как набор статических и динамических компонентов [9, 10]. Следовательно, в цифровом двойнике потребитель должен описываться не только величиной спроса, но и типом нагрузки, приоритетом отключения, допустимыми отклонениями качества электроэнергии и характером изменения во времени.

2. Цифровая модель цифрового двойника. Цифровой двойник отличается от обычной имитационной модели тем, что виртуальное представление объекта тесно связано с реальным объектом. В литературе цифровой двойник рассматривается, как модель сложной физической системы, которая использует данные, физические и статистические модели для анализа, прогнозирования и поддержки управления [13–16]. Для ветровой электростанции это означает, что цифровая модель должна объединять модели подсистем не как справочник, а как взаимосвязанная и настраиваемая система расчета.

В цифровой модели цифрового двойника можно выделить несколько уровней. Первый уровень – интерфейсы связи с информационными моделями, то есть принятый для всех участников разработки подход к реализации связей с компонентами цифровой тени цифрового двойника (цифровая тень представляет собой систему связей и зависимостей, приближенно описывающих поведение реального объекта/продукта, как правило, в нормальных условиях работы и содержащихся в избыточных больших данных, которые получают от реального объекта/продукта при помощи большого количества датчиков и технологий промышленного интернета, изменение физического объекта/продукта приводит к изменению его цифровой тени, но не наоборот [17]).

Цифровая тень содержит сведения о составе объекта: ветровые установки, дизельные генераторы, аккумуляторные батареи, трансформаторы, устройства управления, датчики, потребители и связи между ними. Второй уровень – расчетный. На нем размещаются математические модели, алгоритмы расчета необходимых характеристик, прогнозирования, оценки состояния и оптимизации управления. Третий уровень – сценарный. Он описывает условия применения моделей: расчет годовой выработки, моделирование переходного процесса, анализ отказа, подбор мощности батарей, оценка топливного расхода дизельной

электростанции. Четвертый уровень – интерфейсный. Он обеспечивает визуализацию состояния, результатов расчетов и рекомендаций.

Проблема состоит в том, что эти уровни обычно разрабатываются разными специалистами. Математик описывает расчетную зависимость и набор параметров. Программист проектирует программный код и API. Инженер-энергетик контролирует корректность допущений. Специалист по визуализации определяет, какие значения должны отображаться пользователю. Если общая модель предметной области отсутствует, каждый участник команды создает собственное представление объекта. Например, скорость ветра может существовать, как поле в таблице погодных данных, как параметр функции расчета мощности, как ось графика, как входной сигнал модели управления и как атрибут в окне визуализации. При неформальном согласовании высока вероятность того, что эти представления начнут расходиться по размерности и временному шагу.

Для цифрового двойника ветровой электростанции особенно важна прозрачность данных. Если расчет коэффициента используемой установленной мощности (КИУМ) или прогноза выработки опирается на скорость ветра, разработчик должен понимать, откуда получены данные, какой высоте они соответствуют, как обработаны и какая модель ветровой установки использована. Аналогично для батареи требуется знать, является ли состояние заряда аккумуляторной батареи измеренной величиной, результатом модели или прогнозным значением. Для дизельной электростанции нужно различать установленную мощность, доступную мощность, минимально допустимую нагрузку и фактическую нагрузку. Эти различия очевидны для предметного специалиста, но не всегда очевидны в структуре программной системы.

Таким образом, цифровая модель цифрового двойника должна выполнять интеграционную функцию. Она связывает сущности предметной области, параметры моделей, источники данных, сценарии расчета и программные модули. Если эта функция реализуется только на уровне кода, то знания оказываются распределенными между исходными файлами, базой данных, документацией и опытом отдельных сотрудников. При расширении проекта такая организация становится причиной ускоренного накопления хаоса в программной системе.

3. Сложности интеграции моделей при программной реализации. Интеграция моделей в цифровом двойнике отличается от простой сборки модулей. Каждый расчетный модуль имеет собственную область применимости, набор входов и выходов, ограничения и требования к качеству данных. При этом модули должны работать совместно: аэродинамическая модель формирует механическую мощность, механическая модель передает момент генератору, электрическая модель рассчитывает мощность на шинах, модель батареи компенсирует дисбаланс, дизельная электростанция включается при дефиците, а модель нагрузки задает спрос. Даже если каждая модель по отдельности корректна, цифровой двойник может работать неверно из-за ошибок согласования между ними.

К типовым проблемам относятся: дублирование данных в разных подсистемах; неодинаковые имена и разные единицы измерения для одних и тех же величин; неявные преобразования данных; отсутствие связи между параметром в техническом задании и его реализацией в коде; смешение измеренных, расчетных и прогнозных значений; несогласованное обновление справочников оборудования; отсутствие единого описания ограничений. Например, паспортная кривая мощности может храниться в расчетном модуле как массив точек, в базе данных как таблица, в интерфейсе как график, а в документации как изображение. Если при обновлении оборудования изменится только одна копия, цифровой двойник начнет выдавать противоречивые результаты.

Отдельная проблема связана с развитием цифрового двойника. На начальном этапе команда может реализовать расчет выработки одной ветровой установки. Затем добавляется учет нескольких установок, батарей, дизельной электростанции, потребителей, сценариев аварий, визуализации, импорта данных SCADA, прогноза погоды и оптимизации режимов. Каждая новая функция требует новых сущностей и связей. Если архитектура не содержит явного слоя предметной семантики, разработчики вынуждены искать нужные значения в существующем коде, повторно согласовывать смысл терминов и вручную контролировать совместимость модулей.

В терминах программной инженерии такая ситуация приводит к росту сложности, связанности и стоимости изменений. Закон возрастающей сложности программной эволюции указывает, что изменяемые программные системы со временем становятся сложнее, если специально не выполнять работы по снижению сложности [11, 12]. Цифровой двойник энергетического объекта относится именно к таким системам: он должен изменяться вместе с объектом, данными, расчетными методами и требованиями пользователей. Поэтому задача состоит не в том, чтобы полностью исключить рост сложности, а в том, чтобы сделать его управляемым.

Управляемость может быть достигнута за счет переноса части знаний в формализованный слой. В цифровом двойнике таким слоем может выступать онтология предметной области, связанная с онтологиями сопутствующих объектов, онтологией информационной модели и онтологией математической модели. Эти онтологии позволяют зафиксировать, что «скорость ветра на высоте ступицы», «скорость ветра на высоте измерения», «среднегодовая скорость ветра» и «входной параметр кривой мощности» являются связанными, но не тождественными сущностями. Аналогично можно различать установленную мощность ветровой установки, доступную мощность, фактическую мощность и прогнозную мощность. Такое различие снижает вероятность неявной подмены понятий при разработке.

4. Онтологический подход к интеграции моделей. Онтология в рассматриваемой задаче понимается как формализованное описание классов объектов, их свойств, отношений, ограничений и правил интерпретации данных. Языки семейства Semantic Web, включая OWL, ориентированы на представление классов, свойств, индивидов и значений данных, а также на проверку согласованности и вывод неявных знаний [18]. В рамках цифрового двойника ветровой электростанции онтология может использоваться не только как справочник терминов, но и как проектный артефакт, связывающий предметную область с архитектурой программной системы.

Предлагаемый подход включает несколько взаимосвязанных онтологических слоев. Предметная онтология описывает физические объекты: ветровая установка, ротор, лопасть, генератор и так далее. Онтология математической и информационной моделей фиксирует, какие модели существуют, какие входные и выходные параметры они используют и откуда их берут, для каких задач применимы и какие допущения имеют. Онтология сопутствующих объектов может подразделяться на онтологию сценариев, описывающую расчет выработки, оптимизацию расположения, моделирование переходного процесса, прогнозирование, диагностику, расчет энергобаланса, оценку экономичности и на онтологию программных компонентов, которая связывает сущности других онтологий с расчетными сервисами, API и интерфейсами.

Практический смысл такого разделения состоит в том, что при добавлении новой модели команда не начинает с обсуждения структуры таблиц или сигнатуры функции. Сначала определяется место модели в онтологии: какие объекты она описывает, какие параметры требует, какие результаты возвращает, какие данные являются измеренными, а какие

расчетными, какие ограничения должны выполняться. После этого проектируются изменения в базе данных, расчетном модуле и интерфейсе. С учетом этого онтология используется на ранней стадии – при подготовке технического задания и проектировании новой функциональности. Время разработки при этом не обязательно существенно увеличивается: часть работы, которая обычно выполняется в виде неформальных обсуждений и исправлений после ошибок интеграции, переносится в формализованную проектную модель.

Онтологический подход особенно полезен при интеграции моделей с разным временным шагом. Например, метеорологические данные могут иметь часовой шаг, модель переходных процессов – миллисекундный, расчет годовой выработки – месячный или годовой шаг, а диспетчерская визуализация – шаг обновления в несколько секунд. Онтология не заменяет механизмы интерполяции и агрегации, но позволяет явно описать, какой временной масштаб относится к конкретной модели и какие преобразования допустимы. Это важно для предотвращения ошибок, когда среднее значение используется там, где нужна мгновенная величина, или прогнозное значение смешивается с фактическим измерением.

Еще один эффект связан с повторным использованием знаний. Если онтология содержит описание типовых объектов ветровой электростанции и их параметров, то при разработке нового сценария не требуется заново согласовывать базовые термины. Например, сценарии «оценка КИУМ», «подбор емкости батарей» и «оценка доли дизельной генерации» используют разные расчетные модели, но опираются на общий набор сущностей: ветровая установка, потребитель, аккумуляторная батарея, дизельный генератор, временной ряд ветра, график нагрузки, баланс мощности. Это снижает риск появления несовместимых моделей данных для близких задач.

В литературе, рассматривающей требования к программному обеспечению, онтологии предлагаются как средство поддержки выявления, спецификации, анализа, проверки полноты и непротиворечивости требований [19–21]. Для цифрового двойника это особенно существенно, поскольку техническое задание должно описывать не только пользовательские функции, но и смысл расчетных данных. При отсутствии онтологии требования вида «добавить модель аккумуляторной батареи» или «добавить визуализацию мощности» остаются слишком общими. Онтология позволяет уточнить, какие параметры батареи обязательны, какие состояния допустимы, какие данные должны поступать из датчиков, какие – рассчитываться, какие значения отображаются пользователю и какие проверки выполняются перед расчетом. В таблице 1 представлены этапы использования онтологии при проектировании новой функциональности цифрового двойника.

Таблица 1. Этапы использования онтологии при проектировании новой функциональности цифрового двойника

Этап	Роль онтологии	Результат для разработки
Формулирование задачи	Уточнение объекта, сценария, входов, выходов и ограничений	Техническое задание без неоднозначных терминов
Выбор модели	Сопоставление модели с объектами, параметрами и областью применимости	Понятный состав расчетного модуля и требуемых данных
Проектирование данных	Привязка параметров к источникам, единицам измерения и временным рядам	Меньше дублирования в базе данных и API
Реализация	Согласование классов, таблиц, сервисов и интерфейсов с общей семантикой	Снижение числа интеграционных ошибок
Развитие и сопровождение	Контроль связей между новыми и существующими сущностями	Более управляемое изменение цифрового двойника

5. Практические предпосылки применения онтологий. Онтологический подход не является только академическим инструментом. Крупные цифровые платформы и промышленные проекты уже используют онтологии, семантические модели и графы знаний для интеграции данных, описания активов и построения цифровых решений. В документации Palantir Foundry онтология определяется, как операционный слой организации, соединяющий цифровые активы платформы с реальными объектами, включая оборудование, продукты, процессы и действия [22]. В Azure Digital Twins компания Microsoft предлагает отраслевые DTDL-онтологии, разработанные совместно с предметными экспертами и предназначенные для уменьшения повторной разработки моделей и упрощения решений [23]. Компания DNV предлагает услуги построения энергетических онтологий и указывает, что такие модели позволяют автоматически проверять качество данных и соответствие требованиям [24]. В рамках OntoCommons описан промышленный пример Siemens, где создана библиотека онтологий для доменов производства и автоматизации, а также преобразования промышленных стандартов [25].

Приведенные выше примеры не означают, что для цифрового двойника ветровой электростанции следует напрямую использовать готовую корпоративную онтологию. Их значение состоит в другом: промышленная разработка все чаще приходит к необходимости отдельного семантического слоя, который связывает данные, модели, реальные объекты и пользовательские процессы. Для энергетических объектов это требование имеет принципиальное значение, поскольку ошибки в расчетных моделях могут приводить к существенным техническим и экономическим последствиям. Дополнительную сложность создает большое число взаимосвязанных параметров, а также необходимость согласованной работы специалистов разных профилей – инженеров, математиков, разработчиков программного обеспечения и эксплуатационного персонала.

При проектировании цифрового двойника ветровой электростанции онтология может стать центральным элементом технического задания. В ней фиксируется, какие объекты существуют в системе, какие атрибуты обязательны, какие модели допускается применять, какие данные должны быть сохранены, какие результаты должны отображаться и какие ограничения должны контролироваться. Это позволяет уменьшить разрыв между языком предметного специалиста и языком программной реализации. Программист получает не абстрактное описание «рассчитать режим батареи», а формализованный набор сущностей: батарея, емкость, напряжение, допустимая мощность заряда, допустимая мощность разряда, КПД, температурный диапазон, связь с балансом мощности и условия включения дизельного генератора.

За счет этого онтология не обязательно увеличивает продолжительность разработки. Напротив, при росте цифрового двойника она может сокращать время согласования новых функций, потому что команда работает с уже заданным набором понятий и связей. Дополнительные затраты возникают на этапе первичного построения и сопровождения онтологии. Однако эти затраты компенсируются уменьшением числа неявных решений, повторных обсуждений и исправлений, связанных с ошибками интеграции моделей. Для прототипа с одной моделью такая выгода может быть небольшой, но для цифрового двойника, который должен объединять аэродинамику, механику, электрическую часть, накопители, дизельную генерацию и потребителей, она становится принципиальной.

6. Эффективность цифрового двойника при использовании онтологий для интеграции моделей. Эффективность цифрового двойника ветровой электростанции следует оценивать по двум группам показателей. Первая группа связана с энергетическим и эксплуатационным эффектом: повышение точности оценки выработки, снижение простоев, улучшение планирования обслуживания, оптимизация участия батарей и дизельной

электростанции, уменьшение топливного расхода и повышение наблюдаемости объекта. Вторая группа связана с разработкой: сокращение числа интеграционных ошибок, уменьшение дублирования данных, ускорение добавления новых моделей и новой функциональности и снижение зависимости от уникальных сотрудников.

Практические примеры применения цифровых решений в ветроэнергетике показывают значимость такого направления. GE при запуске концепции Digital Wind Farm указывала, что сочетание сенсоров, сетей данных и аналитики может повысить годовую выработку ветровой электростанции до 20% за счет настройки турбин под конкретные условия эксплуатации [26]. Обобщенные оценки для цифровых двойников в промышленной разработке показывают, что такие технологии способны сокращать время вывода продукта на рынок до 50%, повышать качество продукции и создавать дополнительные источники дохода при правильной организационной и технологической реализации [27].

В нашем случае эффективность предлагаемого подхода целесообразно рассматривать применительно к процессам проектирования, разработки и сопровождения цифрового двойника, а не связывать непосредственно с увеличением выработки электроэнергии или изменением эксплуатационных показателей ветровой электростанции. Энергетический эффект определяется точностью используемых математических моделей, алгоритмами управления и качеством исходных данных. Онтологическое моделирование оказывает влияние на другой уровень системы – на согласованность знаний, данных и программных компонентов, посредством которых эти модели объединяются в цифровом двойнике.

В основу интеграции положена система взаимосвязанных онтологий, включающая онтологию моделируемого объекта, онтологию информационной модели, онтологию математической модели и онтологию сопутствующих объектов. Их концептуальные связи представлены на рисунке 1. Онтология моделируемого объекта содержит базовое описание ветровой электростанции, ее подсистем, оборудования, параметров и отношений между объектами. Она служит основанием для формирования специализированных онтологий информационной и математической моделей. Эти онтологии описывают моделируемый объект и используют определенные в предметной онтологии параметры.

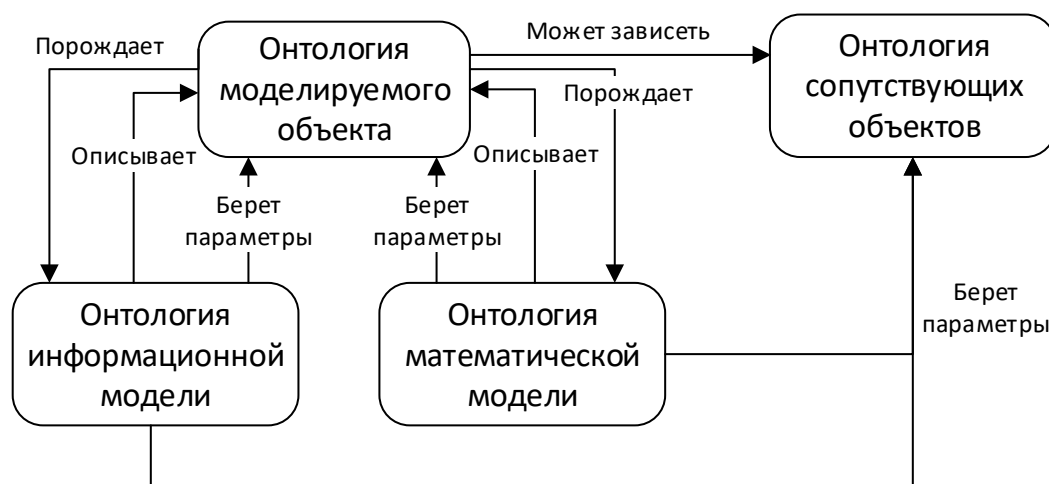


Рис. 1. Концептуальная структура связей между онтологиями цифрового двойника ветровой электростанции

Онтология сопутствующих объектов предназначена для тех фрагментов знаний, которые необходимы для расчета и программной реализации, но не относятся непосредственно к конструктивному составу ветровой электростанции. К ним могут быть отнесены погодные характеристики, требования к программному продукту, элементы архитектуры цифрового двойника, сценарии расчетов, способы визуализации и правила обмена данными. Предметная

онтология может зависеть от этих описаний, а онтологии математической и информационной моделей получают из них необходимые параметры. Например, модель мощности ветровой установки использует скорость ветра и плотность воздуха, которые описываются в онтологии погодных характеристик; одновременно информационная модель должна фиксировать источник этих данных, высоту измерения, временной шаг и единицы измерения.

Такое разделение позволяет не объединять все сведения в одну чрезмерно крупную онтологию. Каждый компонент сохраняет собственное назначение, но остается связанным с другими через явно заданные отношения. Онтология математической модели описывает входные и выходные параметры, расчетные зависимости, ограничения, допущения и область применимости модели. Онтология информационной модели определяет, в каком виде соответствующие параметры представлены в базе данных, программных интерфейсах и цифровой тени. Онтология моделируемого объекта задает предметный смысл этих параметров, а онтология сопутствующих объектов уточняет внешние условия и требования. В результате один параметр может быть прослежен от понятия предметной области до источника данных и конкретного расчетного модуля.

Практический эффект такой структуры проявляется прежде всего в снижении вероятности семантических ошибок. Без формального описания одна и та же величина может быть представлена в разных компонентах под различными именами либо, напротив, одним термином могут обозначаться разные величины. Для скорости ветра необходимо различать скорость на высоте измерения, скорость на высоте ступицы, усредненное и мгновенное значения. Для мощности – установленную, доступную, фактическую и прогнозную мощности. Фиксация этих различий в онтологии уменьшает риск того, что расчетный модуль получит формально допустимое, но содержательно неверное значение.

Другой эффект связан с проектированием новой функциональности. При добавлении математической модели сначала определяется, какой объект она описывает, какие параметры получает из онтологии моделируемого объекта и сопутствующих онтологий, какие результаты формирует и где они должны быть представлены в информационной модели. После этого могут проектироваться структура хранения данных, программный интерфейс и визуализация. Такой порядок дает возможность выявить недостающие или дублирующиеся сущности до реализации программного кода. Онтология в данном случае выступает не дополнительной документацией, создаваемой после разработки, а проектным артефактом, используемым при подготовке требований и согласовании решений между инженерами, математиками и программистами.

Формализованные связи могут использоваться и для частичной автоматизации разработки. На основе онтологии информационной модели возможно формирование структуры таблиц базы данных, классов передачи данных и правил проверки значений. Связи с онтологией математической модели позволяют находить параметры, необходимые для конкретного расчета, проверять наличие источников данных и формировать описание интерфейса расчетного модуля. Представление онтологии в форме графа знаний может применяться при оформлении документации цифрового двойника, поиска зависимостей и анализа последствий изменений. При замене модели или добавлении нового показателя становится возможным определить, какие сущности, расчеты, таблицы и компоненты визуализации связаны с изменяемым элементом.

Эффективность онтологической интеграции может оцениваться по показателям, непосредственно связанным с развитием программного продукта. К ним относятся количество выявленных дублирующихся параметров, доля параметров с явно заданными единицами измерения, источниками и временными характеристиками, число несогласованных связей между моделями и между программными компонентами, продолжительность подключения

нового программного модуля, объем изменений в существующих компонентах и полнота трассировки требований до данных и программной реализации.

Апробация предложенной структуры на примере цифрового двойника ветровой электростанции показала возможность связать описание объекта с параметрами математических моделей и элементами информационной модели. В частности, параметры ветрового потока, характеристики ветровой установки и коэффициенты эффективности могут быть представлены, как связанные сущности, используемые одновременно в расчетах и в структурах хранения данных. Это создает основу для согласованного расширения цифрового двойника при добавлении новых моделей, источников информации и средств визуализации. Следовательно, основным результатом применения онтологического подхода является не устранение необходимости в предметных специалистах, а сохранение и структурирование их знаний в форме, пригодной для совместного использования, проверки и повторного применения.

Интеграция через онтологические модели повышает управляемость развития цифрового двойника за счет явного разделения и согласования предметных, математических, информационных и сопутствующих знаний. Ожидаемый эффект выражается в уменьшении дублирования данных, повышении трассируемости параметров, снижении числа интеграционных ошибок и сокращении объема повторного согласования при изменении программной системы. Количественная оценка этих эффектов требует отдельного вычислительного и программно-инженерного эксперимента, однако представленная структура определяет показатели и условия, необходимые для его проведения.

Заключение. В статье рассмотрена задача интеграции математических моделей, необходимых для построения цифровой модели цифрового двойника ветровой электростанции. Показано, что ветровая электростанция требует совместного учета аэродинамических, механических, электрических, энергетических, аккумуляторных, дизель-генераторных и нагрузочных моделей. Каждая из этих моделей имеет собственные входные данные, выходные величины, ограничения и область применимости, поэтому их объединение в цифровом двойнике является самостоятельной программно-инженерной задачей. При отсутствии явного семантического слоя развитие цифрового двойника приводит к дублированию данных, расхождению терминов, неявным зависимостям между модулями и ускоренному росту сложности программной системы. Предложено использовать для интеграции моделей онтологический подход, при котором знания о составе объекта, параметрах, моделях, задачах и программных компонентах фиксируются в формализованной модели. Такой подход целесообразно применять уже на этапе подготовки технического задания и проектирования новой функциональности, включая внедрение новых расчетных моделей и средств визуализации. Онтология позволяет снизить зависимость проекта от уникальных специалистов, одновременно владеющих математическим моделированием и программной инженерией, поскольку часть их знаний сохраняется в общей модели и становится доступной для разных участников команды. В результате онтологическая интеграция может рассматриваться, как средство повышения воспроизводимости, сопровождаемости и экономической устойчивости разработки цифрового двойника ветровой электростанции.

Благодарности. Работа выполнена в рамках проекта по госзаданию ИСЭМ СО РАН № FWEU-2026-0007. Автор выражает глубокую благодарность д.т.н. Людмиле Васильевне Массель за научное руководство, ценные консультации и поддержку, оказанную при выполнении исследования.

Список источников

1. Burton T., Sharpe D., Jenkins N. et al. Wind Energy Handbook. John Wiley & Sons, 2011, 617 p.
2. Yves-Marie S.-D., Romain B., Malte J. et al. A parametric model for wind turbine power curves incorporating environmental conditions. *Renewable Energy*, 2020, pp. 754-768, DOI: 10.1016/j.renene.2020.04.123.
3. Slootweg J., Haan S.W.H., Polinder H., Kling W.L. General model for representing variable-speed wind turbines in power system dynamics simulations. *IEEE Power Engineering Review*, 2002, vol. 18, pp. 144-151, DOI: 10.1109/MPER.2002.4311816.
4. Chinchilla M., Arnalte S., Burgos J. Control of permanent-magnet generators applied to variable-speed wind-energy systems connected to the grid. *IEEE Transactions on Energy Conversion*, 2006, pp. 130-135, DOI: 10.1109/TEC.2005.853735.
5. Krause P., Wasynczuk, Sudhoff S. et al. *Analysis of Electric Machinery and Drive Systems*, 3rd ed, Hoboken: John Wiley & Sons, 2013, 680 p.
6. Бельский А.А. Оптимизация параметров автономного ветродизельного комплекса, эксплуатируемого на горных территориях / А.А. Бельский, Е.А. Емельянов // Устойчивое развитие горных территорий. – 2026. – Т. 18. – № 1. – с. 69-80. – DOI: 10.21177/1998-4502-2026-18-1-69-80.
7. Djohra S., Koussa M., Rennane A. et al. Hybrid diesel-wind system with battery storage operating in standalone mode: control and energy management – experimental investigation. *Energy*, 2017, DOI: 10.1016/j.energy.2017.04.127.
8. Tremblay O., Dessaint L.-A. Experimental validation of a battery dynamic model for EV application. *World Electric Vehicle Journal*, 2009, vol. 3, pp. 289-298, DOI: 10.3390/wevj3020289.
9. Гуревич Ю.Е. Применение математических моделей электрической нагрузки в расчетах устойчивости энергосистем и надежности электроснабжения промышленных потребителей / Ю.Е. Гуревич, Л.Е. Либова // М.: ЭЛЕКС-КМ, 2008. – 248 с.
10. Arif A., Wang Z., Wang J. et al. Load modeling – a review. *IEEE Transactions on Smart Grid*, 2018, vol. 9, no. 6, pp. 5986–5999, DOI: 10.1109/TSG.2017.2700436.
11. Lehman M.M. Programs, life cycles, and laws of software evolution. *Proceedings of the IEEE*, 1980, vol. 68, no. 9, pp. 1060–1076, DOI: 10.1109/PROC.1980.11805.
12. Levén W., Broman H., Besker T., Torkar R. The broken windows theory applies to technical debt. *Empirical Software Engineering*, 2024, DOI: 10.1007/s10664-024-10456-6.
13. Glaessgen E., Stargel D. The digital twin paradigm for future NASA and U.S. Air Force vehicles. *53rd Structures, Structural Dynamics and Materials Conference*, 2012, DOI: 10.2514/6.2012-1818.
14. Tao F., Cheng J., Qi Q. et al. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 2018, vol. 94, pp. 3563–3576, DOI: 10.1007/s00170-017-0233-1.
15. Rasheed A., San O., Kvamsdal T. Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access*, 2020, vol. 8, DOI: 10.1109/ACCESS.2020.2970143.
16. Semeraro C., Lezoche M., Panetto H., Dassisti M. Digital twin paradigm: a systematic literature review. *Computers in Industry*, 2021, vol. 130, DOI: 10.1016/j.compind.2021.103469.
17. Боровков А.И. Цифровые двойники: определение, подходы и методы разработки / А.И. Боровков, Ю.А. Рябов // Цифровые Двойники. – СПб.: СПбПУ, 2019. – С. 234-245.
18. W3C OWL Working Group. OWL 2 web ontology language document overview (second edition). Available at: <https://www.w3.org/TR/owl2-overview/> (accessed: 06/24/2026).
19. Dermeval D., Vilela J., Bittencourt I.I. et al. Applications of ontologies in requirements engineering: a systematic review of the literature. *Requirements Engineering*, 2015, vol. 20, pp. 1-35, DOI: 10.1007/s00766-015-0222-6.
20. Siegemund K., Assmann U., Pan J., Thomas E., Zhao Y. Towards ontology-driven requirements engineering. *Proceedings of the 10th International Semantic Web Conference*, 2011.
21. Муртазина М. Ш. Система поддержки принятия решений при гибком подходе к инженерии требований на основе OWL-онтологии / М. Ш. Муртазина // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. – 2018. – № 4. – с. 43-55. – DOI: 10.24143/2072-9502-2018-4-43-55.
22. Palantir Technologies. Ontology building overview. Available at: <https://www.palantir.com/docs/foundry/ontology/overview/> (accessed: 06/24/2026).
23. Microsoft. Adopting a DTDL-based industry ontology. *Azure Digital Twins documentation*. Available at: <https://learn.microsoft.com/en-us/azure/digital-twins/concepts-ontologies-adopt> (accessed: 06.24.2026).
24. DNV. Ontology engineering services. Available at: <https://www.dnv.com/services/ontology-engineering-services-199602/> (accessed: 06/24/2026).

25. OntoCommons. Siemens digital manufacturing. Available at: <https://ontocommons.eu/demonstrators/siemens-digital-manufacturing> (accessed: 06/24/2026).
26. GE News. GE launches the next evolution of wind energy making renewables more efficient, economic: the digital wind farm. Available at: <https://www.ge.com/news/press-releases/ge-launches-next-evolution-wind-energy-making-renewables-more-efficient-economic> (accessed: 06/24/2026).
27. McKinsey & Company. PLM systems and the digital twin journey. Available at: <https://www.mckinsey.com/capabilities/operations/our-insights/digital-twins-the-art-of-the-possible-in-product-development-and-beyond> (accessed: 06/24/2026).

Щукин Никита Игоревич. Младший научный сотрудник отдела Трубопроводных систем энергетики, Институт систем энергетики им. Л. А. Мелентьева СО РАН, niksha14@mail.ru, AuthorID: 1122536, SPIN: 6038-0410, ORCID: 0000-0001-8331-335X. Россия, Иркутск, Лермонтова д.130.

UDC 004.94:621.311.245:004.82

DOI:10.25729/ESI.2026.43.3.010

Digital model of a digital twin of a wind power plant development using ontological modeling

Nikita I. Shchukin

Melentiev Energy Systems Institute SB RAS, Russia, Irkutsk, niksha14@mail.ru

Abstract. The paper addresses the problem of integrating heterogeneous mathematical models used in the formation of a digital model of a digital twin of a wind power plant. The relevance of the study is determined by the fact that a wind power plant is not a single computational object, but a complex technical system comprising a number of interrelated subsystems: a wind turbine, an aerodynamic circuit, a mechanical transmission, power electronics, an energy transmission system, a diesel power plant, battery energy storage systems, and consumers. For each of these subsystems, the scientific literature presents various classes of models that differ in terms of input data, level of detail, computational time step, adopted assumptions, constraints, and scope of applicability. In the development of a digital twin, such models should not only be implemented within a single software system, but also coordinated with one another in terms of parameter semantics, units of measurement, temporal characteristics, boundary conditions, and rules for transferring computational results. The aim of the study is to substantiate the feasibility of applying an ontological approach to model integration within the digital model of a digital twin of a wind power plant. It is shown that, in the absence of explicit formalization of the subject area, the increasing complexity of a digital twin is accompanied by data duplication, the emergence of implicit dependencies between computational modules, difficulties in maintaining the software system, and growing costs associated with the development of new functionality. An ontological model is proposed as a means of limiting these effects, defining a common layer of terms, relationships, constraints, and data requirements. It is advisable to apply such a model at the stage of designing new digital twin functionality, including the implementation of new computational models, modification of visualization components, database expansion, and refinement of relationships between subsystems. The novelty of the proposed approach lies in considering ontology not only as a means of representing knowledge about the subject area, but also as a tool for reducing organizational and software engineering risks in the integration of digital twin models. Particular attention is paid to the fact that an ontological model makes it possible to capture specialists' knowledge in a formalized form and thereby reduce the project's dependence on a limited number of employees who simultaneously possess expertise in software engineering, mathematical modeling, and the specific features of energy facilities. The paper substantiates that the use of an ontological approach can contribute to improving the consistency of digital model components, reducing the probability of errors in data exchange between computational modules, and simplifying the further development of the digital twin of a wind power plant.

Keywords: digital twin, wind power plant, mathematical modeling, ontological modeling, model integration

Acknowledgements: The work was carried out within the framework of the project under the state assignment ESI SB RAS No. FWEU-2026-0007. The author expresses deep gratitude to Liudmila Vasilievna Massel for scientific guidance, valuable consultations and support provided during the research.

References

1. Burton T., Sharpe D., Jenkins N., Bossanyi E. Wind Energy Handbook. John Wiley & Sons, 2011, 617 p.
2. Yves-Marie S.-D., Romain B., Malte J. et al. A parametric model for wind turbine power curves incorporating environmental conditions. *Renewable Energy*, 2020, pp. 754-768, DOI: 10.1016/j.renene.2020.04.123.
3. Slootweg J., Haan S.W.H., Polinder H., Kling W.L. General model for representing variable-speed wind turbines in power system dynamics simulations. *IEEE Power Engineering Review*, 2002, vol. 18, pp. 144-151, DOI: 10.1109/MPER.2002.4311816..
4. Chinchilla M., Arnalte S., Burgos J. Control of permanent-magnet generators applied to variable-speed wind-energy systems connected to the grid. *IEEE Transactions on Energy Conversion*, 2006, pp. 130-135, DOI: 10.1109/TEC.2005.853735.
5. Krause P., Wasynczuk, Sudhoff S., Pekarek S. Analysis of Electric Machinery and Drive Systems, 3rd ed. Hoboken: John Wiley & Sons, 2013, 680 p.
6. Belsky A.A., Emelyanov E.A. Optimizatsiya parametrov avtonomnogo vetrodizelnogo kompleksa, ekspluatiruyemogo na gornyykh territoriyakh [Optimization of parameters of an autonomous wind-diesel complex operated in mountainous areas]. *Ustoychivoye razvitiye gornyykh territoriy [Sustainable Development of Mountain Territories]*, 2026, vol. 18, no. 1, pp. 69-80, DOI: 10.21177/1998-4502-2026-18-1-69-80.
7. Djohra S., Koussa M., Rennane A. et al. Hybrid diesel-wind system with battery storage operating in standalone mode: control and energy management – experimental investigation. *Energy*, 2017, DOI: 10.1016/j.energy.2017.04.127.
8. Tremblay O., Dessaint L.-A. Experimental validation of a battery dynamic model for EV application. *World Electric Vehicle Journal*, 2009, vol. 3, pp. 289-298, DOI: 10.3390/wevj3020289.
9. Gurevich Yu.E., Libova L.E. Primeneniye matematicheskikh modeley elektricheskoy nagruzki v raschetakh ustoychivosti energosistem i nadezhnosti elektrosnabzheniya promyshlennykh potrebiteley [Application of mathematical models of electrical load in calculations of power system stability and reliability of power supply to industrial consumers]. Moscow, ELEKS-KM Publ., 2008, 248 p.
10. Arif A., Wang Z., Wang J. et al. Load modeling – a review. *IEEE Transactions on Smart Grid*, 2018, vol. 9, no. 6, pp. 5986–5999, DOI: 10.1109/TSG.2017.2700436.
11. Lehman M.M. Programs, life cycles, and laws of software evolution. *Proceedings of the IEEE*, 1980, vol. 68, no. 9, pp. 1060–1076, DOI: 10.1109/PROC.1980.11805.
12. Levén W., Broman H., Besker T., Torkar R. The broken windows theory applies to technical debt. *Empirical Software Engineering*, 2024, DOI: 10.1007/s10664-024-10456-6.
13. Glaessgen E., Stargel D. The digital twin paradigm for future NASA and U.S. Air Force vehicles. *53rd Structures, Structural Dynamics and Materials Conference*, 2012, DOI: 10.2514/6.2012-1818.
14. Tao F., Cheng J., Qi Q. et al. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 2018, vol. 94, pp. 3563–3576, DOI: 10.1007/s00170-017-0233-1.
15. Rasheed A., San O., Kvamsdal T. Digital twin: values, challenges and enablers from a modeling perspective. *IEEE Access*, 2020, vol. 8, DOI: 10.1109/ACCESS.2020.2970143.
16. Semeraro C., Lezoche M., Panetto H., Dassisti M. Digital twin paradigm: a systematic literature review. *Computers in Industry*, 2021, vol. 130, DOI: 10.1016/j.compind.2021.103469.
17. Borovkov A.I., Ryabov Yu.A. Tsifrovyye dvoyniki: opredeleniye, podkhody i metody razrabotki [Digital twins: definition, approaches and development methods]. *Tsifrovyye dvoyniki [Digital Twins]*. Saint Petersburg, SPbPU Publ., 2019, pp. 234-245.
18. W3C OWL Working Group. OWL 2 web ontology language document overview (second edition). Available at: <https://www.w3.org/TR/owl2-overview/> (accessed: 06/24/2026).
19. Dermeval D., Vilela J., Bittencourt I.I. et al. Applications of ontologies in requirements engineering: a systematic review of the literature. *Requirements Engineering*, 2015, vol. 20, pp. 1-35, DOI: 10.1007/s00766-015-0222-6.
20. Siegemund K., Assmann U., Pan J. et al. Towards ontology-driven requirements engineering. *Proceedings of the 10th International Semantic Web Conference*, 2011.
21. Murtazina M.Sh. Sistema podderzhki prinyatiya resheniy pri gibkom podkhode k inzhenerii trebovaniy na osnove OWL-ontologii [Decision support system for flexible requirements engineering based on OWL ontology]. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravleniye, vychislitel'naya tekhnika i informatika [Bulletin of Astrakhan State Technical University. Series: Management, Computer Science and Informatics]*, 2018, no. 4, pp. 43-55, DOI: 10.24143/2072-9502-2018-4-43-55.
22. Palantir Technologies. Ontology building overview. Available at: <https://www.palantir.com/docs/foundry/ontology/overview/> (accessed: 06/24/2026).

23. Microsoft. Adopting a DTDL-based industry ontology. Azure Digital Twins documentation. Available at: <https://learn.microsoft.com/en-us/azure/digital-twins/concepts-ontologies-adopt> (accessed: 06/24/2026).
24. DNV. Ontology engineering services. Available at: <https://www.dnv.com/services/ontology-engineering-services-199602/> (accessed: 06/24/2026).
25. OntoCommons. Siemens digital manufacturing. Available at: <https://ontocommons.eu/demonstrators/siemens-digital-manufacturing> (accessed: 06/24/2026).
26. GE News. GE launches the next evolution of wind energy making renewables more efficient, economic: the digital wind farm. Available at: <https://www.ge.com/news/press-releases/ge-launches-next-evolution-wind-energy-making-renewables-more-efficient-economic> (accessed: 06/24/2026).
27. McKinsey & Company. PLM systems and the digital twin journey. Available at: <https://www.mckinsey.com/capabilities/operations/our-insights/digital-twins-the-art-of-the-possible-in-product-development-and-beyond> (accessed: 06/24/2026).

Shchukin Nikita Igorevich. Junior Research Fellow of Pipeline Energy Systems, Melentiev Energy Systems Institute SB RAS, niksha14@mail.ru, AuthorID: 1122536, SPIN: 6038-0410, ORCID: 0000-0001-8331-335X. Russia, Irkutsk, Lermontov st., 130.

Статья поступила в редакцию 26.06.2026; одобрена после рецензирования 22.07.2026; принята к публикации 30.07.2026.

The article was submitted 06/26/2026; approved after reviewing 07/22/2026; accepted for publication 07/30/2026.

Цифровая экономика и управление

УДК 001.57:004.5:657

DOI:10.25729/ESI.2026.43.3.011

Построение целевой функции учетной информационной системы публично-правового образования

Ерженин Роман Валерьевич

ООО «Криста-Иркутск», Россия, Иркутск, rerzhenin@gmail.com

Аннотация. Динамичная трансформация человеко-машинного взаимодействия постоянно нуждается в разработке новых моделей принятия решений для поддержания эффективности и качества любого информационного процесса. Публично-правовому образованию, как владельцу процесса исполнения бюджета, в условиях динамичного развития информационных технологий и высокого уровня текучести управленческих кадров остро необходима универсальная методология оценки функционирования информационной системы бухгалтерского учета (ИСБУ). Для решения поставленной задачи с использованием системного подхода были всесторонне проанализированы процессы, связанные с обработкой информацией в централизованных бухгалтериях (ЦБ), и процессы управления подобными учреждениями в условиях бюджетных ограничений. Деятельность пользователей ИСБУ определена рядом показателей, сгруппированных по четырем направлениям: предметные, качественные, бюджетные и экономические. В результате исследования, на основе принципов транспортной задачи была описана и схематически представлена общая математическая модель и целевая функция функционирования ИСБУ, учитывающая основные функции, зависимости и показатели информационного процесса на уровне пользователя. Схематическая модель расчета целевой функции, представленная в формате IDEF0, включает процессы количественного расчета обработанных пользователем данных, расчета непроизводительных потерь рабочего времени и совокупных непроизводительных затрат. Наиболее важными показателями, оказывающими непосредственное влияние на производительность пользователя ИСБУ, определены временные издержки, связанные с поиском собственных ошибок и повторным вводом (корректировкой исходной информации). Выбор целевого критерия ЦБ задает не только задачи оперативного управления, но и стратегию повышения эффективности бюджетных расходов, позволяющую в процессе функционирования ИСБУ направлять бюджетные средства на те цели, которые улучшают ее целевые функции. Целевой критерий публично-правового образования, как учредителя ЦБ, определяет бюджетное ограничение для целевой функции функционирования ИСБУ, обусловленное сокращением любых расходов на поддержку функций, сопровождающих основную деятельность. Предложенная в исследовании модель в дальнейшем может использоваться для анализа и классификации пользователей по признаку профессионализма, разработки алгоритмов и интеллектуальных инструментальных средств поддержки принятия решений на различных уровнях управления.

Ключевые слова: моделирование, оценка эффективности, ИСБУ, AIS, пользовательская аналитика, транспортная задача

Цитирование: Ерженин Р.В. Построение целевой функции учетной информационной системы публично-правового образования / Р.В. Ерженин // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 134-143. – DOI:10.25729/ESI.2026.43.3.011.

Введение. Одним из важных результатов развития организационно-управленческих нововведений в публично-правовых образованиях новой России стала современная централизованная бухгалтерия (ЦБ). Её характерную «инновационность» можно рассмотреть в значительных результатах использования ИТ-технологий – существенном повышении ценности результата работы бухгалтерии, в появлении нового информационного продукта как для государства, так и для общества в целом [1].

Процессы развития информационных технологий не стоят на месте, современная информационная система бухгалтерского учета (ИСБУ) в формате программной системы за последние десятилетия в разы «распухла» у всех разработчиков не только в размере программного кода, но и обросла многочисленными сторонними функциями, некоторые из них не облегчили труд, а, наоборот, нагроулили пользователя дополнительной работой. Непрерывные нормативно-методологические нововведения генерируются с такой частотой,

что ни разработчики программ, ни пользователи не успевают отойти еще от предыдущих порций новизны, как законодатели обрушивают на них новый шквал изменений.

В подобных условиях самым нагруженным и самым незащищенным звеном оказывается рядовой пользователь, загоняемый в условия заранее невыполнимых требований. Динамика текучести кадров в централизованных бухгалтериях за последние три года превышает показатели самой низкоквалифицированной группы грузчиков и уборщиц и достигает в среднем 40-45% в год, при условии, что нормальным показателем для квалифицированных специалистов в области управления подобный показатель варьируется в пределах 5-15%.

Непрерывное обновление кадров является актуальной проблемой управления в любом государственном или муниципальном учреждении, в особенности, если речь идет о таком технологическом центре обработки информации, как централизованная бухгалтерия, требующей от специалиста высокий уровень квалификации, профессионализма и ответственности.

Неподготовленный пользователь – это вызов современности для систем класса ИСБУ, ответом на который должна стать смена приоритетов в управлении развитием системы. На взгляд автора, для решения возникающих проблем уже в ближайшее время потребуются усиление интеллектуальных способностей системы, позволяющих компенсировать недостаток знаний у специалиста за счет адаптации, оказания помощи, обучения и ориентации действий пользователя на траекторию действий, заданную опытным путем его коллег-пользователей. Приобретение ИСБУ новых функциональных свойств, базирующихся на подходах к проектированию интерфейсов, ориентированных на опыт пользователя (*User eXperience, UX*) и принципах искусственного интеллекта, невозможно без разработки методологических основ оценки эффективности функционирования ИСБУ с акцентом на оценку поведения пользователей. В этом ключе к первоочередной задаче следует отнести выбор и обоснование целевой функции функционирования ИСБУ.

1. Системный анализ управления и функционирования учетной информационной системы. Применение системного подхода в бухгалтерском (бюджетном) учете имеет некоторую специфику, связанную с определением учета как системы, как процесса, как области знаний и как профессии [2].

Рассматривая с позиций системного подхода введенное в начале 2000-х годов понятие «Информационная система бухгалтерского учета» (ИСБУ) следует заметить, что отечественные исследователи выделили компоненты (системы) отличные от тех, которые приняты выделять в зарубежной литературе для понятия «Accounting Information Systems», AIS. Следует отметить, что персонал, программное обеспечение и информационная технология являются общими для обоих подходов к описанию состава системы (см. таблицу 1).

Таблица 1. Компоненты (системы) AIS и ИСБУ

Accounting Information Systems (AIS) [3,4]	Информационная система бухгалтерского учета (ИСБУ) [5]
people	Персонал, реализующий информационный процесс бухгалтерского учета
procedures and instructions	Учетная информационная система как предмет и продукт труда
data	
software	Комплекс программно-технических средств для реализации информационной технологии
information technology infrastructure	Информационная технология обработки учетных данных
internal controls	Внутренний контроль

Развернув фокус системного взгляда от внутреннего исследования структуры ИСБУ к позиционированию системы во внешней среде, можно увидеть сложную систему взаимоотношений между компонентами бюджетного информационного процесса (рисунок 1а) и выделить две основные функции систем управления (рисунок 1б):

1. Функции управления финансами публично-правового образования (ППО). Согласно бюджетному законодательству, они возложены на финансовый орган (ФО), поэтому, в рамках полномочий по организации бюджетного процесса, можно выделить отдельный контур управления функционированием и развитием условной информационной системы управления финансами (ИСУФ) и ее отдельной подсистемы – ИС финансового органа (ИСФО).
2. Функция управления ЦБ, как предприятием, в рамках которой выделяется отдельный контур управления функционированием и развитием подсистемы ИСБУ.

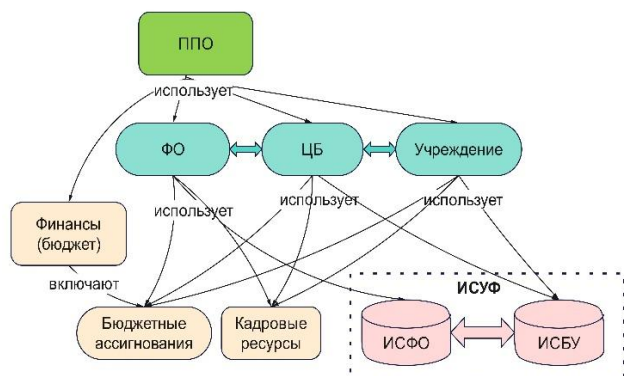


Рис. 1а. Система отношений

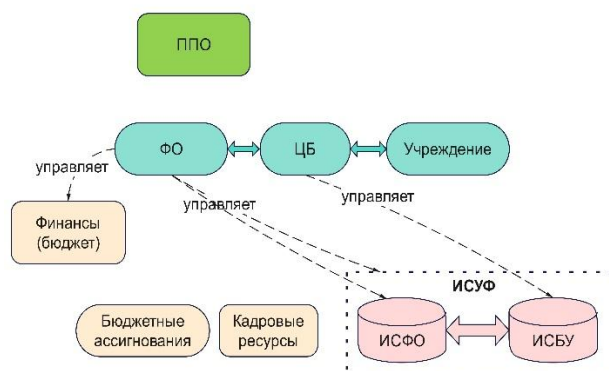


Рис. 1б. Функции систем управления

Таким образом, из фокуса управления ИСБУ нельзя упускать ни факторы, влияющие на управление финансами ППО, ни процедуры и функции, связанные с бюджетным процессом, ни факторы, влияющие на управление функционированием и развитием ИС. Важно учитывать все аспекты отношений, возникающие в процессах функционирования информационной технологии и использования организациями ключевых ресурсов – финансов и кадров.

Факторы, влияющие на функционирование систем класса ИСБУ. Оценке экономической эффективности информационных систем посвящено достаточно много отечественных и зарубежных исследований [6-11]. Преимущественно в предложенных различными авторами методиках сопоставляются результаты использования ИС с затратами на ее внедрение и эксплуатацию, где сопоставимость затрат и результатов предполагает их представление в денежной форме. В данном же случае, когда предметом исследования является не просто деятельность некоммерческого предприятия, т.е. не заинтересованного в получении прибыли, но и в том, что основной целью этого предприятия является обработка различного рода информации о деятельности других организаций с использованием вычислительной техники и предоставление ее заинтересованным пользователям.

Учитывая эти два ключевых фактора, коренным образом отличающие ЦБ от общепринятого понимания предприятия, где ИС управления является основной, а не вспомогательной системой, выделим основные группы функций, оказывающие влияние на управление и функционирование ИСБУ:

«Предметные» – группа функций, представляющих из себя зависимости от возникающего в сфере влияния характеристик предмета деятельности ЦБ на объем, обрабатываемой в ИСБУ информации. Это факторы напрямую связаны с параметрами хозяйственной деятельности обслуживаемых учреждений, информация о которой обрабатывается ЦБ и хранится в регистрах учета ИСБУ.

«Качественные» – представляют собой характеристики функционирования ИСБУ, как организационно-технической системы. Характеристики, касающиеся организации деятельности специалистов-пользователей, функционирования программных систем и вычислительной техники.

«Бюджетные» – группа функций, представляющих из себя зависимости от возникающего в сфере влияния характеристик ЦБ на объем финансирования, выделяемого на его функционирование. Как правило, бюджетные ассигнования (расходы на содержание ЦБ, как учреждения), утверждаются Законом о бюджете и оказывают значительное влияние на ограничения функционирования ИСБУ.

«Экономические» – представляют различные издержки, образующиеся в процессе использования централизованной бухгалтерией информационной технологии. Требуют наибольшего внимания в процессе решения оптимизационных задач.

Таким образом, базовые функции и процессы, влияющие на управление и функционирование ИСБУ, можно описать с использованием принципов модели транспортной системы [12,13].

$$F_{AIS} = (\vec{V}, \vec{P}, \vec{B}, \vec{Q}) \quad (1)$$

где V – вектор параметров, характеризующих объем информации, требующей обработки (предметные факторы), P – вектор параметров, описывающих характеристики информационного процесса (качественные факторы), B – вектор параметров, характеризующих размер ограничений бюджетного финансирования (бюджетные факторы); Q – вектор параметров, характеризующих издержки процесса обработки информации (экономические факторы).

При этом связи в системе управления ИС можно представить в виде следующей системы:

$$F_{AIS} = \begin{Bmatrix} V(P, t) \\ P(Q, B) \\ B(V, t) \\ Q(P, B) \end{Bmatrix} \quad (2)$$

Такая система является базовым представлением процесса обработки информации в общем виде, однако рассмотрение на таком уровне не позволяет выбрать стратегию повышения эффективности управления ИСБУ. Для анализа структуры системы и решения задач управления необходимо представить процесс управления более детально, опираясь на наиболее влияющие на эффективность ИСБУ факторы.

Характеристика ИСБУ наиболее явно связана с информационной технологией, то есть с организацией человеко-машинного взаимодействия, с программной архитектурой и информационной инфраструктурой. В свою очередь, эти факторы напрямую зависят от бюджетного финансирования и экономических параметров, тесно связанных с расходом рабочего времени специалистов, а также с другими издержками, определяющими расходы на владение ИС. Поскольку рассматриваемые параметры модели не имеют прямых линейных зависимостей между собой, то структуру процесса управления целесообразно описывать через более детальные характеристики деятельности пользователей, сгруппированные по функциям, предложенным в формуле 1:

«Предметная» V -группа функций может характеризоваться такими, например, подсистемами, как:

- Общий объем и вес данных основной деятельности. Объем и вес данных, введенных пользователем первично.
- Общий объем и вес данных дополнительной деятельности. Объем и вес данных, измененных пользователем после первичного ввода.
- Количество ошибок и предупреждений системы.

- Количество обслуживаемых в ЦБ учреждений.
«Качественная» Т-группа функций может содержать:
- Функциональность прикладного программного обеспечения.
- Количество рабочих мест ИСБУ.
- Мощность серверов.
- Производительность системы.
- Количество сбоев системы.
- Количество ошибок в контроле бюджетной отчетности.
«Бюджетная» В-группа функций включает:
- Объем бюджетного финансирования.
«Экономическая» Q-группа функций может содержать:
- Различные производительные затраты ЦБ. Заработную плату, налоги, расходы на владение ИС.
- Непроизводительные затраты. Потери рабочего времени. Сумма штрафов и взысканий за нарушения и ошибки и другие случайные факторы.

Очевидно, что сложность и стохастичность системы управления ИСБУ с указанными выше разнородными параметрами существенно возрастет, поэтому для упрощения и дальнейшего анализа выделим только те характеристики деятельности пользователей ИСБУ, которые отвечают критериям измеримости, валидности и существенности (таблица 2).

Таблица 2. Характеристики деятельности пользователей ИСБУ

	Параметр	Расшифровка	Размерность
1	$V_{\text{нов}}$	Количество введенных пользователем новых данных в t	ед.
2	$V_{\text{изм}}$	Количество измененных пользователем данных в t	ед.
3	$V_{\text{удл}}$	Количество удаленных пользователем данных в t	ед.
4	$V_{\text{ошб}}$	Количество ошибок и предупреждений системы в t	ед.
5	M	Количество обслуживаемых в ЦБ учреждений	ед.
6	N	Количество пользователей ИСБУ	ед.
7	$R_{\text{ис}}$	Простой ИС по причине сбоя системы	час
8	$B_{\text{фин}}$	Годовой объем финансирования ЦБ	руб.
9	$Q_{\text{прз}}$	Производительные затраты в t	руб.
10	$Q_{\text{нпрз}}$	Непроизводительные затраты в t	руб.
11	$Q_{\text{нпвр}}$	Непроизводительные потери рабочего времени в t	час
12	$Q_{\text{штрф}}$	Штрафы и удержания за нарушения законодательства в t	руб.
13	$Q_{\text{фин}}$	Финансовые потери обслуживаемых в ЦБ учреждений в связи с некачественной обработкой данных.	руб.
14	$Q_{\text{ис}}$	Финансовые потери ЦБ от простоя ИС	руб.

4. Построение целевой функции. Важнейшим фактором построения целевой функции является отражение противоречия, связанного, с бюджетными и временными ограничениями, при которых повышение непроизводительных затрат приводит к снижению качественных функций ИСБУ. Чем меньше производительного времени, тем меньше возможности обработать первичные данные и сформировать качественную отчетную информацию существующей технологией в полном объеме с заданным параметрами.

Таким образом, задачей ЦБ, обслуживающей учреждения за счет средств бюджета, является качественная обработка информации с наименьшими непроизводительными потерями, а требования по снижению расходов бюджета на исполнение функции учета

отходят на второй план, т.к. потери бюджета от неисполнения функции учета учреждениями намного превышают расходы на ее исполнение.

Со стороны публично-правового образования, за счет средств которого исполняется функция учета в подведомственных учреждениях, первостепенно важным является не размер суммы расходов на содержание ЦБ, а возможность за эти деньги как можно более технологично (эффективно) обработать информацию в соответствии с заявленными требованиями. Задача ЦБ – сократить свои издержки, задача управления ИСБУ – обеспечить максимально технологичную обработку информации.

Вся обработанная пользователем информация в ИСБУ представляет собой совокупность введенных в систему новых, измененных и удаленных данных:

$$V_{\text{общ}} = \sum_{i=1}^N V_{\text{нов}} + \sum_{i=1}^N V_{\text{изм}} + \sum_{i=1}^N V_{\text{удл}} \quad (3)$$

где N – количество пользователей, $V_{\text{общ}}$ – общее количество обработанных i -м пользователем данных, $V_{\text{нов}}$ – количество введенных i -м пользователем новых данных, $V_{\text{изм}}$ – количество измененных i -м пользователем данных, $V_{\text{удл}}$ – количество удаленных i -м пользователем данных.

С точки зрения эффективности управления функционированием ИСБУ задача предельно проста – количество измененных и удаленных данных должно стремиться к минимуму. Это условие можно назвать одним из самых главных требований к организации эффективной обработки учетной информации:

$$\sum_{i=1}^N V_{\text{изм}} + \sum_{i=1}^N V_{\text{удл}} \rightarrow 0 \quad (4)$$

С точки зрения ЦБ, как некоммерческого предприятия, целевым критерием является эффективность бюджетных расходов, то есть сокращение любых непроизводительных затрат.

Таким образом целевая функция ЦБ имеет вид:

$$B_{\text{фин}} - Q_{\text{нпрз}} - \sum_{j=1}^M Q_{\text{фин}} \rightarrow \max \quad (5)$$

где $B_{\text{фин}}$ – годовой объем финансирования ЦБ, $Q_{\text{нпрз}}$ – совокупные годовые непроизводительные затраты, M – количество обслуживаемых в ЦБ учреждений, $Q_{\text{фин}}$ – финансовые потери j -го учреждения.

Подобная стратегия позволяет направлять бюджетные средства на те цели, которые улучшают целевые функции ИСБУ:

- на стимулирование персонала, на создание более выгодных условий для привлечения более квалифицированных специалистов;
- на инвестирование и развитие информационной технологии.

Совокупные годовые непроизводительные затраты $Q_{\text{нпрз}}$ можно представить, как:

$$Q_{\text{нпрз}} = \sum_{i=1}^N Q_{\text{нпр}} \times q + Q_{\text{штрф}} + Q_{\text{ис}} + Q_{\text{изд}} \quad (6)$$

$$Q_{\text{ис}} = \sum_{i=1}^N R_{\text{ис}} \times q + \sum_{i=1}^N R_{\text{ис}} \quad (7)$$

Где N – количество пользователей ИСБУ, q – средняя стоимость часа работы i -го пользователя, $R_{\text{ис}}$ – время простоя ИСБУ по причине сбоя системы, $Q_{\text{нпр}}$ – непроизводительные потери рабочего времени, $Q_{\text{штрф}}$ – общая сумма штрафов и взысканий, наложенных фискальными органами на ЦБ за нарушения, $Q_{\text{изд}}$ – прочие издержки, $Q_{\text{ис}}$ – финансовые потери ЦБ от простоя ИС.

Модель расчета целевой функции в формате IDEF0 представлена на рисунке 2.

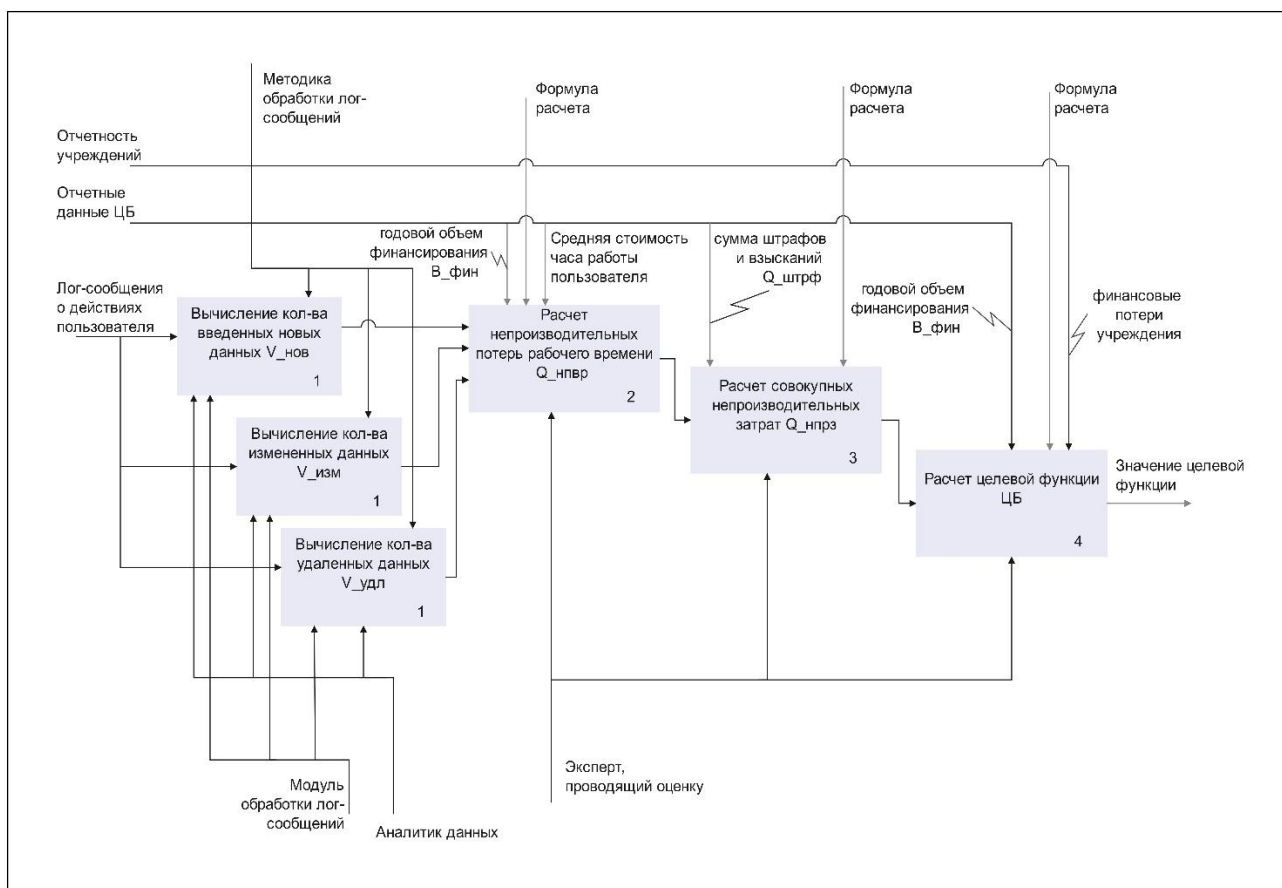


Рис. 2. Модель расчета целевой функции

С точки зрения публично-правового образования, как учредителя ЦБ, целевым критерием является та же эффективность бюджетных расходов, но с другими целями – сокращение любых расходов на сопровождающие функции, к которым и относится учетная функция, исполняемая ЦБ:

$$B_{\text{фин}} + \sum_{j=1}^M Q_{\text{фин}} \rightarrow \min \quad (7)$$

где $B_{\text{фин}}$ – годовой объем финансирования ЦБ, M – количество обслуживаемых в ЦБ учреждений, $Q_{\text{фин}}$ – финансовые потери обслуживаемых ЦБ учреждений.

Заключение. Предложенная в данном исследовании методика определения целевой функции функционирования ИС управления особой – некоммерческой организацией, выполняющей аутсорсинговые функции по обработке учетных данных и предоставлению информации заинтересованным потребителям, направлена на решение проблемы повышения производительности специалиста-пользователя ИС.

Ключевой особенностью подходов к рассмотрению этой проблемы является ориентация оценки функционирования ИСБУ одновременно с двух ракурсов: с позиции архитектуры предприятия, т.е. через рассмотрение процесса управления учреждением принадлежащей ей информационной системой, и с позиции процесса управления бюджетными ресурсами, выделяемыми учредителем для функционирования этого учреждения. Подобный комплексный подход, с одной стороны, позволяет учитывать все аспекты отношений, возникающих при анализе и оценке использования информационной технологии в учетной функции (ведении бухгалтерского учета) публично-правового образования, а с другой – позволяет использовать эти выявленные отношения для определения параметров и ограничений целевой функции.

Применение специального метода решения классической транспортной задачи для описания целевой функции позволило задать критерий эффективности управления

функционированием ИСБУ – достигнуть минимума непроизводительных затрат при заданных ресурсных ограничениях. Таким образом, в результате предложенной нестандартной постановки задачи ядром рассматриваемой проблемы и целью предложенных управленческих решений стал неподготовленный (неквалифицированный) специалист, т.к. именно этот кластер пользователей ИСБУ генерирует избыточное количество непроизводительных действий, тратит свое рабочее время на совершение ошибок, длительный их поиск и исправление первичных документов, и порой неоднократно.

Однако выработка точных и выверенных управленческих решений в этом направлении невозможна без изучения действий пользователей ИСБУ. Подобный анализ поведения и их оценка, в свою очередь, невозможны без интеллектуальных средств помощи принятия решений, предполагающих использование технологии нейросетевой обработки поведенческих данных с целью отделения непроизводительной деятельности пользователей ИСБУ от производительной. Таким образом, в целом предложенная в данном исследовании модель расчета целевой функции не только продекларировала заданную цель управления ИСБУ, но и сформировала представление об общем плане реализации комплекса мероприятий, который еще предстоит реализовать на практике и доказать состоятельность предложенного подхода к повышению эффективности функционирования ИСБУ.

Список источников

1. Ерженин Р.В. Централизованная бухгалтерия в секторе государственного управления (теория и практика) / Р.В. Ерженин. – Москва: Издательский Центр РИОР, 2017. – 204 с.
2. Нарibaев К.Н. Организация и методология бухгалтерского учета в условиях АСУ: Монография / К.Н. Нарibaев. – М.: Финансы и статистика, 1983. 135 с.
3. Hall J.A. Introduction to accounting information system. 8th International ed. South-Western, Cengage Learning, 2013, 800 p.
4. Granof M.H., Khumawala S.B. Government and not-for-profit accounting: concepts and practices. 6th ed. Wiley, 2013, 814 p.
5. Харитонов С.А. Информационные системы бухгалтерского учета: учебное пособие / С.А. Харитонов, Д.В. Чистов, Е.Л. Шуремов. – М.: ФОРУМ; ИНФРА-М, 2010. – 160 с.
6. Andresen J.L. A framework for selecting an IT evaluation method: in the context of construction. Danmarks Tekniske Universitet, 2001, 257 p.
7. Деверадж С. Тайны ИТ: Измерение отдачи от инвестиций в информационные технологии / С. Деверадж, Р. Кохли – М.: Букпресс, 2006. – 192 с.
8. Ерженин Р.В. Теоретические основы оценки эффективности сложной информационной системы / Р.В. Ерженин // Вестник Астраханского государственного технического университета. Серия: Управление, вычислительная техника и информатика. 2024. – № 2. – С. 105-110. – DOI: 10.24143/2072-9502-2024-2-105-110.
9. Cronk M., Fitzgerald E. A conceptual framework for furthering understanding of 'IT business value' and its dimensions. PACIS 1997 Proceedings, 1997, pp. 405-415.
10. Remenyi D., Money A., Sherwood-Smith M. The effective measurement and management of IT costs and benefits. Oxford, Butterworth-Heinemann, 2000, 362 p.
11. Скрипкин К.Г. Экономическая эффективность информационных систем / К.Г. Скрипкин. – М.: ДМК-Пресс, 2002, 256 с.
12. Майоров Н.Н. Практические задачи моделирования транспортных систем / Н.Н. Майоров, В.А. Фетисов. – СПб.: ГУАП, 2012. – 185 с.
13. Kanto Kantorovich L. On the translocation of masses. C. R. (Doklady) Acad. Sci. URSS (N. S.), 1942, vol. 37, pp. 199-201.

Ерженин Роман Валерьевич. Кандидат экономических наук, генеральный директор ООО «Кристалл-Иркутск», исследователь-практик проблем организации человеко-машинного взаимодействия в сфере управления государственными финансами. AuthorID: 752554, SPIN: 6438-1418, ORCID: 0000-0001-9380-0987, Scopus Author ID: 57204457721, erzhenin@gmail.com. 664047, Россия, Иркутск, Трудовая д.60, офис 010.

Construction of the target function of the accounting information system of a public-law entity

Roman V. Erzhenin

Krista-Irkutsk LLC, Russia, Irkutsk, *rerzhenin@gmail.com*

Abstract. The dynamic transformation of human-machine interaction constantly requires the development of new decision-making models to maintain the efficiency and quality of any information process. Public institutions, as the owners of the budget execution process, urgently need a universal methodology for assessing the performance of their accounting information systems (AIIS) in the face of dynamic IT development and high management turnover. To address this challenge, a systems approach was used to comprehensively analyze the processes associated with information processing in centralized accounting departments (CADs) and the management processes of such institutions under budget constraints. The activities of AIIS users are determined by a number of indicators grouped into four areas: subject, qualitative, budgetary, and economic. As a result of the study, a general mathematical model and target function for the functioning of the AIIS, based on the principles of the transport problem, were described and schematically presented. This model takes into account the key functions, dependencies, and indicators of the information process at the user level. A schematic model for calculating the objective function, presented in IDEF0 format, includes processes for quantifying user-processed data, calculating unproductive time losses, and total unproductive costs. The most important indicators directly impacting the productivity of the ISBU user are the time costs associated with identifying their own errors and re-entering (correcting the original information). The selection of the target criterion of the central bank determines not only operational management objectives but also a strategy for improving the efficiency of budget expenditures, allowing budget funds to be directed toward those purposes that improve the objective functions of the ISBU during its operation. The target criterion of the public entity, as the founder of the central bank, determines the budget constraint for the objective function of the ISBU, conditioned by the reduction of any expenses on supporting functions accompanying the core business. The model proposed in the study can be further used to analyze and classify users based on their professionalism, and to develop algorithms and intelligent decision-support tools at various management levels.

Keywords: modeling, performance evaluation, ISBU, AIS, user analytics, transport task

References

1. Erzhenin R.V. Tsentralizovannaya bukhgalteriya v sektore gosudarstvennogo upravleniya (teoriya i praktika) [Centralized accounting in the public sector: theory and practice]. Moscow, RIOR Publ., 2017, 204 p.
2. Naribaev K.N. Organizatsiya i metodologiya bukhgalterskogo ucheta v usloviyakh ASU: monografiya [Organization and methodology of accounting under automated control systems: monograph]. Moscow, Finansy i statistika [Finance and Statistics] Publ., 1983, 135 p.
3. Hall J.A. Introduction to accounting information system. 8th International ed. South-Western, Cengage Learning, 2013, 800 p.
4. Granof M.H., Khumawala S.B. Government and not-for-profit accounting: concepts and practices. 6th ed. Wiley, 2013, 814 p.
5. Kharitonov S.A., Chistov D.V., Shuremov E.L. Informatsionnyye sistemy bukhgalterskogo ucheta: uchebnoye posobiye [Accounting information systems: textbook]. Moscow, FORUM; INFRA-M Publ., 2010, 160 p.
6. Andresen J.L. A framework for selecting an IT evaluation method: in the context of construction. Danmarks Tekniske Universitet, 2001, 257 p.
7. Devaraj S., Kohli R. Tainy IT: Izmereniye otdachi ot investitsiy v informatsionnyye tekhnologii [The IT payoff: measuring the business value of information technology investments]. Moscow, Bukpress Publ., 2006, 192 p.
8. Erzhenin R.V. Teoreticheskiye osnovy otsenki effektivnosti slozhnoy informatsionnoy sistemy [Theoretical foundations for evaluating the effectiveness of a complex information system]. Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta. Seriya: Upravleniye, vychislitel'naya tekhnika i informatika [Bulletin of Astrakhan State Technical University. Series: Management, Computer Science and Informatics], 2024, no. 2, pp. 105-110, DOI: 10.24143/2072-9502-2024-2-105-110.
9. Cronk M., Fitzgerald E. A conceptual framework for furthering understanding of 'IT business value' and its dimensions. PACIS 1997 Proceedings, 1997, pp. 405-415.
10. Remenyi D., Money A., Sherwood-Smith M. The effective measurement and management of IT costs and benefits. Oxford, Butterworth-Heinemann, 2000, 362 p.

11. Skripkin K.G. Ekonomicheskaya effektivnost' informatsionnykh sistem [Economic efficiency of information systems]. Moscow, DMK-Press Publ., 2002, 256 p.
12. Mayorov N.N., Fetisov V.A. Prakticheskiye zadachi modelirovaniya transportnykh sistem [Practical problems of modeling transport systems]. Saint Petersburg, GUAP Publ., 2012, 185 p.
13. Kantorovich L. On the translocation of masses. C. R. (Doklady) Acad. Sci. URSS (N. S.), 1942, vol. 37, pp. 199-201.

Erzhenin Roman Valerievich. *Candidate of Economic Sciences, General Director of Krista-Irkutsk LLC, researcher and practitioner in the field of organizing human-machine interaction in public finance management. AuthorID: 752554, SPIN: 6438-1418, ORCID: 0000-0001-9380-0987, Scopus Author ID: 57204457721, rerzhenin@gmail.com. 664047, Russia, Irkutsk, Trudovaya 60, office 010.*

Статья поступила в редакцию 14.12.2025; одобрена после рецензирования 18.01.2026; принята к публикации 01.08.2026.

The article was submitted 12/14/2025; approved after reviewing 01/18/2026; accepted for publication 08/01/2026.

Интеллектуальный анализ спроса и оптимизация распределения оборотного капитала селлера на электронных маркетплейсах: методологическая основа, алгоритм и дизайн экспериментов

Бастанов Артур Эмильевич

Уфимский университет науки и технологий, Россия, Уфа, gradinv@gmail.com

Аннотация. Статья представляет методологический результат первого этапа исследовательской программы, направленной на повышение эффективности управления товарными запасами и оборотным капиталом селлеров на электронных маркетплейсах. Актуальность исследования определяется тем, что задачи распределения ограниченного оборотного капитала между многочисленными товарными позициями на практике решаются в условиях стохастического спроса, выраженной сезонности, неоднородной волатильности и ограниченных вычислительных ресурсов, в то время как традиционные методы управления запасами либо остаются статичными, либо требуют априорного задания параметров спроса и затрат. В работе предложена комплексная методика анализа спроса и распределения оборотного капитала, включающая три взаимосвязанных компонента: интеллектуальный анализ спроса на основе сезонно-трендовой декомпозиции и моделей машинного обучения для временных рядов (DLinear, NLinear, XGBoost); расширенный анализ волатильности спроса с использованием двух вариантов коэффициента вариации и адаптивной кластеризации ассортиментной матрицы; алгоритм распределения оборотного капитала, основанный на интегральном показателе «индекс доходности» и дополненный учётом сезонности и риска. Существенным элементом методики является разделение предсказуемой сезонной компоненты и остаточной волатильности, что позволяет более корректно оценивать риск товарных позиций и использовать эти оценки при формировании решений о распределении оборотного капитала. Методика формирует структурированный вектор признаков, включающий прогнозные, статистические и экономические характеристики товарных позиций, пригодный как для одноэтапных систем поддержки принятия решений, так и в качестве входного пространства состояний для дальнейших стохастических и динамических моделей управления запасами. Научная новизна работы заключается в интеграции известных инструментов прогнозирования, оценки волатильности и распределения оборотного капитала в единую математически согласованную методологическую модель, учитывающую специфику электронных маркетплейсов и специфические ограничения селлеров. Статья обосновывает предложенную методику, описывает её архитектуру и раскрывает дизайн вычислительных экспериментов, предназначенных для последующей эмпирической оценки её эффективности.

Ключевые слова: маркетплейс, селлер¹, прогнозирование спроса, сезонность, коэффициент вариации, индекс доходности, управление запасами, управление оборотным капиталом, методы машинного обучения

Цитирование: Бастанов А.Э. Интеллектуальный анализ спроса и оптимизация распределения оборотного капитала селлера на электронных маркетплейсах: методологическая основа, алгоритм и дизайн экспериментов / А.Э. Бастанов // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 144-153. – DOI:10.25729/ESI.2026.43.3.012.

Введение. Современный этап развития розничной торговли характеризуется стремительной цифровизацией и ростом доли онлайн-каналов, прежде всего электронных маркетплейсов. Это привело к демократизации доступа на рынок для множества продавцов и одновременно к резкому обострению конкуренции. В таких условиях критически важным для выживания и развития бизнеса селлера становится эффективное управление внутренними ресурсами, прежде всего оборотным капиталом, направляемым на формирование товарных запасов.

Актуальность исследования обусловлена потребностью селлеров в формализованных, но вычислительно недорогих алгоритмах, которые позволяют не только реагировать на

¹ Селлер (от англ. *seller* – продавец) – это предприниматель или компания, которые продают товары через цифровые каналы, чаще всего на маркетплейсах (таких, как Wildberries, Ozon, Яндекс Маркет, Мегамаркет)

рыночные колебания, но и тактически распределять ограниченный оборотный капитал с учётом прогнозируемого спроса, рентабельности и оборачиваемости в условиях риска. Традиционные методы управления запасами и ассортиментом: ABC/XYZ-анализ, матрицы BCG и McKinsey/GE, классические EOQ-модели – решают прежде всего задачи стратегической классификации номенклатуры или опираются на детерминированные зависимости спроса и затрат, слабо учитывая неопределённость и требуя сложных оптимизационных вычислений.

Строгие стохастические модели управления запасами, напротив, учитывают вероятностный характер спроса и позволяют описывать оптимальные политики пополнения (включая (s, S)-политики, политики уровня пополнения и однопериодную модель управления запасами – модель газетчика), но предполагают наличие априорной информации о распределении спроса и параметрах затрат либо отдельного статистического этапа их оценки. На практике продавцы маркетплейсов работают с большим числом SKU (Stock Keeping Unit – уникальный внутренний код товара), неполной и зашумлённой информацией о спросе и ограниченными вычислительными ресурсами, поэтому ни чисто детерминированные, ни классические стохастические постановки в явном виде не закрывают их тактические потребности. В результате формируется методологический разрыв между теоретически строгими моделями управления запасами и прикладными задачами распределения оборотного капитала на маркетплейсах.

В ранее выполненной работе был предложен алгоритм распределения оборотного капитала продавца на основе интегрального показателя «индекс доходности», рассчитываемого, как произведение коэффициента оборачиваемости и маржинальной рентабельности [1]. Алгоритм реализует итерационную процедуру ранжирования номенклатуры по индексу доходности и формирования бюджета закупок в пределах минимальных и целевых уровней запасов. При этом алгоритм остаётся одношаговым и статичным: опирается на точечные оценки оборачиваемости, рентабельности и прогноза продаж и не учитывает динамику спроса, сезонность и риски, связанные с волатильностью.

Цель настоящей работы – формализовать и обосновать методологические аспекты, расширяющие указанный подход за счёт интеграции методов анализа временных рядов и стохастических представлений о спросе.

Обзор литературы и постановка задачи. Распределение оборотного капитала продавца на маркетплейсах, по существу, представляет собой задачу управления запасами в условиях стохастического спроса, осложнённую бюджетными ограничениями. Это определяет необходимость рассмотреть литературные источники в трёх взаимосвязанных направлениях: классические подходы к управлению запасами и практики распределения капитала; методы прогнозирования спроса, обеспечивающие количественные оценки параметров моделей; стохастические модели и модели управления запасами, основанные на данных, задающие теоретическую и вычислительную рамки для разработки искомой методики.

Классические инструменты управления запасами и распределения капитала – это ABC/XYZ-анализ [2], матрицы BCG и McKinsey/GE [3], а также широко известные EOQ-модели и методы математической оптимизации [4, 5]. Эти инструменты решают прежде всего задачи стратегической классификации номенклатуры или оптимизации отдельных товарных позиций при заданных параметрах спроса и затрат. На практике, при распределении оборотного капитала между большим количеством SKU (товарных позиций), продавцы чаще используют простые алгоритмы: приоритизацию по маржинальности [6], по оборачиваемости [7], ABC-классификацию по прибыли [2]). Предложенный ранее алгоритм [1] на основе интегрального индекса доходности объединяет два основных критерия эффективности

(маржинальность и оборачиваемость), что повышает экономическую обоснованность решений. Однако и он остаётся статичным: все решения принимаются на основе точечных, ретроспективных оценок без явного использования прогнозов и без количественного учёта волатильности спроса.

Для того, чтобы перейти от простых статических алгоритмов к более строгим решениям, необходимо иметь модель прогнозирования будущего спроса и оценки степени его неопределённости. Классические модели временных рядов (ARIMA, экспоненциальное сглаживание) обеспечивают приемлемую точность на умеренно волатильных рядах [5], но хуже работают в условиях сильной сезонности, промо-активности и структурных сдвигов. Коэффициент вариации, традиционно применяемый в XYZ-анализе, при наличии сезонности смешивает регулярную сезонную вариацию и остаточный шум, завышая оценку риска для сезонных товаров [8]. Современные исследования показывают, что линейные архитектуры DLinear/NLinear для долгосрочного прогнозирования временных рядов могут конкурировать с трансформерами по точности при меньшей вычислительной сложности [9, 10, 11], а XGBoost остаётся эффективным методом табличного прогнозирования спроса [12, 13, 14]. Сочетание сезонно-трендовой декомпозиции (Seasonal-Trend decomposition using Loess, STL) (для выделения тренда и сезонности), линейных моделей DLinear/NLinear и XGBoost с расширенным анализом коэффициента вариации позволяет разделять предсказуемый сезонный компонент и остаточную волатильность [8]. Тем самым формируется именно тот набор параметров (прогнозы, интервалы, меры волатильности), который потенциально может быть использован в алгоритмах распределения капитала и стохастических моделях запасов.

Классическая теория стохастического управления запасами описывает оптимальные политики при вероятностном спросе и заданной структуре затрат. В дискретном времени при стационарном и независимом по периодам спросе, выпуклых и возрастающих затратах хранения и дефицита и наличии фиксированных затрат заказа доказана оптимальность пороговых (s, S) -политик: существует пара уровней $s < S$, при которой оптимальная политика состоит в заказе до уровня S , если запас ниже s , и отказе от заказа в противном случае. При отсутствии фиксированных затрат естественными являются политики уровня пополнения (base-stock-политики), предусматривающие доведение запаса до постоянного целевого уровня после каждого пополнения. Для однопериодных задач важную роль играет однопериодная модель управления запасами («модель газетчика», newsvendor model), формализующая выбор объёма заказа при асимметричных издержках дефицита и излишка. Современные работы развивают робастные и дистрибуционно-робастные модификации этих постановок, описывая неопределённость спроса через множества допустимых распределений и минимизацию наихудшего ожидаемого риска [15].

Эти модели задают теоретически строгие эталоны управления запасами, но исходно предполагают наличие априорно известного распределения спроса, либо возможность его надёжной оценки на отдельном статистическом этапе. Для селлеров маркетплейсов с большим числом SKU, короткой историей продаж и зашумленными данными это допущение редко выполняется в полной мере, что ограничивает прямое практическое применение классических стохастических методов.

На стыке рассмотренных классических моделей и машинного обучения появились подходы, основанные на данных (data-driven), трёх типов. Во-первых, схемы «прогноз и оптимизация», в которых выходы моделей прогнозирования спроса (статистических или ML-моделей) используются для параметризации классических или робастных задач управления запасами [16, 17]. Во-вторых, методы обучения политики, включая глубокое обучение с подкреплением (Deep Reinforcement Learning, DRL), где стратегия управления

аппроксимируется параметрами обучаемой модели [18, 19, 20]. В-третьих, эвристические алгоритмы, основанные на заранее сконструированных признаках спроса и риска (например, различные варианты ABC/XYZ-классификации по выручке, оборачиваемости и коэффициенту вариации), определяющих правило ранжирования и распределения ресурсов. Сложные DRL-подходы демонстрируют высокий потенциал, но требуют значительных вычислительных ресурсов и больших исторических выборок, что ограничивает их применение в малом и среднем бизнесе, тогда как простые алгоритмы, будучи практически удобными, редко опираются на систематическое использование информации о сезонности и волатильности и слабо связаны с теорией стохастического управления запасами.

Таким образом, анализ трёх групп подходов: классических инструментов управления запасами и распределения капитала, методов прогнозирования спроса и стохастических/основанных на данных моделей управления запасами – выявляет общий методологический разрыв. Теоретически хорошо изучены оптимальные политики при заданном распределении спроса и параметрах затрат, а в практической деятельности продавцы опираются на простые алгоритмы, основанные на ретроспективных показателях и почти не использующие прогнозную и статистическую информацию. При этом отсутствует компактная и формально определённая методика, которая бы одновременно: строила прогнозы спроса и оценки волатильности для множества SKU с учётом сезонности; переводила эти оценки в количественные характеристики риска; интегрировала показатели ожидаемой доходности и риска в алгоритм распределения оборотного капитала на основе экономически интерпретируемого показателя; формировала структурированный вектор признаков, пригодный для последующего использования в более строгих стохастических и динамических моделях.

Решение именно этой задачи, а именно, построение математически согласованной методики, связывающей анализ спроса, оценку неопределённости и алгоритм распределения оборотного капитала для продавцов маркетплейсов, и составляет предмет настоящей статьи.

Методика анализа спроса и распределения оборотного капитала. Исходными данными являются временные ряды продаж по товарным позициям (SKU) продавцов маркетплейсов с горизонтами наблюдений не менее 24 месяцев и шагом дискретизации месяц или неделя. Для каждой SKU доступны данные о фактических продажах, наличии товара, ценах и промо-акциях.

Для выявления сезонности, для каждого ряда y_t применяется STL-декомпозиция [21], разбивающая ряд на три компонента:

$$y_t = T_t + S_t + R_t \quad (1)$$

где T_t – тренд (долгосрочная динамика), S_t – сезонный компонент (регулярные периодические колебания), R_t – остаток (нерегулярные колебания, шум).

Для количественной оценки выраженности сезонности вводится показатель *SeasonShare*:

$$SeasonShare = \frac{Var(S_t)}{Var(y_t)} \quad (2)$$

Значение *SeasonShare* принадлежит отрезку $[0,1]$ и характеризует долю общей дисперсии, объясняемую сезонностью. При малых значениях сезонность незначительна; при значениях выше заданного порога сезонность является значимым фактором вариации спроса.

Дополнительно рассчитывается автокорреляционная функция (Autocorrelation Function, ACF) на сезонном лаге (12 для месячных данных, 52 для недельных). Товарная позиция классифицируется как сезонная, если *SeasonShare* превышает эмпирически подобранный порог $\theta_s \in [0,2; 0,5]$ и значения автокорреляции на сезонном лаге выходят за пределы 95%-доверительного интервала. Порог θ_s рассматривается как настраиваемый параметр, подлежащий последующей калибровке в ходе вычислительных экспериментов.

Для прогнозирования спроса автор предлагает использовать три класса моделей:

- 1) линейные архитектуры DLinear и NLinear – модели долгосрочного прогнозирования временных рядов, в которых применяется внутренняя декомпозиция и/или нормализация, а затем линейное отображение на будущие значения;
- 2) XGBoost – реализация градиентного бустинга по деревьям решений, на вход которой подаются лаговые значения продаж, скользящие средние, календарные индикаторы, признаки промо-активности, ценовые переменные, а также показатели сезонности;
- 3) базовые эвристические модели прогнозирования (наивный прогноз, экспоненциальное сглаживание) в качестве эталонных моделей-бенчмарков.

Обучение выполняется по схеме расширяющегося окна для оценки устойчивости моделей во времени. Качество прогнозов предлагается оценивать метриками MAPE, MAE, RMSE с агрегацией по портфелю и подгруппам (сильная сезонность, слабая, стабильный спрос).

Далее, для количественной оценки волатильности, по каждому SKU рассчитываются коэффициенты вариации по исходному ($CV_i^{(orig)}$) и сезонно–скорректированному ($CV_i^{(adj)}$) рядам:

$$CV_i^{(orig)} = \frac{\sigma_i}{\mu_i + \varepsilon} \quad (3)$$

$$CV_i^{(adj)} = \frac{\sigma_i^{(adj)}}{\mu_i^{(adj)} + \varepsilon} \quad (4)$$

где μ_i, σ_i – среднее значение и стандартное отклонение исходного ряда y_t , $\mu_i^{(adj)}, \sigma_i^{(adj)}$ – среднее значение и стандартное отклонение по сезонно-скорректированному ряду $T_t + R_t$, ε – малая константа (например, 10^{-6}) для обеспечения численной устойчивости.

Коэффициент $CV^{(orig)}$ отражает общий разброс исходного ряда с учётом сезонности и полезен прежде всего для несезонных товаров. Коэффициент $CV^{(adj)}$ позволяет оценить остаточную волатильность после исключения регулярного сезонного компонента и более адекватен для сезонных SKU. Таким образом, пара $(CV^{(orig)}, CV^{(adj)})$ отражает различие между общей и истинно случайной вариацией.

Классификация товарных позиций по уровню риска производится путем адаптивной кластеризации:

- для несезонных SKU ($SeasonShare < \theta_s$) используется коэффициент $CV^{(orig)}$;
- для сезонных SKU ($SeasonShare \geq \theta_s$) применяется коэффициент $CV^{(adj)}$.

Кластеризация выполняется методом К-средних в пространстве $\log(CV)$, что позволяет стабилизировать масштаб данных и улучшить свойства выборки.

Мы предлагаем использовать три интуитивно интерпретируемых кластера:

- 1) L (Low) / SL (Seasonal Low) – низкая волатильность;
- 2) M (Medium) / SM (Seasonal Medium) – средняя волатильность;
- 3) H (High) / SH (Seasonal High) – высокая волатильность.

Границы кластеров, определённые по результатам кластеризации, при необходимости могут быть сглажены до удобных порогов (например, $CV < 0.3$ для категории L, $0.3 \leq CV \leq 0.8$ для M, $CV > 0.8$ для H). Параметры кластеризации рассматриваются как настраиваемые и подлежат анализу чувствительности.

В основе предложенного одношагового алгоритма распределения оборотного капитала лежит интегральный показатель – индекс доходности R_i :

$$R_i = T_i \cdot M_i. \quad (5)$$

где T_i – коэффициент оборачиваемости, M_i – маржинальная рентабельность.

Алгоритм распределения оборотного капитала включает следующие шаги [1]:

1. Определяется объём свободных финансовых ресурсов F_0 , заполняется таблица исходных данных: T_i , M_i , прогнозный объём продаж S_i , R_i .
2. Товары упорядочиваются по убыванию R_i .
3. Для каждого товара определяется минимальный запас $N_{i,\min}$, закупочная стоимость единицы товара F_i , вычисляется стоимость минимального запаса $N_{i,\min} \cdot F_i$. Рассчитывается суммарная потребность в финансовых ресурсах на закупку минимального запаса:

$$F_{\min} = \sum_{i=1}^A N_{i,\min} \cdot F_i \quad (6)$$

и остаток ресурсов для распределения:

$$F = F_0 - F_{\min}. \quad (7)$$

4. Для товара с текущим приоритетом i вычисляется количество единиц для дополнительной закупки:

$$w = \text{round}\left(\frac{F}{F_i}\right). \quad (8)$$

Если $w < 1$, алгоритм завершается. Иначе объём закупки определяется по правилу:

$$N_i = \begin{cases} N_{i,\min} + w, & \text{если } N_{i,\min} + w \leq S_i, \\ S_i, & \text{иначе.} \end{cases} \quad (9)$$

5. Обновляется остаток ресурсов по формуле:

$$F = F - F_i \cdot (N_i - N_{i,\min}), \quad (10)$$

после чего переход осуществляется к следующему товару.

Для учёта волатильности и сезонности целевой уровень запаса S_i корректируется в зависимости от класса риска:

$$S_i = N_{i,\min} + k_c \cdot \sigma_i^{(\text{prog})}, \quad (11)$$

где $\sigma_i^{(\text{prog})}$ – оценка стандартного отклонения прогнозного спроса, а k_c – коэффициент, зависящий от кластерной принадлежности (L, M, H).

Вводятся портфельные ограничения на суммарную долю бюджета, направляемого в высоковолатильные кластеры (например, H/SH), что позволяет ограничивать совокупный риск портфеля. Параметры k_c и пороговые доли бюджета рассматриваются как настраиваемые и подлежат последующей калибровке.

С вычислительной точки зрения алгоритм распределения капитала представляет собой однопроходную процедуру по упорядоченному списку товарных позиций. На этапе предварительной сортировки по индексу доходности R_i используется стандартный алгоритм сортировки со сложностью $O(n \log n)$, где n – число SKU; последующие итерации по списку выполняются за линейное время $O(n)$, поскольку каждая SKU обрабатывается не более одного раза. Таким образом, суммарная вычислительная сложность алгоритма составляет $O(n \log n)$. Алгоритм всегда завершается не более чем за n итераций, так как на каждом шаге либо полностью исчерпывается свободный бюджет F , либо индекс позиции i монотонно увеличивается, и при $i > n$ дальнейшее распределение становится невозможным при любом значении F .

На основе описанных процедур для каждой пары (SKU, период) формируется структурированный вектор признаков, объединяющий характеристики спроса, неопределённости и экономические параметры. Он служит единым формальным описанием состояния товарной позиции и используется в алгоритмах распределения оборотного капитала, а также как кандидат на пространство состояний для последующего применения в стохастических и динамических моделях управления запасами. Основные компоненты вектора представлены в таблице 1.

Дизайн вычислительных экспериментов. Поскольку в настоящей статье делается акцент на методологической составляющей, количественная верификация выносится в отдельный этап. Планируется проведение вычислительных экспериментов, включающих:

- сравнение точности прогнозов DLinear, NLinear, XGBoost и базовых статистических моделей на реальных данных селлеров маркетплейсов;
- верификацию корректности классификации сезонности по *SeasonShare* и автокорреляционной функции ACF;
- оценку устойчивости кластеров волатильности при изменении временного окна анализа;
- сравнение алгоритма распределения капитала с альтернативными стратегиями (по маржинальности, по оборачиваемости, по ABC-классификации) по показателям прибыли, рентабельности и дисперсии результата;
- анализ чувствительности по ключевым параметрам методики: порогу *SeasonShare* θ_s , границам кластеров, коэффициентам k_c и портфельным ограничениям.

Таблица 1. Структурированный вектор признаков

Признак	Описание
$\hat{y}$	Прогнозный спрос (точечная оценка)
CI_{low}, CI_{high}	Границы доверительного интервала прогноза
<i>SeasonShare</i>	Доля вариации, объясняемая сезонностью
is_seasonal	Бинарный флаг сезонности (0/1)
$CV^{(orig)}$	Коэффициент вариации исходного ряда
$CV^{(adj)}$	Коэффициент вариации сезонно–скорректированного ряда
cluster_label	Кластер волатильности (L/M/H или SL/SM/SH)
T_i	Оборачиваемость
M_i	Маржинальная рентабельность
R_i	Индекс доходности
$N_{i,min}$	Минимальный уровень запаса
S_i	Целевой (рекомендуемый) уровень запаса

Для оценки статистической значимости различий в качестве прогнозов могут использоваться критерии для сравнения прогнозных моделей, в частности тест Диболда-Мариано, с учётом поправок на множественные сравнения. Для оценки устойчивости кластеризации планируется использовать показатели внутреннего качества кластеров (например, силуэтный коэффициент) и долю сохраняющих кластерную принадлежность позиций при смещении окна анализа.

Обсуждение и заключение. Предложенная методика интегрирует методы анализа временных рядов, стохастические представления о спросе и алгоритм распределения оборотного капитала. Использование STL-декомпозиции, пары коэффициентов вариации и адаптивной кластеризации позволяет отделять регулярный сезонный компонент от остаточной волатильности, что важно для построения устойчивых решений в задачах управления запасами.

Предложенный алгоритм распределения капитала на основе индекса доходности, дополненный учётом прогнозного спроса, кластеризации по волатильности и портфельных ограничений, не выводится как строго оптимальное решение конкретной стохастической задачи. Его следует рассматривать, как структурированную эвристику, использующую агрегированные показатели доходности и риска, согласованные с классической теорией управления запасами, но адаптированные к ограничениям реальной практики (большое число SKU, ограниченные данные и вычислительные ресурсы). Такая постановка позволяет, с одной

стороны, улучшить качество решений по сравнению с простыми однокритериальными правилами, а с другой – подготовить основу для последующего перехода к более строгим стохастическим и DRL-моделям.

Статья носит методологический характер, она не предъявляет экспериментально подтверждённые выводы об оптимальности предложенного подхода, а задаёт формальную рамку, алгоритмическую структуру и программу последующей верификации. Научная значимость работы состоит в интеграции современных методов анализа временных рядов и идей стохастического управления запасами в прикладной задаче распределения оборотного капитала на маркетплейсах. Практическая значимость заключается в возможности реализации предложенной методики в составе систем поддержки принятия решений для продавцов.

Список источников

1. Бастанов А.Э. Оптимизация распределения оборотного капитала продавца на электронных маркетплейсах: алгоритмический подход на основе интегрального показателя / А.Э. Бастанов, И.В. Ахметов // Информационные технологии интеллектуальной поддержки принятия решений ITIDS'2025: труды XI Международной научной конференции. – Уфа: Уфимский университет науки и технологий, 2025. – Т. 1-2.
2. Дзури К. Управленческий и производственный учет; пер. с англ. – М.: ЮНИТИ-ДАНА, 2002. – 1071 с.
3. Стратегии, которые работают: подход BCG: сб. ст.; пер. с англ.; сост. К. Штерн, Д. Сток-мл. – 2-е изд. – М.: Манн, Иванов и Фербер, 2007. – 496 с.
4. Интрилигатор М.М. Математические методы оптимизации и экономическая теория; пер. с англ. – М.: Айрис-пресс, 2002. – 576 с.
5. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward. PLOS ONE, 2018, vol. 13, no. 3, e0194889, DOI: 10.1371/journal.pone.0194889.
6. Хомяченко О.В. Технология маркетинга на фирме. Экономический аспект. – М.: МЦФЭР, 1996.
7. Балашов В.Г. Технологии повышения финансового результата: практика и методы / В.Г. Балашов, В.А. Ириков. – М.: МЦФЭР, 2006. – 672 с.
8. Abolghasemi M. et al. Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. Computers & Industrial Engineering, 2020, vol. 142, p. 106380, DOI: 10.1016/j.cie.2020.106380.
9. Zeng A., Chen M., Zhang L., Xu Q. Are transformers effective for time series forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 2023, vol. 37, no. 9, pp. 11121-11128, DOI: 10.1609/aaai.v37i9.26317.
10. Brigato L., Morand R., Strømmen K. et al. Position: there are no champions in long-term time series forecasting. arXiv preprint arXiv:2502.14045, 2025, DOI:10.48550/arXiv.2502.14045.
11. Kim J., Kim H., Kim HG. et al. A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. Artificial Intelligence Review, 2025, vol. 58, no. 7, p. 216, DOI: 10.48550/arXiv.2411.05793.
12. Sarisa M. et al. An effective predicting e-commerce sales & management system based on machine learning methods. Journal of Artificial Intelligence and Big Data, 2020, vol. 1, no. 1, pp. 75–85, DOI:10.31586/jaibd.2020.1110.
13. Balusani A., Jai Ram Chowdary P., Paruchuri B.V.N. Enhancing retail demand forecasting with Xgboost: a comparative study with machine learning and deep learning models, DOI:10.2139/ssrn.5274510
14. Ji S., Wang X., Zhao W. et al. An application of a three-stage XGBoost-based model to sales forecasting of a cross-border e-commerce enterprise. Mathematical Problems in Engineering, 2019, vol. 2019, p. 8503252 (15 pages), DOI: 10.1155/2019/8503252.
15. Perera S.C., Sethi S.P. A survey of stochastic inventory models with fixed costs: optimality of (s, S) and (s, S)-type policies—continuous-time case. Production and Operations Management, 2023, vol. 32, no. 1, pp. 154-169, DOI:10.1111/poms.13819.
16. Ban G.-Y., Rudin C. The big data newsvendor: practical insights from machine learning. Operations Research, 2018, vol. 67, DOI:10.1287/opre.2018.1757.
17. Shi J. Application of the model combining demand forecasting and inventory decision in feature based newsvendor problem. Computers & Industrial Engineering, 2022, vol. 173, p. 108709, DOI: 10.1016/j.cie.2022.108709.
18. Gijbrecchts J., Boute R., Zhang D. Can deep reinforcement learning improve inventory management? performance on lost sales, dual-sourcing, and multi-echelon problems. Manufacturing & Service Operations Management, 2022, vol. 24, DOI: 10.1287/msom.2021.1064.

19. Dehaybe H., Catanzaro D., Chevalier P. Deep reinforcement learning for inventory optimization with non-stationary uncertain demand. *European Journal of Operational Research*, 2023, vol. 314, DOI: 10.1016/j.ejor.2023.10.007.
20. Хайндман Р. Прогнозирование: принципы и практика / Р. Хайндман, Д. Атанасопулос. – Москва: ДМК Пресс, 2023. – 459 с.

Бастанов Артур Эмильевич. Кандидат экономических наук, преподаватель кафедры экономики предпринимательства, Уфимский университет науки и технологий. AuthorID: 1329354, SPIN: 8314-0561, ORCID: 0009-0000-5922-2220, gradinv@gmail.com, Россия, Республика Башкортостан, Уфа, ул. Карла Маркса, 12.

UDC 658.7

DOI:10.25729/ESI.2026.43.3.012

Intelligent demand analysis and working capital allocation optimization for sellers on electronic marketplaces: methodological foundation, algorithm and design of experiments

Arthur E. Bastanov

Ufa University of Science and Technology, Russia, Ufa, gradinv@gmail.com

Abstract. The article presents the methodological outcome of the first stage of a research program aimed at improving inventory and working capital management for sellers operating on electronic marketplaces. The relevance of the study stems from the fact that practical tasks of allocating limited working capital across numerous product items are solved under conditions of stochastic demand, pronounced seasonality, heterogeneous volatility and constrained computational resources, whereas traditional inventory management methods either remain static or require a priori specification of demand and cost parameters. The paper proposes a comprehensive methodology for demand analysis and working capital allocation that comprises three related components: intelligent demand analysis based on seasonal-trend decomposition and machine learning models for time series (DLinear, NLinear, XGBoost); extended analysis of demand volatility using two variants of the coefficient of variation and adaptive clustering of the assortment matrix; and a working capital allocation algorithm built on an integral «profitability index» and supplemented by seasonality- and risk-aware adjustments. A key element of the methodology is the separation of predictable seasonal variation from residual volatility, which enables more accurate assessment of the risk associated with individual product items and the use of these risk estimates in working capital allocation decisions. The methodology yields a structured feature vector that includes forecast, statistical and economic characteristics of product items and is suitable both for one-step decision-support systems and as a state-space representation for subsequent stochastic and dynamic inventory control models. The scientific contribution of the work lies in integrating known tools for forecasting, volatility assessment and working capital allocation into a single mathematically consistent methodological model that accounts for the specifics of electronic marketplaces and the particular constraints faced by sellers. The article substantiates the proposed methodology, describes its architecture and presents the design of computational experiments intended for subsequent empirical evaluation of its effectiveness.

Keywords: marketplace, seller, demand forecasting, seasonality, coefficient of variation, profitability index, inventory management, working capital management, machine learning methods

References

1. Batanov A.E., Akhmetov I.V. Optimizatsiya raspredeleniya oborotnogo kapitala seller na elektronnykh marketpleysakh: algoritmicheskiy podkhod na osnove integral'nogo pokazatelya [Optimization of seller's working capital allocation on e-marketplaces: an algorithmic approach based on an integral indicator]. *Informatsionnyye tekhnologii intellektual'noy podderzhki prinyatiya resheniy ITIDS'2025: Trudy XI Mezhdunarodnoy nauchnoy konferentsii* [Information Technologies for Intelligent Decision Support ITIDS'2025: Proceedings of the XI International Scientific Conference]. Ufa, Ufa University of Science and Technology Publ., 2025, vols. 1-2.
2. Drury C. Upravlencheskiy i proizvodstvennyy uchet [Management and cost accounting]. Trans. from English. Moscow, UNITI-DANA Publ., 2002, 1071 p.

3. Stern C., Stalk G. (eds.). Strategii, kotoryye rabotayut: podkhod BCG [Strategies that work: the BCG approach]. Trans. from English. 2nd ed. Moscow, Mann, Ivanov i Ferber Publ., 2007, 496 p.
4. Intriligator M.M. Matematicheskiye metody optimizatsii i ekonomicheskaya teoriya [Mathematical optimization and economic theory]. Trans. from English. Moscow, Ayris-press Publ., 2002, 576 p.
5. Makridakis S., Spiliotis E., Assimakopoulos V. Statistical and machine learning forecasting methods: concerns and ways forward. PLOS ONE, 2018, vol. 13, no. 3, e0194889, DOI: 10.1371/journal.pone.0194889.
6. Khomyachenko O.V. Tekhnologiya marketinga na firme. Ekonomicheskiy aspekt [Marketing technology in a firm. Economic aspect]. Moscow, MTsFER Publ., 1996.
7. Balashov V.G., Irikov V.A. Tekhnologii povysheniya finansovogo rezul'tata: praktika i metody [Technologies for improving financial performance: practice and methods]. Moscow, MTsFER Publ., 2006, 672 p.
8. Abolghasemi M. et al. Demand forecasting in supply chain: the impact of demand volatility in the presence of promotion. Computers & Industrial Engineering, 2020, vol. 142, p. 106380, DOI: 10.1016/j.cie.2020.106380.
9. Zeng A., Chen M., Zhang L., Xu Q. Are transformers effective for time series forecasting? Proceedings of the AAAI Conference on Artificial Intelligence, 2023, vol. 37, no. 9, pp. 11121-11128, DOI: 10.1609/aaai.v37i9.26317.
10. Brigato L., Morand R., Strømmen K. et al. Position: there are no champions in long-term time series forecasting. arXiv preprint arXiv:2502.14045, 2025, DOI:10.48550/arXiv.2502.14045.
11. Kim J., Kim H., Kim HG. et al. A comprehensive survey of deep learning for time series forecasting: architectural diversity and open challenges. Artificial Intelligence Review, 2025, vol. 58, no. 7, p. 216, DOI: 10.48550/arXiv.2411.05793.
12. Sarisa M. et al. An effective predicting e-commerce sales & management system based on machine learning methods. Journal of Artificial Intelligence and Big Data, 2020, vol. 1, no. 1, pp. 75–85, DOI:10.31586/jaibd.2020.1110.
13. Balusani A., Jai Ram Chowdary P., Paruchuri B.V.N. Enhancing retail demand forecasting with Xgboost: a comparative study with machine learning and deep learning models, DOI:10.2139/ssrn.5274510
14. Ji S., Wang X., Zhao W. et al. An application of a three-stage XGBoost-based model to sales forecasting of a cross-border e-commerce enterprise. Mathematical Problems in Engineering, 2019, vol. 2019, p. 8503252 (15 pages), DOI: 10.1155/2019/8503252.
15. Perera S.C., Sethi S.P. A survey of stochastic inventory models with fixed costs: optimality of (s, S) and (s, S)-type policies—continuous-time case. Production and Operations Management, 2023, vol. 32, no. 1, pp. 154-169, DOI:10.1111/poms.13819.
16. Ban G.-Y., Rudin C. The big data newsvendor: practical insights from machine learning. Operations Research, 2018, vol. 67, DOI:10.1287/opre.2018.1757.
17. Shi J. Application of the model combining demand forecasting and inventory decision in feature based newsvendor problem. Computers & Industrial Engineering, 2022, vol. 173, p. 108709, DOI: 10.1016/j.cie.2022.108709.
18. Gijbrecchts J., Boute R., Zhang D. Can deep reinforcement learning improve inventory management? performance on lost sales, dual-sourcing, and multi-echelon problems. Manufacturing & Service Operations Management, 2022, vol. 24, DOI: 10.1287/msom.2021.1064.
19. Dehaybe H., Catanzaro D., Chevalier P. Deep reinforcement learning for inventory optimization with non-stationary uncertain demand. European Journal of Operational Research, 2023, vol. 314, DOI: 10.1016/j.ejor.2023.10.007.
20. Hyndman R., Athanasopoulos G. Prognozirovaniye: printsipy i praktika [Forecasting: principles and practice]. Moscow, DMK Press Publ., 2023, 459 p.

Bastanov Artur Emilievich. PhD in economics of the department of economics of entrepreneurship, Ufa university of sciences and technologies, gradinv@gmail.com. AuthorID: 1329354, SPIN: 8314-0561, ORCID: 0009-0000-5922-2220, gradinv@gmail.com, Russia, Republic of Bashkortostan, Ufa, Karl Marx St., 12.

Статья поступила в редакцию 23.01.2026; одобрена после рецензирования 11.03.2026; принята к публикации 02.08.2026.

The article was submitted 01/23/2026; approved after reviewing 03/11/2026; accepted for publication 08/02/2026.

Оптимизация расписания авиакомпании с учётом моделирования трансферных пассажиропотоков

Карамушко Денис Маркович, Чернов Алексей Владимирович

Московский физико-технический институт,

Россия, Долгопрудный, karamushko.dm@phystech.edu

Аннотация. Планирование расписания авиакомпании можно представить, как несколько оптимизационных задач: определение набора рейсов и времени их выполнения, назначение типов воздушных судов (ВС) и построение их маршрутов, разработка расписания экипажей. В настоящей работе предлагается математическая модель для назначения типов воздушных судов на рейсы и определения времени их вылета в хабовой маршрутной сети (hub-and-spoke). Модель сформулирована в виде задачи смешанного целочисленного линейного программирования на основе пространственно-временного графового представления расписания полётов. Ключевой отличительной особенностью модели является явный учёт стыковочных маршрутов пассажиров: в целевую функцию включены доходы как от прямых (т.н., point-to-point), так и от трансферных перевозок, наряду с операционными расходами на эксплуатацию воздушных судов. Размерность задачи снижена за счет учета особенностей оборота ВС для хабовой маршрутной сети: рейсы туда-обратно между хабом и регионом рассматривались как единое целое. Модель верифицирована на реальных данных крупной российской авиакомпании, охватывающих 260 ежедневных рейсов, агрегированных в 146 оборотных рейсов и рейсов между хабами, 61 воздушное судно 5 различных типов. Для валидации использовались два источника данных о спросе: собственные данные авиакомпании и данные, синтезированные по гравитационной модели. Предложена двухшаговая схема оптимизации: на первом шаге выполняется поиск времени вылета в диапазоне в более широком диапазоне с большим интервалом, на втором оно уточняется в узком диапазоне с шагом 5 минут. Оптимизированное расписание обеспечило увеличение прогнозируемой операционной прибыли на 4-9%, а также рост числа доступных направлений с временем стыковки не более трёх часов на 19%. Полученные результаты свидетельствуют о том, что одновременная оптимизация временных параметров расписания и назначения флота позволяет достичь существенного коммерческого эффекта и обеспечивает пассажирам более комфортные условия для совершения пересадок

Ключевые слова: оптимизация расписания авиакомпании, назначение типов воздушных судов, линейное программирование, математическое моделирование

Цитирование: Карамушко Д.М. Оптимизация расписания авиакомпании с учётом моделирования трансферных пассажиропотоков / Д.М. Карамушко, А.В. Чернов // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 154-164. – DOI:10.25729/ESI.2026.43.3.013.

Введение. В авиационной отрасли, как и во многих других, возникает задача планирования производства. Её неотъемлемой частью является задача планирования сезонного расписания. Как правило, разработка расписания начинается приблизительно за год до начала сезона. Эксперты определяют маршрутную сеть, частоты полетов по направлениям, время выполнения каждого рейса, продолжительность полетов, тип воздушного судна (ВС) для выполнения рейса. Пример такого решения: рейс 3017 из Москвы в Иркутск будет выполняться ежедневно на Airbus A320, вылетать в 21:30 и прилетать в 8:15.

Задача планирования расписания авиакомпании традиционно декомпозируется на ряд подзадач, представленных в таблице 1.

За 1 год фиксируют набор рейсов для выполнения, а также времена вылета и прилета для каждого рейса. Также в этот период распределяют типы воздушных судов (ВС) по рейсам, оценивается выполнимость оборота воздушных судов и формируется черновой план оборота на уровне типов ВС. Задача маршрутизации на уровне конкретных самолетов, как правило, решается приблизительно за месяц до вылета. Параметры рейса, определённые на годовом горизонте планирования, включая время выполнения и тип воздушного судна, могут

подвергаться изменениям вплоть до момента вылета в ответ на колебания спроса и воздействие внешних факторов.

Таблица 1. Подзадачи планирования расписания авиакомпании

Подзадача (Этап)	Описание	Время принятия решения (до отправления)
Определение времен выполнения рейсов	Множество рейсов с временем вылета/прилета	1 год
Распределение флота	Для каждого рейса определен тип ВС	1 год
Построение оборота	Маршруты (набор рейсов) для каждого ВС	1–3 месяца
Планирование экипажей	Маршруты экипажей, формирование ростера	1 месяц

Обзор основных публикаций, посвященных задаче планирования расписания авиакомпании, можно найти в [1-3]. В работах рассматриваются разные подзадачи или их комбинации, разные целевые функции и ограничения из-за специфики определенных авиакомпаний. Своими особенностями обладает постановка задачи планирования для авиакомпании с хабовой структурой маршрутной сети (т.н. hub-and-spoke). В ней существует несколько ключевых аэропортов (хабов), и все рейсы выполняются между хабами и регионами или между хабами [4]. Также в маршрутной сети такого типа высокая доля трансфера в хабах: значительная часть пассажиров делает пересадку и продолжает путешествие на другом рейсе. Таким образом, авиакомпания стремится обслужить как можно большее количество прибыльных рынков, формируемых из соединяемых через хаб пар городов. Некоторые перевозчики группируют рейсы в хабах в волны стыковок – временные окна, в которых несколько рейсов прибывают и отправляются с небольшим интервалом, давая возможность пересадки пассажирам.

В [5] была представлена модель для построения расписания и назначения типов ВС на рейсы с учетом трансферных маршрутов пассажиров. Основываясь на ней, авторы [6] предложили робастную модель для расписания без волн стыковок, которая максимизирует количество стыковок с учетом прогнозируемой выручки, а также учитывает возможные корректировки из-за колебаний спроса. Авторы [7] рассмотрели увеличение устойчивости расписания за счет минимизации числа пассажиров с сорванными стыковками. В работе [8], одной из немногих, совместно оптимизируется трансферный потенциал и операционная прибыль авиакомпании. Также допускается изменение времени вылета и прилета рейсов, при этом заранее генерируются все возможные трансферные маршруты пассажиров над копиями рейсов и не учитываются особенности хабовой структуры сети, что существенно увеличивает размер модели. В [9] представлена модель оптимизации волн стыковок для уменьшения суммарного стыковочного времени пассажиров с учетом ограничений по слотам в аэропортах. Многокритериальная оптимизация расписания аэропорта, где одновременно максимизируется число доступных стыковок и минимизируется количество требуемых изменений слотов, предложена в [10]. Автор [11] решает задачи проектирования парка ВС и расстановки ВС по авиалиниям совместно, считая спрос заданным экспертно в нечёткой форме.

Целью настоящей работы является разработка модели оптимизации расписания авиакомпании, которая одновременно учитывает возможность стыковок в зависимости от выбранного времени выполнения рейсов и операционную прибыль перевозчика с хабовой структурой маршрутной сети.

1. Постановка задачи

1.1. Описание процесса построения расписания. На практике разработка нового расписания полётов представляет собой итеративный процесс, который начинается с анализа

операционных и финансовых показателей действующего расписания. Для каждого рейса рассчитываются ключевые метрики: коэффициент занятости кресел, выручка на креслокилометр (RASK) и операционные расходы на креслокилометр (CASK). Это позволяет выявить убыточные рейсы и направления с неудовлетворённым спросом, требующие корректировки провозных ёмкостей.

На основе проведённого анализа начинается формирование нового расписания путем внесения изменения в старое. Во-первых, тип воздушного судна, назначенного на рейс, может быть изменен на более вместительный при устойчиво высоком спросе либо на менее вместительный в случае систематического избытка провозной емкости, ведущего к снижению доходности. Во-вторых, корректируется частота выполнения рейсов: по убыточным направлениям она сокращается, по прибыльным – увеличивается.

При корректировках расписания нужно валидировать его исполнимость: проверить, будет ли достаточно имеющегося флота для выполнения всех рейсов, выстроится ли оборот воздушных судов. Если полученное расписание исполнить нельзя, то пересматривают ранее принятые решения по флоту и частоте или сдвигают время выполнения рейсов.

Таким образом, текущий процесс построения расписания авиакомпании носит выраженный итеративный характер. Каждая итерация приближает эксперта к расписанию, удовлетворяющему одновременно экономическим требованиям и операционным ограничениям, однако гарантия нахождения глобально оптимального решения при таком подходе отсутствует, что обуславливает актуальность применения формальных методов математической оптимизации.

В рамках оптимизационной модели, представленной в данной статье, частоты выполнения рейсов рассматриваются, как фиксированные входные параметры, допускается изменять времена выполнения рейсов и переназначать типы воздушных судов, тогда как добавление новых рейсов в расписание и исключение существующих из него не предусмотрены.

1.2. Трансферные перевозки. Ключевой особенностью маршрутной сети с хабовой структурой является возможность сквозных перевозок через узловые аэропорты по единому билету. Это позволяет авиакомпаниям соединять существенно большее количество городов по сравнению с маршрутной сетью типа point-to-point.

Вместе с тем далеко не каждая комбинация прибывающего и вылетающего рейсов в узловом аэропорту образует дает возможность пересадки: наличие стыковки между рейсами зависит от времени между рейсами. Снизу оно ограничено минимальным временем пересадки (MCT – minimum connection time). Это значение устанавливается аэропортом и может зависеть от пар терминалов и типов стыкующихся рейсов (внутренний – внутренний, внутренний – международный и т.д.). Гарантируется, что пассажир успеет пройти все предполетные формальности за минимальное время пересадки.

Верхняя граница времени стыковки (MaxCT – maximum connection time) технически ограничена 24 часами, но авиакомпания может условно считать стыковки активными с временем пересадки не более 4-6 часов. Превышение данного порога делает суммарное время в пути непривлекательным по сравнению с альтернативными маршрутами.

Будем называть стыковку между рейсами $i, j \in F$ допустимой если время стыковки между ними, которое рассчитывается как разность между временем отправления вылетающего рейса j (STD_j) и временем прибытия прилетающего рейса i (STA_i), лежит в интервале между MCT и $MaxCT$.

$$CT_{ij} = STD_j - STA_i$$

$$MCT \leq CT_{ij} \leq MaxCT$$

Таким образом, в ходе построения расписания определяется набор активных стыковок.

1.3. Источники данных для построения расписания. Для оптимизации расписания используются следующие данные:

1. Расписание авиакомпании – множество рейсов F предыдущего периода планирования.

Под рейсом понимается перелет воздушного судна между аэропортами отправления и назначения. Для каждого рейса известны:

- уникальный номер;
- аэропорты отправления и прибытия;
- время вылета и время прилета;
- границы для изменения времени вылета;
- множество типов ВС, на которых этот рейс может быть выполнен;

Будем считать, что все рейсы выполняются ежедневно.

2. Флот авиакомпании – множество типов воздушных судов K . Для каждого типа определено M_k – количество доступных самолётов.

3. c_k^i – операционные расходы при выполнении рейса i воздушным судном типа k .

4. Данные о тарифах:

- $fare_i$ – средний тариф на перелет рейсом $i \in F$;
- $fare_{ij}$ – средний тариф на перелет по маршруту, состоящему из рейсов $i, j \in F$.

Обычно $fare_{ij} \neq fare_i + fare_j$, так как тариф определяется в первую очередь спросом, а не расходами [12].

5. Данные о спросе:

- $demand_i$ – спрос на перелет рейсом $i \in F$;
- $demand_{ij}$ – спрос на перелет по маршруту, состоящему из рейсов $i, j \in F$.

Они представляют собой усредненные оценки, построенные на основе исторических данных бронирований за несколько предыдущих аналогичных сезонов.

Для каждого рейса $f \in F$ требуется найти наиболее оптимальный тип воздушного судна $k \in K$ и время вылета по критерию максимизации операционной прибыли. Операционная прибыль рассчитывается, как разность выручки от перевозки пассажиров и расходов на выполнение рейсов.

2. Сетевое представление расписания. Представим расписание рейсов в виде пространственно-временного графа. Такой подход является одним из общепринятых при моделировании движения самолетов и используется, например, в [13]. По горизонтали откладывается временная шкала, по вертикали перечисляются аэропорты. В графе узлы отражают события вылета и прилета рейсов, а ребра бывают нескольких видов:

- рейсовые – обозначают рейсы, соединяют узлы в разных аэропортах;
- наземные – обозначают стоянку и разворот воздушного судна в аэропорту, соединяют узлы в одном аэропорту;
- ночные – подвид наземных ребер, соединяют начало и конец периода в каждом аэропорту, в котором допускается ночная стоянка воздушных судов для отражения цикличности. На рисунке 1 это аэропорты 1, 4.

Для авиакомпании с хабовой структурой маршрутной сети можно упростить граф. При таком подходе рейсы всегда разворотные, выполняются между хабом (крупным базовым аэропортом) и регионом. Поэтому можно заменить рейсы на круги – парные рейсы туда-обратно. Если хабов несколько, то рейсы между ними в пары не объединяются для сохранения возможности размена воздушных судов между базами. Таким образом, на графе расписания

авиакомпания с хабовой структурой достаточно отобразить хабовые аэропорты. Рейсовое ребро будет обозначать либо круг, либо рейс между хабами.

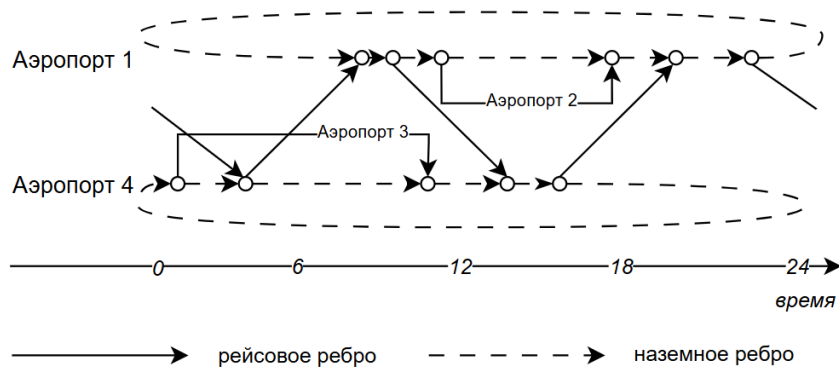


Рис. 1. Сетевое представление расписания на сутки для хабовой структуры маршрутной сети

3. Математическая модель

3.1. Переменные

- $x_k^{i,t_i} = \begin{cases} 1, & \text{если рейс } i \text{ с отправлением в момент времени } t_i \text{ выполняется типом ВС } k, \\ 0, & \text{иначе} \end{cases}$
- y_a^k – число ВС типа k на наземном ребре a ,
- время отправления рейса i :

$$D_i = \sum_{t \in T_i} \sum_{k \in K} t \cdot x_k^{i,t}$$

- BT_i – длительность (block time) рейса i (от момента отправления ВС со стоянки до прибытия на стоянку, включая время руления), предполагаем, что BT не зависит от типа ВС,
- время прибытия рейса i : $A_i = D_i + BT_i$
- $\delta_{ij} = \begin{cases} 1, & \text{если стыковка между рейсами } i \text{ и } j \text{ возможна,} \\ 0, & \text{иначе.} \end{cases}$
- rax_i – количество point-to-point пассажиров, перевозимых на рейсе i (следующих только на этом рейсе),
- rax_{ij} – количество трансферных пассажиров, перевозимых на рейсах i, j .

3.2. Множества

- L – множество аэропортов в маршрутной сети авиакомпании;
- K – множество типов воздушных судов;
- F – множество рейсов;
- F_k – множество рейсов, которые могут выполняться типом ВС $k \in K$;
- Q – множество пар рейсов $(i, j) \in F$ между которыми возможна активная стыковка (аэропорт прилета одного рейса совпадает с аэропортом вылета другого, а время на пересадку лежит в интервале между минимальным и максимальным стыковочным временем);
- $Q_{market} \subseteq Q$ – множество трансферных рынков – пар рейсов $(i, j) \in Q$ с активной стыковкой и с достаточным трансферным потоком (большей некоторой заранее фиксированной константы) для учета этой пары в модели: $demand_{ij} \geq const$;
- T_i – множество возможных времен вылета рейса i ;
- $K_i \subseteq K$ – множество типов воздушных судов, на которых возможно выполнение рейса i , с учетом эксплуатационных ограничений;
- N – множество узлов графа расписания;

- G – множество наземных ребер графа расписания.

3.3. Параметры

- $dep_i \in L$ – аэропорт вылета рейса i ;
- $arr_i \in L$ – аэропорт прилета рейса i ;
- $fare_i$ – средний тариф на перелет рейсом i ;
- $fare_{ij}$ – средний тариф на перелет рейсами i, j . Причем $fare_{ij} \neq fare_i + fare_j$;
- M^k – число воздушных судов типа k ;
- c_k^i – операционные расходы при выполнении рейса i воздушным судном типа k ;
- $demand_i$ – спрос на перелет рейсом i ;
- $demand_{ij}$ – спрос на перелет рейсами i, j .

3.4. Целевая функция. Максимизация операционных доходов.

$$\sum_{i \in F} fare_i * pax_i + \sum_{i \in F} \sum_{j \in F} fare_{ij} * pax_{ij} - \sum_{i \in F} \sum_{t_i \in T_i} \sum_{k \in K} c_k^i * x_k^{i, t_i} \rightarrow \max$$

3.5. Ограничения

1. Покрытие: на каждый рейс назначен ровно один тип воздушного судна.

$$\sum_{t \in T_i} \sum_{k \in K} x_k^{i, t} = 1, \quad \forall i \in F$$

2. Непрерывность потока воздушных судов: число входящих самолетов в каждый узел равно числу исходящих.

$$y_{n-}^k + \sum_{i \in I(k, n)} x_k^{i, t - BT_i} = y_{n+}^k + \sum_{i \in O(k, n)} x_k^{i, t}, \quad \forall n \in N, \quad \forall k \in K$$

где:

- $n = (l, t)$, т.е. $l \in L$ – пространственный компонент узла n , а именно – аэропорт из маршрутной сети, а t – временной;
- y_{n+}^k – число ВС типа k на наземном ребре, исходящем из узла n , т.е. соединяющим узлы $n = (l, t)$ и $n_+ = (l, t + 1)$;
- $O(k, n)$ – множество рейсов, выполняемых типом ВС k , исходящих из узла n : $O(k, n) = \{i \in F_k \mid dep(i) = l, t \in T_i\}$;
- y_{n-}^k – число ВС типа k на наземном ребре, входящем в узел n , т.е. соединяющим узлы $n = (l, t)$ и $n_- = (l, t - 1)$;
- $I(k, n)$ – множество рейсов, выполняемых типом ВС k , входящих в узел n : $I(k, n) = \{i \in F_k \mid arr(i) = l, t - BT_i \in T_i\}$;

6. Размер флота: потребное число воздушных судов для выполнения рейсов не превышает доступное.

$$\sum_{a \in CG(k, t_0)} y_a^k + \sum_{i \in F} \sum_{t \in T_i} x_k^{i, t} \leq M^k, \quad \forall k \in K$$

где:

- y_a^k – число ВС типа k на наземном ребре a ;
 - t_0 – произвольный момент времени, через который проводится линия счета. Достаточно проверки ограничения на размер флота в произвольный момент времени, так как если оно будет выполнено в нем, то и в любые другие моменты времени оно также будет выполнено из-за наличия условия непрерывности потока;
 - $CG(k, t_0)$ – множество наземных ребер, пересекаемых линией счета в момент времени t_0 ;
7. Вместимость самолета: сумма пассажиров point-to-point, входящего трансфера и исходящего трансфера не превышает вместимости выбранного типа ВС.

$$\sum_{j \in F} \text{pax}_{ij} + \text{pax}_i + \sum_{j \in F} \text{pax}_{ji} \leq \sum_{t_i \in T_i} \sum_{k \in K} \text{cap}_k * x_k^{i,t_i} \quad \forall i \in F$$

8. Условие активности стыковки:

$$\begin{aligned} D_j - A_i &\geq MCT - M(1 - \delta_{ij}) \quad \forall (i, j) \in Q \\ D_j - A_i &\leq \text{MaxCT} + M(1 - \delta_{ij}) \quad \forall (i, j) \in Q \end{aligned}$$

где $M = 25$ часов.

9. Количество перевозимых пассажиров не превосходит спрос (для всех рейсов и активных стыковок):

$$\begin{aligned} \text{pax}_{ij} &\leq \text{demand}_{ij} \delta_{ij} \quad \forall (i, j) \in Q_{\text{market}} \\ \text{pax}_i &\leq \text{demand}_i \quad \forall i \in F \end{aligned}$$

4. Анализ размерности задачи. Оценим размер задачи. В качестве примера приведём характерные значения размерности для расписания крупной российской авиакомпании со следующими параметрами: $|F_{\text{all}}| = 260$ исходных плеч (до агрегации в обороты), $|F| = 146$ оборотных рейсов и рейсов между хабами, $T = 3$ варианта времени выполнения каждого рейса, $|K| = 5$ типов ВС, $|N| = 522$ узла и $|G| = 522$ наземных рёбер в графе расписания.

4.1. Число переменных. Модель содержит пять групп переменных (переменные D_i и A_i являются линейными выражениями через $x_k^{i,t}$ и самостоятельными переменными не считаются), которые перечислены в таблице 2.

Таблица 2. Группы переменных, их количество и значения для примера расчёта

Группа переменных	Объекты индексации	Количество	Пример
$x_k^{i,t}$	рейс $i \in F$, время вылета $t \in T_i$, допустимый тип $k \in K_i$	$\sum_{i \in F} T_i K_i $	1833
y_a^k	наземное ребро $a \in G$ (включая ночные), тип ВС $k \in K$	$ K \cdot G $	2610
δ_{ij}	стыковочная пара-кандидат $(i, j) \in Q$	$ Q $	1457
pax_{ij}	трансферный рынок $(i, j) \in Q^{\text{market}}$	$ Q^{\text{market}} $	1457
pax_ℓ	плечо $\ell \in F_{\text{all}}$, вариант вылета $t \in T_\ell$	$ F_{\text{all}} \cdot T$	780

Суммарное число переменных:

$$V = \sum_{i \in F} |T_i| |K_i| + |K| \cdot |G| + |Q| + |Q^{\text{market}}| + |F_{\text{all}}| \cdot T$$

При равномерном числе вариантов времени вылета для каждого рейса $|T_i| = T$ и количестве возможных типов ВС для выполнения рейса $|K_i| \approx K$ суммарное число переменных будет выражаться как:

$$V = |F| T K + |K| \cdot |G| + |Q| + |Q^{\text{market}}| + |F_{\text{all}}| \cdot T$$

Размерность переменных растёт линейно по каждому из параметров $|F|$, $|F_{\text{all}}|$, T , $|K|$, $|G|$, $|Q|$. При этом в общем случае количество пар рейсов $|Q|$, между которыми возможна стыковка, будет расти квадратично по числу рейсов. Однако, оно ограничено временным фильтром (рассматриваем только стыковки с временем пересадки в пределах от MCT до MaxCT) и шириной окна разрешенных сдвигов рейсов по времени. Также на $|Q|$ влияет наличие волн стыковок в исходном расписании и распределение рейсов между хабами. Эти ограничения существенно уменьшают число возможных пар рейсов для пересадки, что удерживает задачу в практически решаемых размерах.

4.2. Число ограничений. Модель содержит семь групп ограничений, которые перечислены в таблице 3.

Таблица 3. Группы ограничений, их количество и значения для примера расчёта

Ограничение	Объекты индексации	Количество	Пример
Покрытие рейсов	рейс $i \in F$	$ F $	146
Непрерывность потока ВС	узел $n \in N$, тип ВС $k \in K$	$ N \cdot K $	2610
Размер флота	тип ВС $k \in K$	$ K $	5
Вместимость ВС	плечо $\ell \in F_{all}$, вариант $t \in T_\ell$	$ F_{all} \cdot T$	780
Активность стыковки (нижняя и верхняя границы)	стыковочная пара $(i, j) \in Q$	$2 Q $	2914
Трансферный спрос	трансферный рынок $(i, j) \in Q^{market}$	$ Q^{market} $	1457
Спрос point-to-point	плечо $\ell \in F_{all}$, вариант $t \in T_\ell$	$ F_{all} \cdot T$	780

Суммарное число ограничений:

$$C = |F| + |N| \cdot |K| + |K| + 2|F_{all}|T + 2|Q| + |Q^{market}|$$

5. Вычислительные эксперименты. Расчет выполнялся на данных крупной российской авиакомпании с несколькими хабовыми аэропортами. Парк воздушных судов авиакомпании состоит из 61 ВС 5 типов. Расписание составлялось на сезон Зима 2025-2026 на основе данных сезона Зима 2024-2025. Расписание включало 260 ежедневных рейсов. Их количество было сокращено до 146 за счет объединения в оборотные рейсы пары "туда-обратно". При этом рейсы между хабами в оборотные рейсы не объединялись.

Для проверки модели использовались два источника данных о спросе на перевозку: собственные данные авиакомпании и сгенерированные по гравитационному методу на основе данных о населении в городах и расстоянии между ними [14]. Для каждого оборотного рейса рассматривались несколько вариантов изменения времени вылета: ± 1 час с интервалом 1 час; ± 1 час с интервалом 30 минут; ± 1 час с интервалом 5 минут. Помимо этого, использовалась двухшаговая оптимизация: сначала задача решается в диапазоне ± 1 час с интервалом 30 минут, затем на основе полученного решения ищется новое в диапазоне ± 30 минут с интервалом 5 минут. Это позволяет избежать роста размерности задачи при уменьшении рассматриваемых интервалов. Расчет производился с помощью солвера HiGHS версии 1.14.0.

Результаты численных экспериментов представлены в таблицах 4, 5. В них можно увидеть эффект от применения двухшаговой схемы оптимизации. В эксперименте 2 одноэтапная оптимизация с поиском времени вылета в диапазоне ± 1 час с интервалом 5 минут была остановлена по лимиту времени (1 час) с MIP-gap = 7,82% и значением целевой функции 273,41, в то время как использование двухшаговой схемы позволило найти лучшее решение (288,62), при этом MIP-gap в 2,91% был достигнут спустя 10 минут работы решателя.

Таблица 4. Параметры и характеристики экспериментов

№ эксп.	Кол-во шагов	Источник спроса	Окно, интервал, мин.	Число переменных	Число ограничений	Разреженность	t реш., с	MIP-gap, %
1	1	грав.	$\pm 60, 30$	9799	9242	$3,43 \cdot 10^{-4}$	971	1,5
2	1	грав.	$\pm 60, 5$	78589	96012	$3,40 \cdot 10^{-5}$	3607	7,82
3	2	грав.	Шаг 1: $\pm 60, 30$; Шаг 2: $\pm 60, 5$;	32790	37019	$8,84 \cdot 10^{-5}$	3603	1,64
4	1	грав.	$\pm 60, 60$	6767	6637	$4,62 \cdot 10^{-4}$	3601	31,88

Таблица 5. Результаты экспериментов: изменение целевой функции и числа стыковок

№ эксп.	ЦФ до, у.е.	ЦФ после, у.е.	Изменение	Стыковки до, шт	Стыковки после, шт.	Изменение
1	26340	28759	9,2%	57	68	19,3%

2	26340	27341	3,8%	57	64	12,3%
3	26340	28862	9,6%	68	69	1,5%
4	-4086	-3882	5,0%	122	126	3,3%

На собственных данных авиакомпании расчет происходил с меньшей скоростью сходимости: при лимите в 1 час решение завершилось с MIP-gap = 31,9%. Причина такого поведения может заключаться в структуре трансферных потоков: в собственных данных их больше по количеству, что приводит к кратному росту LP-итераций. Увеличение устойчивости модели к данным по спросу является одной из тем дальнейших исследований. Тем не менее полученное решение обеспечивает прирост операционной прибыли на 5%, что подтверждает практическую ценность подхода даже при незакрытом зазоре оптимальности.

В результате прогнозируемая операционная прибыль выросла на 4-9% за счет корректировки времен вылета и оптимального подбора типов ВС под каждый рейс. Общее число трансферных направлений, по которым время пересадки не превышает 3 часов, выросло на 19%, что повысило транспортную доступность и связность сети авиакомпании.

Заключение. В работе рассматривается задача построения расписания и назначения воздушных судов для авиакомпаний с хабовой структурой маршрутной сети (hub-and-spoke). Предложена модель смешанного целочисленного линейного программирования, в которой учитывается выбор времени вылета, выбор типа воздушного судна для каждого рейса, а в качестве критерия оптимизации выступает операционная прибыль, рассчитываемая с учетом выручки от трансферных перевозок.

Вычислительные эксперименты, проведенные на реальных данных крупной российской авиакомпании, подтвердили практическую применимость предложенного подхода. Оптимизированное расписание обеспечило увеличение операционной прибыли и рост числа трансферных направлений с временем пересадки, не превышающим трёх часов, на 19% относительно базового расписания. Полученные результаты свидетельствуют о том, что незначительные корректировки времени вылета рейсов в сочетании с переназначением воздушных судов способны существенно повысить связность сети и коммерческие показатели без изменения маршрутной сети и частоты выполнения рейсов.

Вместе с тем необходимо обозначить ряд ограничений настоящего исследования. Модель предполагает фиксированную частоту рейсов, в связи с чем на начальном этапе может потребоваться её ручная корректировка. Кроме того, показатели пассажирского спроса и тарифов рассматриваются, как детерминированные параметры. С одной стороны, использование средних значений представляется обоснованным, поскольку исследование ориентировано на планирование расписания, а не на ценообразование и управление доходами. С другой стороны, спрос и тарифы могут сами зависеть от расписания.

В рамках дальнейших исследований возможно развитие модели по нескольким направлениям. Во-первых, включение рейсов, которые выполняются не ежедневно, в предложенный подход расширит область его практического применения. Во-вторых, моделирование изменения спроса на изменение расписания обеспечит более реалистичные оценки выручки. Во-третьих, включение мер по обеспечению устойчивости – таких, как буферное время и штрафы за распространение задержек – повысит надёжность формируемых расписаний.

Список источников

1. Xu Y., Wandelt S., Sun X. Airline scheduling optimization: literature review and a discussion of modelling methodologies. Intelligent Transportation Infrastructure, 2024, vol. 3, DOI: 10.1093/iti/liad026.
2. Zhou L., Liang Z., Chou C.A. et al. Airline planning and scheduling: models and solution methodologies. Frontiers of Engineering Management, 2020, vol. 7, no. 1, pp. 1–26, DOI: 10.1007/s42524-020-0093-5.

3. Eltoukhy A.E.E., Chan F.T.S., Chung S.H. Airline schedule planning: a review and future directions. *Industrial Management & Data Systems*, 2017, vol. 117, no. 6, pp. 1201–1243, DOI: 10.1108/IMDS-09-2016-0358.
4. Goedeking P. *Networks in aviation: strategies and structures*. Berlin, Heidelberg, Springer, 2010.
5. Lohatepanont M., Barnhart C. Airline schedule planning: integrated models and algorithms for schedule design and fleet assignment. *Transportation Science*, 2004, vol. 38, no. 1, pp. 19–32, DOI:10.1287/trsc.1030.0026.
6. Jiang H., Barnhart C. Robust airline schedule design in a dynamic scheduling environment. *Computers & Operations Research*, 2013, vol. 40, no. 3, pp. 831–840, DOI: 10.1016/j.cor.2011.06.018.
7. Cadarso L., Marín Á. Robust passenger oriented timetable and fleet assignment integration in airline planning. *Journal of Air Transport Management*, 2013, vol. 26, pp. 44–49, DOI: 10.1016/j.jairtraman.2012.10.004.
8. Sherali H.D., Bae K.H., Haouari M. A benders decomposition approach for an integrated airline schedule design and fleet assignment problem with flight retiming, schedule balance, and demand recapture. *Annals of Operations Research*, 2013, vol. 210, pp. 213–244, DOI: 10.1007/s10479-011-0906-3.
9. Çiftçi U., Özkır V. Optimising flight connection times in airline bank structure through simulated annealing and tabu search algorithms. *Journal of Air Transport Management*, 2020, vol. 87, p. 101858, DOI: 10.1016/j.jairtraman.2020.101858.
10. Mei T., Ma L., Ma Y. Flight schedule adjustment for hub airports using multi-objective optimization. *Journal of Intelligent Systems*, 2021, vol. 30, no. 1, pp. 931–946, DOI: 10.1515/jisys-2020-0114.
11. Романенко В.А. Нечёткая оптимизация управления парком воздушных судов с учётом трансферных пассажиропотоков / В.А. Романенко // *Управление большими системами*, 2025. – Вып. 117. – С. 74–102.
12. Belobaba P., Odoni A., Barnhart C. *The global airline industry*. Chichester, John Wiley & Sons, 2009, DOI: 10.1002/9780470744734.
13. Hane C., Barnhart C., Johnson E.L. et al. The fleet assignment problem: solving a large-scale integer program. *Mathematical Programming*, 1995, vol. 70, no. 1–3, pp. 211–232, DOI: 10.1007/BF01585938.
14. Grosche T., Rothlauf F., Heinzl A. Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, 2007, vol. 13, no. 4, pp. 175–183, DOI: 10.1016/j.jairtraman.2007.02.001.

Карамушко Денис Маркович. Аспирант МФТИ. AuthorID: 1353167, SPIN: 6139-6197, ORCID: 0009-0006-7905-5216, 141700, Россия, Московская область, Долгопрудный, Институтский пер., 9.

Чернов Алексей Владимирович. К.ф.-м.н., с. н. с., зав. лаб. численных методов прикладной структурной оптимизации МФТИ. AuthorID: 987214, SPIN: 7878-3230, ORCID: 0000-0003-2122-054X, 141700, Россия, Московская область, Долгопрудный, Институтский пер., 9.

UDC 519.852, 656.7.043

DOI:10.25729/ESI.2026.43.3.013

Airline schedule optimization with transfers modelling

Denis M. Karamushko, Alexey V. Chernov

Moscow Institute of Physics and Technology,

Russia, Dolgoprudny, karamushko.dm@phystech.edu

Abstract. Airline schedule planning is a complex multi-stage optimization process, including schedule design, fleet assignment, and aircraft routing. This paper proposes an integrated mathematical model for the joint optimization of fleet type assignment and flight departure times determination within a hub-and-spoke network structure. The model is formulated as a mixed-integer linear program built upon a space-time graph representation of the flight schedule. A key distinguishing feature of the model is the handling of connecting passenger itineraries: revenues from both point-to-point and transfer passengers are included in the objective function, along with fleet operating costs. Problem dimensionality is reduced by exploiting the rotation structure of hub-and-spoke networks: outbound and return legs between a hub and a spoke region are treated as a single rotation unit. The model was validated on real-world data from a major Russian airline, comprising 260 daily flights aggregated into 146 rotations and inter-hub flights, operated by a fleet of 61 aircraft across 5 fleet types. Two independent demand data sources were used for validation: the airline's own booking data and estimates generated via a gravity model. A two-step optimization scheme is proposed: in the first step, departure times are searched over a wider range with

a coarser interval; in the second step, the solution is refined over a narrower range with a 5-minute resolution. The optimized schedule achieved a 4–9% increase in forecasted operational profit and a 19% improvement in the number of reachable transfer markets within a three-hour connection window. These results demonstrate that the simultaneous optimization of schedule timing and fleet assignment yields meaningful commercial gains and gives more convenient connection opportunities for passengers.

Keywords: airline schedule optimization, fleet assignment model, linear programming, mathematical modelling

References

1. Xu Y., Wandelt S., Sun X. Airline scheduling optimization: literature review and a discussion of modelling methodologies. *Intelligent Transportation Infrastructure*, 2024, vol. 3, DOI: 10.1093/iti/liad026.
2. Zhou L., Liang Z., Chou C.A. et al. Airline planning and scheduling: models and solution methodologies. *Frontiers of Engineering Management*, 2020, vol. 7, no. 1, pp. 1–26, DOI: 10.1007/s42524-020-0093-5.
3. Eltoukhy A.E.E., Chan F.T.S., Chung S.H. Airline schedule planning: a review and future directions. *Industrial Management & Data Systems*, 2017, vol. 117, no. 6, pp. 1201–1243, DOI: 10.1108/IMDS-09-2016-0358.
4. Goedeking P. *Networks in aviation: strategies and structures*. Berlin, Heidelberg, Springer, 2010.
5. Lohatepanont M., Barnhart C. Airline schedule planning: integrated models and algorithms for schedule design and fleet assignment. *Transportation Science*, 2004, vol. 38, no. 1, pp. 19–32, DOI: 10.1287/trsc.1030.0026.
6. Jiang H., Barnhart C. Robust airline schedule design in a dynamic scheduling environment. *Computers & Operations Research*, 2013, vol. 40, no. 3, pp. 831–840, DOI: 10.1016/j.cor.2011.06.018.
7. Cadarso L., Marín Á. Robust passenger oriented timetable and fleet assignment integration in airline planning. *Journal of Air Transport Management*, 2013, vol. 26, pp. 44–49, DOI: 10.1016/j.jairtraman.2012.10.004.
8. Sherali H.D., Bae K.H., Haouari M. A benders decomposition approach for an integrated airline schedule design and fleet assignment problem with flight retiming, schedule balance, and demand recapture. *Annals of Operations Research*, 2013, vol. 210, pp. 213–244, DOI: 10.1007/s10479-011-0906-3.
9. Çiftçi U., Özkır V. Optimising flight connection times in airline bank structure through simulated annealing and tabu search algorithms. *Journal of Air Transport Management*, 2020, vol. 87, p. 101858, DOI: 10.1016/j.jairtraman.2020.101858.
10. Mei T., Ma L., Ma Y. Flight schedule adjustment for hub airports using multi-objective optimization. *Journal of Intelligent Systems*, 2021, vol. 30, no. 1, pp. 931–946, DOI: 10.1515/jisys-2020-0114.
11. Romanenko V.A. Nechetkaya optimizatsiya upravleniya parkom vozdukhnykh sudov s uchetom transfornykh passazhiropotokov [Fuzzy optimization of fleet assignment with transfer passenger flows]. *Upravleniye bol'shimi sistemami [Large-Scale Systems Management]*, 2025, iss. 117, pp. 74–102.
12. Belobaba P., Odoni A., Barnhart C. *The global airline industry*. Chichester, John Wiley & Sons, 2009, DOI: 10.1002/9780470744734.
13. Hane C., Barnhart C., Johnson E.L. et al. The fleet assignment problem: solving a large-scale integer program. *Mathematical Programming*, 1995, vol. 70, no. 1–3, pp. 211–232, DOI: 10.1007/BF01585938.
14. Grosche T., Rothlauf F., Heinzl A. Gravity models for airline passenger volume estimation. *Journal of Air Transport Management*, 2007, vol. 13, no. 4, pp. 175–183, DOI: 10.1016/j.jairtraman.2007.02.001.

Karamushko Denis Markovich. PhD student at MIPT. AuthorID: 1353167, SPIN: 6139-6197, ORCID: 0009-0006-7905-5216. 9 Institutsky per., Dolgoprudny, MO, 141700

Chernov Alexey Vladimirovich. PhD, senior researcher, chief of Laboratory of Numerical Methods of Applied Structural Optimization at MIPT, AuthorID: 987214, SPIN: 7878-3230, ORCID: 0000-0003-2122-054X, 9 Institutsky per., Dolgoprudny, MO, 141700.

Статья поступила в редакцию 11.06.2026; одобрена после рецензирования 16.06.2026; принята к публикации 01.08.2026.

The article was submitted 06/11/2026; approved after reviewing 06/16/2026; accepted for publication 08/01/2026.

Программные системы и комплексы

УДК 004.032.26

DOI:10.25729/ESI.2026.43.3.014

Разработка гибридного модуля сопровождения воздушных объектов на основе GRU и фильтра Калмана

Фелагерева Евгения Юрьевна, Рудоман Нэлли Радиковна,

Новиков Александр Дмитриевич

Кубанский государственный университет, Россия, Краснодар, efelagereva@gmail.com

Аннотация. В статье рассматривается задача устойчивого сопровождения воздушных объектов в видеопотоке, с акцентом на предсказание их траектории при наличии шумов измерений и кратковременных потерь детекции. Целью исследования является разработка гибридного метода трекинга, объединяющего классический фильтр Калмана и рекуррентную нейронную сеть типа GRU для повышения точности и устойчивости прогноза движения объектов. Предлагаемый подход основывается на использовании фильтра Калмана для оценки текущего состояния объекта, заданного вектором координат ограничивающей рамки YOLO и скоростями $[c_x, c_y, w, h, v_x, v_y]$, и нейросетевой модели GRU, которая обучается предсказывать изменение скорости $\Delta v_x, \Delta v_y$ на основе последовательности предыдущих состояний. В отличие от традиционных методов, где модель движения задается априорно, в предлагаемой работе динамика движения частично извлекается из данных. Предсказания GRU интегрируются в модель фильтра Калмана через механизм адаптивного сглаживания с коэффициентом α , что позволяет гибко регулировать вклад нейросетевого прогноза в итоговую оценку состояния. Основное содержание работы включает описание архитектуры гибридной модели, методики подготовки обучающего датасета, а также алгоритма интеграции нейросетевых предсказаний в процесс фильтрации. Особое внимание уделено задаче долгосрочного прогноза – до 30 кадров вперед – что является критически важным для систем активного сопровождения и управления. Научная новизна работы заключается в предложении способа комбинирования GRU и фильтра Калмана через адаптивное сглаживание управляющего воздействия, что позволяет учитывать нелинейные и нестационарные характеристики движения объектов без отказа от интерпретируемости классической модели. В отличие от существующих решений, предлагаемый метод обеспечивает более стабильную работу в условиях шумных измерений и пропусков детекций. В результате проведенных экспериментов показано, что предложенный подход повышает точность предсказания траектории на 16,57% и снижает накопление ошибки при многошаговом прогнозировании по сравнению с классическим фильтром Калмана и базовыми нейросетевыми моделями на 21%. Разработанный метод может быть применен в задачах детекции и сопровождения беспилотных летательных аппаратов, системах видеонаблюдения и автономной навигации.

Ключевые слова: фильтр Калмана, GRU (Gated Recurrent Unit – управляемый рекуррентный блок), БПЛА (беспилотный летательный аппарат), позиционирование

Цитирование: Фелагерева Е.Ю. Разработка гибридного модуля сопровождения воздушных объектов на основе GRU и фильтра Калмана / Е.Ю. Фелагерева, Н.Р. Рудоман, А.Д. Новиков // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 165-177. – DOI:10.25729/ESI.2026.43.3.014.

Введение. Задача автоматической детекции и сопровождения воздушных объектов является одной из ключевых в современных системах компьютерного зрения и автономного управления. Особую сложность представляет сопровождение малых и быстро движущихся объектов, таких, как беспилотные летательные аппараты (БПЛА), на значительных дистанциях, где наблюдаются низкое разрешение цели, шумы измерений и нестабильность детекции. В таких условиях стандартные методы трекинга часто демонстрируют снижение точности и устойчивости, особенно при наличии пропусков обнаружения и резких изменений траектории движения.

Классическим подходом к решению задачи сопровождения является использование фильтра Калмана, который обеспечивает эффективную оценку состояния объекта на основе линейной динамической модели и зашумленных измерений. Несмотря на широкое распространение, данный метод предполагает априорно заданную модель движения, которая,

как правило, не отражает реальную нелинейную и нестационарную динамику воздушных объектов. Это приводит к накоплению ошибки при длительном прогнозировании и снижению качества сопровождения в сложных сценариях.

В последние годы активно развиваются методы, основанные на применении нейронных сетей, в частности рекуррентных архитектур, таких, как GRU (Gated Recurrent Unit – управляемый рекуррентный блок) и LSTM (Long Short-Term Memory – долгая краткосрочная память), способных извлекать временные зависимости из данных и моделировать сложные динамические процессы. Однако полностью нейросетевые подходы, как правило, требуют значительных объемов данных, обладают меньшей интерпретируемостью и могут демонстрировать нестабильность при работе в условиях шумных измерений.

В связи с этим перспективным направлением является разработка гибридных методов, сочетающих преимущества классических алгоритмов фильтрации и нейросетевых моделей. Такие подходы позволяют сохранить устойчивость и интерпретируемость фильтра Калмана, одновременно обогащая модель движения за счет данных.

В настоящей работе предлагается гибридный метод сопровождения объектов, в котором фильтр Калмана используется для оценки текущего состояния объекта, а рекуррентная нейронная сеть типа GRU – для предсказания изменения скорости на основе последовательности предыдущих состояний. Отличительной особенностью подхода является способ интеграции нейросетевого прогноза: предсказанные изменения скорости рассматриваются, как управляющее воздействие и вводятся в модель фильтра Калмана с использованием механизма адаптивного сглаживания. Это позволяет гибко регулировать влияние нейросети и повышает устойчивость системы к ошибкам предсказания.

Дополнительно в работе рассматривается задача многошагового прогнозирования траектории движения объекта на горизонт до 30 кадров вперед, что имеет практическое значение для систем активного сопровождения и управления. Предлагаемый метод ориентирован на работу в условиях ограниченных вычислительных ресурсов и может быть применен в реальном времени.

Таким образом, основной вклад работы заключается в разработке устойчивого гибридного алгоритма сопровождения, учитывающего сложную динамику движения объектов и обеспечивающего повышение точности прогноза в условиях шумных и неполных данных.

1. Смежные работы. В области сопровождения и прогнозирования траектории движения воздушных объектов активно применяются как классические методы фильтрации, так и современные подходы на основе глубокого обучения. Рассмотрим наиболее релевантные из них.

Традиционно для задачи трекинга используется линейный фильтр Калмана, так как движение объекта между соседними кадрами почти линейно, и использовании нелинейного алгоритма избыточно и не даст нужного прироста качества. Линейный фильтр предполагает заранее заданную модель движения с постоянной скоростью. Такой подход демонстрирует высокую эффективность при краткосрочном прогнозировании, однако при увеличении горизонта предсказания его точность существенно снижается. Это связано с тем, что линейная модель не учитывает ускорения и манёвры объекта, вследствие чего ошибка оценки состояния накапливается, особенно при прогнозировании на десятки кадров вперед.

В работе «VECTOR: Velocity-Enhanced GRU Neural Network for Real-Time 3D UAV Trajectory Prediction» [1] предложен подход к прогнозированию траектории БПЛА с использованием рекуррентной нейронной сети GRU, обученной на последовательностях скоростей. В отличие от моделей, использующих только координаты, авторы показывают, что моделирование изменений скорости позволяет значительно повысить точность предсказания,

особенно при нелинейной динамике движения и резких манёврах. Экспериментальные результаты подтверждают возможность применения модели в реальном времени.

Схожий подход представлен в статье «GRU-Based Deep Learning Framework for Real-Time UAV Trajectory Prediction» [2], где GRU применяется для анализа временных зависимостей в траекториях движения БПЛА. Авторы демонстрируют, что нейросетевые методы превосходят классические модели при моделировании сложных траекторий, однако отмечают снижение устойчивости при наличии шумов и ограниченного объема обучающих данных.

Наряду с полностью нейросетевыми подходами, активно развиваются гибридные методы, объединяющие рекуррентные сети и фильтр Калмана. В работе «Multiple Variational Kalman-GRU for Ship Trajectory Prediction» [3] предложена архитектура, в которой GRU используется для оценки параметров динамической модели, а фильтр Калмана выполняет вероятностную оценку состояния. Показано, что такая комбинация позволяет учитывать нелинейные зависимости и снижает ошибку прогнозирования по сравнению с классическим фильтром.

Аналогичный подход рассмотрен в статье «GRU-UKF Trajectory Prediction with Attention» [4], где выход рекуррентной сети интегрируется в модифицированный фильтр Калмана. Авторы показывают, что объединение нейросетевого предсказания и фильтрации позволяет существенно повысить точность и устойчивость системы в условиях шумных измерений.

Таким образом, существующие исследования показывают, что:

- линейный фильтр Калмана обеспечивает устойчивость, но ограничен простой моделью движения [5];
- рекуррентные нейронные сети эффективно моделируют сложную динамику, но менее устойчивы к шумам [6];
- гибридные подходы позволяют объединить преимущества обоих методов.

В отличие от рассмотренных работ, в настоящей статье предлагается использовать GRU для предсказания изменения скорости, а не абсолютного состояния объекта, с последующей интеграцией этого прогноза в линейный фильтр Калмана через механизм адаптивного сглаживания. Это позволяет компенсировать накопление ошибки при длительном прогнозировании и повысить устойчивость сопровождения без усложнения базовой модели движения.

2. Предлагаемый гибридный метод. Общая схема работы алгоритма включает три основных этапа: получение измерений от детектора [7], оценку состояния с использованием фильтра Калмана и коррекцию скорости объекта с помощью нейросетевого прогноза при отсутствии наблюдений.

В качестве детектора выступает дообученная модель YOLO11s, предсказывающая 3 класса объектов: самолетный и мультироторный типы, птицы. Птицы нужны для того, чтобы при их детекции не сопровождать.

В качестве наблюдений используются выходы YOLO – координаты ограничивающей рамки:

$$z_t = [c_x, c_y, w, h], \quad (1)$$

где c_x – центр рамки по оси x ; c_y – центр рамки по оси y ; w – ширина рамки; h – высота рамки.

Вектор состояния системы задаётся в следующем виде:

$$x_t = [c_x, c_y, w, h, v_x, v_y], \quad (2)$$

где v_x – скорость по оси x ; v_y – скорость по оси y .

2.1 Модель фильтра Калмана. Для оценки состояния используется линейный фильтр Калмана с моделью равномерного движения [8]. Был выбран именно линейный фильтр,

потому что движение объектов между соседними кадрами около линейное, и использование нелинейного алгоритма здесь избыточно. Динамика системы описывается выражением:

$$x_{t+1} = Fx_t + w_t \quad (3)$$

где F – матрица перехода, w_t – шум процесса.

Матрица перехода состояния имеет вид:

$$F = \begin{bmatrix} 1 & 0 & 0 & 0 & dt & 0 \\ 0 & 1 & 0 & 0 & 0 & dt \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

Шум процесса:

$$w_t \sim N(0, Q), \quad (5)$$

где Q – ковариация шума процесса.

Модель наблюдения:

$$z_t = Hx_t + v_t, \quad (6)$$

где H – матрица наблюдения; v_t – шум измерения.

Матрица наблюдения имеет вид:

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (7)$$

Шум измерения:

$$v_t \sim N(0, R), \quad (8)$$

где R – ковариация шума измерений.

Фильтр выполняет стандартные этапы прогнозирования и коррекции. При наличии детекции происходит обновление состояния по измерению, что позволяет компенсировать шумы и стабилизировать оценку [9].

2.2 Нейросетевая модель GRU. Для учета сложной динамики движения используется рекуррентная нейронная сеть типа GRU [10], состоящая из последовательных блоков (рис.1). На вход сети подаётся последовательность длины N , состоящая из предыдущих оценок состояния: $\{x_{t-N}, \dots, x_t\}$, где каждый вектор состояния представлен формулой 2.

Выходом сети является вектор приращения скорости:

$$\Delta v_t = \begin{bmatrix} \Delta v_x \\ \Delta v_y \end{bmatrix} \quad (9)$$

Таким образом, GRU моделирует изменение скорости объекта на основе временного контекста, что позволяет учитывать нелинейные траектории и манёвры, но нелинейность здесь скрыта не между соседними кадрами, а в последовательности кадров [11].

2.3 Гибридная схема интеграции. В отличие от классических гибридных подходов, в предлагаемом методе нейросетевой прогноз используется только в условиях отсутствия детекции.

Если на текущем кадре измерение отсутствует, выполняется только шаг прогнозирования фильтра Калмана по формуле 3.

При наличии достаточной истории состояний (N шагов) формируется входная последовательность для GRU, на основе которой вычисляется приращение скорости Δv_t .

Коррекция скорости осуществляется по правилу адаптивного сглаживания:

$$v_x = (1 - \alpha)v_x^{KF} + \alpha\Delta v_x, \quad (10)$$

$$v_y = (1 - \alpha)v_y^{KF} + \alpha\Delta v_y \quad (11)$$

где v_x^{KF} , v_y^{KF} – скорость, полученная после шага предсказания фильтра Калмана; Δv_x , Δv_y – изменение скорости, предсказанное GRU; α – коэффициент сглаживания, определяющий вклад нейросетевой коррекции. Структура блока GRU показана на рис. 1.

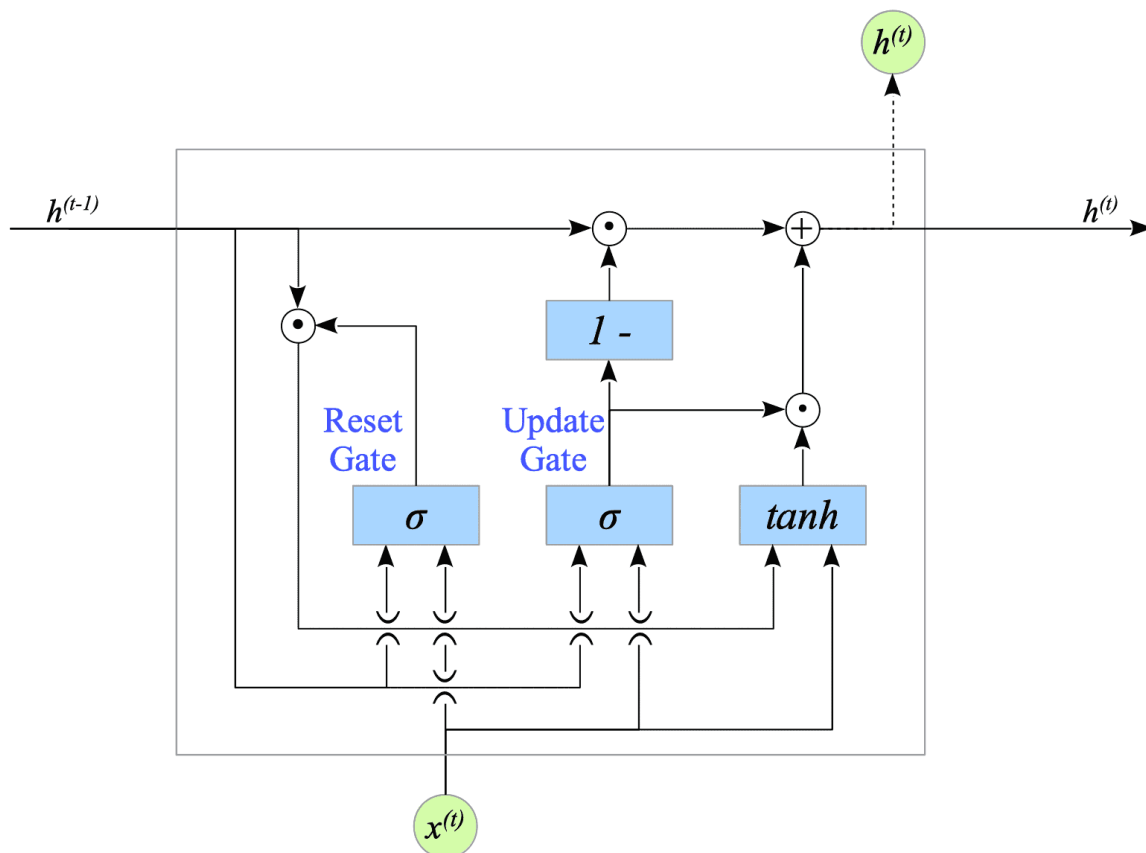


Рис. 1. Блок GRU

Величина α адаптивно изменяется в зависимости от длительности отсутствия детекции:

- при кратковременных пропусках используется большее значение α , что позволяет учитывать динамику движения;
- при увеличении горизонта прогнозирования значение α уменьшается, снижая влияние нейросети и предотвращая накопление ошибки.

Дополнительно вводится ограничение на величину скорости:

$$v_x, v_y \in [-v_{max}, v_{max}], \quad (12)$$

где v_{max} – максимальная скорость, определяемая из распределения скоростей реальных данных.

2.4 Алгоритм гибридного сопровождения. Работа метода на каждом временном шаге включает следующие этапы:

1. Получение измерения z_t от детектора (формула 1).
2. Если измерение доступно:
 - выполняется предсказание и обновление фильтра Калмана.
3. Если измерение отсутствует:
 - выполняется только шаг предсказания;
 - формируется последовательность состояний длины N , состоящая из предыдущих оценок состояния: $\{x_{t-N}, \dots, x_t\}$, где каждый вектор состояния представлен формулой 2;
 - предсказывается Δv_t с использованием GRU;
 - корректируются скорости с использованием адаптивного сглаживания.

4. Обновление состояния объекта.
5. Добавление текущего состояния в буфер истории.

Таким образом, фильтр Калмана обеспечивает устойчивую оценку состояния при наличии наблюдений, а GRU компенсирует ограничения линейной модели в условиях пропусков детекции, улучшая качество прогнозирования траектории.

3. Обучение модели GRU. Для создания датасета использовалась обученная YOLO11s, с помощью которой были обработаны видеопотоки мультироторного дрона. Стоит отметить, что для датасета не имеет разницы, какой класс снимать, важны только координаты детекции модели для формирования итоговых последовательностей. При съемке датасета важны были разнообразные маневры дрона, чтобы нейросеть училась на сложных последовательностях.

Чтобы получить итоговый вектор состояния (формула 2), нужно получить скорости, так как на выходе YOLO получаем только вектор из формулы 1. Для этого, а также, чтобы сгладить дрожание рамки, убрать пропуски (несколько секунд) и продлить треки, был применен фильтр Калмана. Как и рассматривалось ранее, была взята линейная модель фильтра, потому что между кадрами объект движется линейно с постоянной скоростью. Фильтр Калмана предсказывал только на 5 кадров вперед, потому что далее из-за линейной модели, рамка начинала «улетать» резко, что далее будет исправлять нейросеть. Итоговая архитектура обработки данных для подачи на вход в модель представлена на рис. 2.

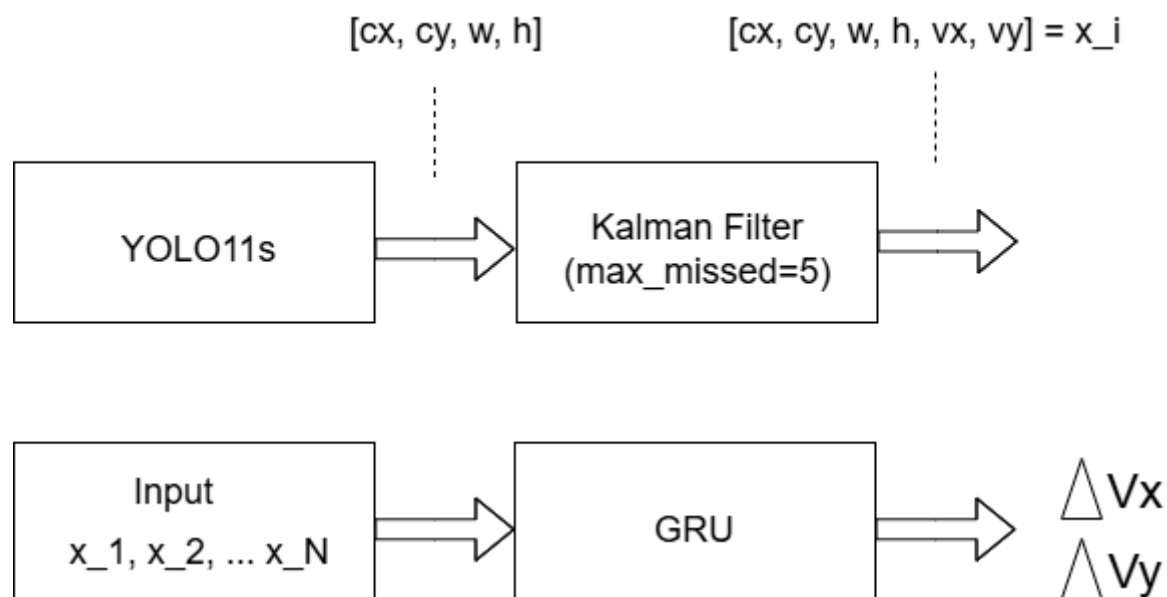


Рис. 2. Архитектура обработки данных для обучения GRU

Для подачи на вход нейросети применялись скользящие окна длиной 30 кадров, на которых сеть обучалась предсказывать приращения скорости на следующие 30 кадров.

Анализ данных показывает интересные закономерности:

- YOLO ограничивающая рамка и скорость: средние координаты объектов примерно по центру кадра ($c_x \approx 0,53$, $c_y \approx 0,43$), средний размер рамки небольшой ($w \approx 0,15$, $h \approx 0,12$), но наблюдаются редкие крупные значения до 0,97 и 0,77, что отражает сильное изменение масштаба цели. Средняя скорость детектора невелика (0,0128), однако встречаются быстрые движения до 0,835. Это говорит о разнообразии траекторий, чтобы GRU обучалась на сложных данных.
- Дрожание и сглаживание: YOLO имеет заметное дрожание ($\text{mean} \approx 0,025$, $\text{max} \approx 0,91$), фильтр Калмана снижает среднее дрожание до 0,021, демонстрируя эффективность линейного сглаживания.

- Ускорения и предсказания GRU: изменения скорости (Δv) достаточно малы (порядка $\sim 0,001$), из-за плавности большинства траекторий, но встречаются резкие скачки до 0,16 по оси x и 0,12 по оси y . Более того, больше половины значений колеблются около 0, а значит, требуется стандартизировать данные. После нормализации $\text{std} = 1$. Это позволяет сети GRU обучаться на единообразной шкале изменений скорости, учитывая как типичные малые приращения, так и редкие резкие манёвры без того, чтобы выбросы доминировали в процессе обучения. Такой подход улучшает сходимость модели и устойчивость предсказаний на сложных траекториях.
- Автокорреляция: $\Delta v \approx 0,53$ на 1 кадр вперед, что показывает сильную временную зависимость. Эти свойства делают рекуррентные модели верным выбором для предсказания будущей динамики.

Так как в данных обнаружились редкие выбросы, и они сильно влияют на распределение и плохо сказываются на нормализации, то были удалены последовательности, выходящие за пределы 1 и 99 перцентилей.

В совокупности статистика подтверждает, что подготовленный датасет обеспечивает достаточное разнообразие движений и сложность, позволяя GRU эффективно изучать временные зависимости, а фильтр Калмана – стабилизировать данные и компенсировать шум детектора.

Нейросеть имеет один GRU-слой с 64 скрытыми нейронами и последовательную голову из двух линейных слоёв с нормализацией и ELU-активацией. Для оптимизации использовался AdamW с начальной скоростью обучения $3 \cdot 10^{-4}$ и уменьшением LR по Plateau. В качестве функции потерь применялось среднеквадратичное отклонение (MSE), чтобы модель корректно училась на выбросах [12].

Модель обучалась на платформе Kaggle с использованием NVIDIA Tesla T4 с батчами по 128 последовательностей. Для предотвращения переобучения использовался ранний стоп при отсутствии улучшения валидационной ошибки в течение 12 эпох, а градиенты обрезались (clip) для стабильности. График обучения приведен на рис. 3.

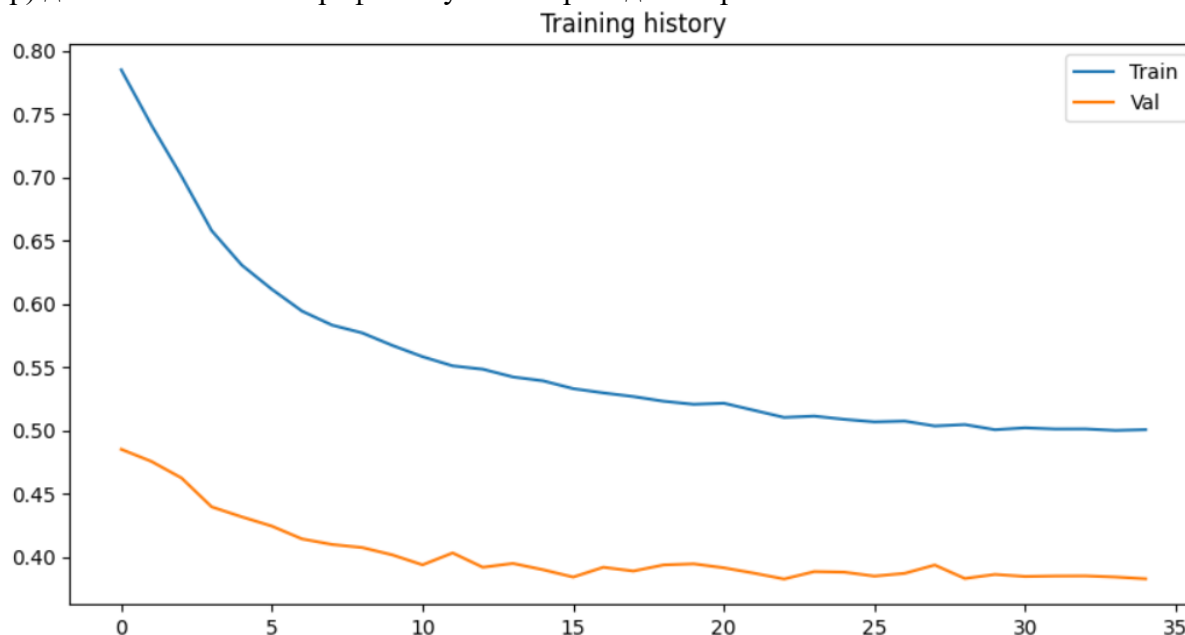


Рис. 3. График обучения GRU

Анализ процесса обучения показал стабильное снижение ошибки как на обучающей, так и на валидационной выборках:

- начальная ошибка (Epoch 1): Train 0,785, Val 0,485;

– наилучшая эпоха (Epoch 35, ранний стоп): Train 0,501, Val 0,383.

Это подтверждает хорошую сходимость модели и способность GRU учить сложные зависимости в последовательностях.

Так как в данных присутствует большое количество значений, близких к нулю, возникла необходимость проверить, не предсказывает ли модель всегда ноль. В противном случае ошибка могла бы оставаться относительно небольшой, но модель не выполняла бы фактическое предсказание скоростей.

Для оценки этого на тестовой выборке были рассчитаны метрики RMSE и MAE для базового сценария, когда выход модели всегда равен нулю, и сравнивались с фактическими предсказаниями модели [13]. Полученные результаты приведены в таблице 1 и демонстрируют, что модель GRU корректно предсказывает изменения скорости, значительно превосходя базовый вариант.

Таблица 1. Сравнение выходов GRU с константным выходом 0

Метрика	Baseline	GRU
MAE	0,396	0,333
RMSE	0,654	0,558

По метрикам видно, что выходы GRU лучше baseline на 19% по MAE и на 17% по RMSE, значит, модель выучила последовательности и эффективнее учитывает редкие резкие манёвры, так как RMSE чувствительна к выбросам.

Таким образом, GRU обеспечивает существенное улучшение точности прогноза по сравнению с наивным подходом, подтверждая целесообразность использования рекуррентной сети в гибридной системе сопровождения.

4. Разработка итогового алгоритма автоматической детекции и сопровождения БПЛА. В рамках работы был реализован итоговый алгоритм сопровождения воздушного объекта, объединяющий результаты детекции, фильтрацию состояния и нейросетевую коррекцию динамики. Основной задачей являлось обеспечение устойчивого трекинга при пропусках детекции и резких изменениях траектории движения.

Алгоритм работает в покадровом режиме и включает два основных состояния: наличие детекции и её отсутствие.

При наличии детекции координаты ограничивающего прямоугольника, полученные от модели YOLO, используются для обновления состояния с помощью фильтра Калмана. Это позволяет компенсировать шумы измерений и стабилизировать положение объекта.

В случае пропуска детекции активируется режим прогнозирования. На первом этапе выполняется предсказание состояния фильтром Калмана. Далее, при наличии достаточной истории (не менее 30 предыдущих состояний), формируется входная последовательность для GRU, которая предсказывает приращение скорости объекта. Полученное значение используется для коррекции скоростей фильтра Калмана с применением адаптивного сглаживания.

Коэффициент влияния GRU зависит от длительности пропуска детекции:

- в первые кадры пропуска используется большее значение (0,7), что позволяет учитывать текущую динамику движения;
- при увеличении горизонта прогнозирования влияние снижается (0,35), предотвращая накопление ошибки.

Дополнительно вводится ограничение на максимальную скорость, что обеспечивает устойчивость алгоритма и исключает выбросы при некорректных предсказаниях.

Таким образом, алгоритм сочетает:

- устойчивость фильтра Калмана при наличии наблюдений;

- способность GRU моделировать сложную динамику при их отсутствии.

Тестирование алгоритма проводилось на видеопоследовательностях с полётами мультироторного дрона в режиме реального времени на графическом процессоре NVIDIA GeForce 4060, что позволило оценить стабильность сопровождения объектов, точность предсказания координат и скоростей, а также поддерживать кадровую частоту видеопотока порядка 45 FPS.

Эксперименты включали сценарии плавного движения, резких манёвров, частичных и длительных пропусков детекции

Визуальный анализ показал, что:

- при наличии детекции алгоритм обеспечивает стабильное сопровождение без заметного дрожания;
- при кратковременных пропусках (до 10 кадров) GRU корректно предсказывает траекторию;
- при длительных пропусках (до 30 кадров) алгоритм сохраняет устойчивость за счёт постепенного снижения влияния нейросети.

Особенно важно, что гибридный подход предотвращает «улёт» ограничивающей рамки, характерный для линейного фильтра при длительном прогнозировании.

При обработке видеопотока с разрешением Full HD (1920×1080) была достигнута производительность:

- детекция (YOLO): ~35–45 FPS;
- GRU + Kalman: >200 FPS (не является узким местом);
- итоговая система: 40–45 FPS;

Таким образом, система работает в режиме, близком к реальному времени, с возможностью обработки видеопотока без существенных задержек.

Следует отметить, что вычислительная нагрузка в основном определяется детектором объектов, тогда как модуль сопровождения (Kalman + GRU) имеет незначительные затраты и может быть эффективно использован даже на менее производительных устройствах.

Для количественной оценки превосходства гибридного модуля трекинга над фильтром Калмана видеопоследовательности были обработаны дообученной моделью YOLO11s с последующим сохранением результатов детекции в формате CSV. Далее формировались фрагменты последовательностей, содержащие не менее 60 последовательных кадров с устойчивой детекцией объекта. Первые 30 кадров использовались для накопления истории наблюдений, необходимой для запуска гибридного алгоритма сопровождения, тогда как детекции на последующих 30 кадрах искусственно маскировались.

Таким образом, оба алгоритма выполняли прогноз положения объекта в условиях отсутствия детекции. Предсказания сравнивались с исходными результатами YOLO, принимаемыми в качестве истинных значений.

Метрики качества рассчитывались по евклидову расстоянию между центрами истинной и предсказанной ограничивающих рамок. Размеры рамок не учитывались, поскольку исследуемая система решает задачу прогнозирования движения объекта, а не восстановления его геометрических параметров. К тому же, ширина и высота приходят из дообученной YOLO11s.

Итоговое сравнение фильтра Калмана и предложенного гибридного метода представлено в таблице 2.

Таблица 2. Сравнение гибридного модуля сопровождения с фильтром Калмана

Метрика	Фильтр Калмана	Гибрид (Калман+GRU)
Mean error	0,094	0,074

RMSE	0,186	0,118
Std	0,161	0,093
Improvement, %	-	16,570

Для оценки качества сопровождения использовался набор метрик, характеризующих точность, устойчивость и стабильность предсказаний при пропадании детекции.

1. Средняя ошибка сопровождения (Mean error).

Метрика отражает среднее отклонение центра предсказанной рамки от истинного положения объекта в кадрах без детекции. Снижение ошибки составляет $\approx 21\%$, это означает, что использование GRU-предсказания движения позволяет значительно уменьшить накопление ошибки положения объекта при отсутствии измерений.

2. Среднеквадратичная ошибка (RMSE).

Метрика RMSE сильнее штрафует большие ошибки и характеризует наличие выбросов и срывов трекинга. Уменьшение RMSE составляет $\approx 36\%$. Полученный результат показывает, что гибридный алгоритм существенно снижает вероятность резкого ухода трека и лучше сохраняет траекторию объекта при длительном прогнозировании.

3. Стандартное отклонение ошибки (Std).

Метрика std характеризует стабильность сопровождения. Снижение дисперсии ошибки составляет $\approx 42\%$. Это свидетельствует о более плавной и устойчивой траектории сопровождения. Гибридный метод демонстрирует значительно меньшие колебания положения объекта относительно истинной траектории.

4. Интегральное улучшение (Improvement).

Показатель отражает относительное снижение средней ошибки сопровождения гибридного подхода относительно фильтра Калмана и составляет 16,57%. Таким образом, предложенный гибридный подход обеспечивает устойчивое повышение точности трекинга по сравнению с классическим фильтром Калмана.

Проведённый эксперимент показывает, что добавление рекуррентной нейросетевой модели движения к классическому фильтру Калмана позволяет:

- снизить среднюю ошибку сопровождения;
- уменьшить вероятность крупных ошибок предсказания;
- повысить устойчивость траектории;
- значительно улучшить качество трекинга при длительной потере детекции.

Полученные результаты подтверждают эффективность гибридного подхода для задач автоматического сопровождения воздушных объектов в видеопотоке и демонстрируют его преимущество над классическими методами фильтрации.

Заключение. В данной работе были рассмотрены классические методы сопровождения объекта на основе фильтра Калмана и выявлено, что они не всегда обеспечивают стабильное отслеживание при пропусках детекции или резких изменениях траектории дрона. Фильтр Калмана эффективно сглаживает траектории и предсказывает координаты при небольших пропусках наблюдений, однако его линейная модель движения ограничена в случаях сложных манёвров, ускорений и внезапных изменений направления. Для повышения точности и устойчивости трекинга был реализован рекуррентный нейросетевой модуль GRU, который корректирует предсказанные скорости объекта на основе истории наблюдений. Архитектура GRU состоит из одного слоя с 64 скрытыми нейронами и последующей линейной головой, предсказывающей приращения скорости по двум координатам. На вход модели подаются скользящие последовательности длиной 30 кадров, что позволяет учитывать динамику движения объекта и предсказывать его траекторию на горизонте до 30 кадров вперёд.

Для подготовки данных перед обучением GRU была использована дообученная модель YOLO11s, которая формировала датасет с мультироторными и самолётными типами дронов, а также птицами для проверки работы на сложных сценах. YOLO позволила выделять ограничивающую рамку объектов и рассчитывать скорости движения, которые затем сглаживались фильтром Калмана и использовались, как целевые приращения скорости для GRU. Такой подход позволил обеспечить высокое качество предсказаний даже при неполной детекции и минимизировать дрожание рамки.

Результаты тестирования алгоритма на видеопоследовательностях с реальными полётами БПЛА показали стабильное сопровождение объектов, точное предсказание координат и скоростей, а также возможность работы в реальном времени на графическом процессоре NVIDIA GeForce 4060. Средняя кадровая частота видеопотока достигла порядка 45 FPS, что обеспечивает достаточную скорость обработки для практического применения. Количественная оценка качества сопровождения выполнялась по евклидову расстоянию между центрами истинной и предсказанной ограничивающих рамок в условиях искусственно смоделированных пропусков детекции. По результатам эксперимента гибридный алгоритм показал снижение средней ошибки сопровождения на 16,57 % по сравнению с классическим фильтром Калмана. Среднеквадратичная ошибка (RMSE) уменьшилась примерно на 36 %, а стандартное отклонение ошибки – более чем на 40 %, что свидетельствует о более стабильной и предсказуемой траектории движения.

Таким образом, комбинированный подход, сочетающий фильтр Калмана и рекуррентную нейросеть GRU, позволяет реализовать гибридный алгоритм автоматического сопровождения БПЛА, который демонстрирует высокую точность, устойчивость к пропускам и возможность работы в реальном времени с кадровой частотой до 45 FPS. Это делает предложенное решение эффективным для практического применения в задачах трекинга объектов в динамичных условиях полёта.

Список источников

1. Nacar O., Abdelkader M., Ghouti L. et al. VECTOR: velocity-enhanced GRU neural network for real-time 3D UAV trajectory prediction. *Drones*, 2025, vol. 9, no. 1, DOI: 10.3390/drones9010008.
2. Yoon S., Jang D., Yoon H. et al. GRU-based deep learning framework for real-time, accurate, and scalable UAV trajectory prediction. *Drones*, 2025, vol. 9, no. 2, DOI:10.3390/drones9020142.
3. Jia C., Ma J., Kouw W.M. Multiple variational Kalman-GRU for ship trajectory prediction with uncertainty. *IEEE Transactions on Aerospace and Electronic Systems*, 2025, vol. 61, no. 2, pp. 3654-3667, DOI: 10.1109/TAES.2024.3491053.
4. Yusheng L., Xinmin T., Xuanming R. GRU-IKF scene taxiing trajectory prediction model based on attention mechanism. *Journal of Applied Artificial Intelligence*, 2024, vol. 2, no. 1, pp. 314–327, DOI:10.59782/aai.v1i2.311.
5. Li Z. et al. GAKP: GRU association and Kalman prediction for multiple object tracking. *arXiv preprint*, 2020, arXiv:2012.14314, DOI:10.48550/arXiv.2012.14314.
6. Li N., Miao Q. LSTM-Kalman filter-based multi-sensor signal fusion for UAV altitude prediction in non-Gaussian environments. *Traitement du Signal*, 2025, vol. 42, no. 2, pp. 1111-1122, DOI:10.18280/ts.420242.
7. Золкин А.Л. Методы идентификации и отслеживания объектов в системах дополненной реальности и компьютерного зрения: учебное пособие для вузов. – СПб.: Лань, 2025. – URL: <https://e.lanbook.com/book/505478> (дата обращения: 22.01.2026).
8. Клетте Р. Компьютерное зрение. Теория и алгоритмы: учебник. – М.: Лань, 2019. – URL: <https://e.lanbook.com/book/131691> (дата обращения: 25.02.2026).
9. Макшанов А.В. Технологии интеллектуального анализа данных: учебное пособие. – СПб.: Лань, 2022. – URL: <https://e.lanbook.com/book/206711> (дата обращения: 12.03.2026).
10. Калин О. Архитектуры глубокого обучения. Математический подход: руководство. – М.: Лань, 2025. – URL: <https://e.lanbook.com/book/514864> (дата обращения: 20.03.2026).
11. Паттерсон Д. Глубокое обучение с точки зрения практика. – М.: Лань, 2018. – URL: <https://e.lanbook.com/book/116122> (дата обращения: 22.03.2026).

12. Парамонов В.А. Методы и модели исследования сложных систем и обработки больших данных / В.А. Парамонов, Н.Е. Смагин, А.Д. Косых, А.Д. Хомоненко. – СПб.: Лань, 2024. – URL: <https://e.lanbook.com/book/425024> (дата обращения: 22.03.2026).
13. Ненашев В.А. Компьютерное зрение. Анализ, обработка и моделирование: учебное пособие. – СПб.: Лань, 2022. – URL: <https://e.lanbook.com/book/341057> (дата обращения: 22.03.2026).

Фелагерева Евгения Юрьевна. Бакалавр в области радиофизики. Основные направления исследования автора: машинное обучение, нейронные сети, радиофизика. AuthorID: 1337017, SPIN: 1499-0400, ORCID: 0009-0001-2406-1225. efelagereva@gmail.com. 350040, г. Краснодар, ул. Ставропольская, 149.

Рудоман Нелли Радиковна. Старший преподаватель кафедры оптоэлектроники ФГБОУ ВО "Кубанский государственный университет". Основные направления исследования автора: физика твердого тела, компьютерное моделирование физических явлений, машинное обучение, искусственный интеллект, квантовые компьютеры, квантовая криптография. AuthorID: 39466, SPIN: 5844-2932, ORCID: 0009-0003-4308-235X. rudnel@ftf.kubsu.ru. 350040, г. Краснодар, ул. Ставропольская, 149.

Новиков Александр Дмитриевич. Аспирант в области оптики. Основные направления исследования: машинное обучение, нейронные сети, радиофизика. ORCID: 0009-0003-1596-8719. nad1100@yandex.ru. 350040, г. Краснодар, ул. Ставропольская, 149.

UDC 004.032.26

DOI:10.25729/ESI.2026.43.3.014

Development of a hybrid module for tracking air objects based on GRU and the Kalman filter

Eugenia Yu. Felagereva, Nelly R. Rudoman, Alexander D. Novikov

Kuban state university, Russia, Krasnodar, efelagereva@gmail.com

Abstract. This paper considers the problem of robust tracking of aerial objects in a video stream, with an emphasis on predicting their trajectories in the presence of measurement noise and short-term detection losses. The aim of the study is to develop a hybrid tracking method combining a classical Kalman filter and a GRU-type recurrent neural network to improve the accuracy and robustness of object motion prediction. The proposed approach is based on the use of a Kalman filter to estimate the current state of an object, specified by the coordinate vector of the YOLO bounding box and the velocities $[c_x, c_y, w, h, v_x, v_y]$, and a GRU-type neural network model, which is trained to predict the change in velocity $\Delta v_x, \Delta v_y$ based on a sequence of previous states. Unlike traditional methods, where the motion model is specified a priori, in this paper the motion dynamics are partially extracted from the data. GRU predictions are integrated into the Kalman filter model through an adaptive smoothing mechanism with a coefficient α , which allows for flexible adjustment of the contribution of the neural network prediction to the final state estimate. The main content of the paper includes a description of the hybrid model architecture, a methodology for preparing a training dataset, and an algorithm for integrating neural network predictions into the filtering process. Particular attention is paid to the problem of long-term forecasting—up to 30 frames ahead—which is critical for active tracking and control systems. The scientific novelty of the paper lies in the proposed method for combining GRU and a Kalman filter through adaptive smoothing of the control action, which allows for the consideration of nonlinear and nonstationary characteristics of object motion without sacrificing the interpretability of the classical model. Unlike existing solutions, the proposed method ensures more stable operation under noisy measurements and missed detections. Experiments demonstrated that the proposed approach improves trajectory prediction accuracy and reduces error accumulation during multi-step forecasting compared to the classical Kalman filter and basic neural network models. As a result of the experiments, it was shown that the proposed approach increases the accuracy of trajectory prediction by 16.57% and reduces the accumulation of errors in multi-step forecasting by 21% compared to the classical Kalman filter and basic neural network models. The developed method can be applied to unmanned aerial vehicle detection and tracking, video surveillance systems, and autonomous navigation.

Keywords: Kalman filter, GRU, UAV, positioning

References

1. Nacar O., Abdelkader M., Ghouti L. et al. VECTOR: velocity-enhanced GRU neural network for real-time 3D UAV trajectory prediction. *Drones*, 2025, vol. 9, no. 1, DOI: 10.3390/drones9010008.
2. Yoon S., Jang D., Yoon H. et al. GRU-based deep learning framework for real-time, accurate, and scalable UAV trajectory prediction. *Drones*, 2025, vol. 9, no. 2, DOI:10.3390/drones9020142.
3. Jia C., Ma J., Kouw W.M. Multiple variational Kalman-GRU for ship trajectory prediction with uncertainty. *IEEE Transactions on Aerospace and Electronic Systems*, 2025, vol. 61, no. 2, pp. 3654-3667, DOI: 10.1109/TAES.2024.3491053.
4. Yusheng L., Xinmin T., Xuanming R. GRU-IKF scene taxiing trajectory prediction model based on attention mechanism. *Journal of Applied Artificial Intelligence*, 2024, vol. 2, no. 1, pp. 314–327, DOI:10.59782/aai.v1i2.311.
5. Li Z. et al. GAKP: GRU association and Kalman prediction for multiple object tracking. *arXiv preprint*, 2020, arXiv:2012.14314, DOI:10.48550/arXiv.2012.14314.
6. Li N., Miao Q. LSTM-Kalman filter-based multi-sensor signal fusion for UAV altitude prediction in non-Gaussian environments. *Traitement du Signal*, 2025, vol. 42, no. 2, pp. 1111-1122, DOI:10.18280/ts.420242.
7. Zolkin A.L. *Metody identifikatsii i otslezhivaniya ob"yektov v sistemakh dopolnennoy real'nosti i komp'yuternogo zreniya: uchebnoye posobiye dlya vuzov* [Methods of identification and tracking of objects in augmented reality and computer vision systems: a textbook for universities]. Saint Petersburg, Lan Publ., 2025. Available at: <https://e.lanbook.com/book/505478> (accessed: 01/22/2026).
8. Klette R. *Komp'yuternoye zreniye. Teoriya i algoritmy: uchebnik* [Computer vision. Theory and algorithms: textbook]. Moscow, Lan Publ., 2019. Available at: <https://e.lanbook.com/book/131691> (accessed: 02.25.2026).
9. Makshanov A.V. *Tekhnologii intellektual'nogo analiza dannykh: uchebnoye posobiye* [Data mining technologies: textbook]. Saint Petersburg, Lan Publ., 2022. Available at: <https://e.lanbook.com/book/206711> (accessed: 03/22/2026).
10. Kalin O. *Arkhitektury glubokogo obucheniya. Matematicheskiy podkhod: rukovodstvo* [Deep learning architectures. A mathematical approach: guide]. Moscow, Lan Publ., 2025. Available at: <https://e.lanbook.com/book/514864> (accessed: 03/22/2026).
11. Patterson D. *Glubokoye obucheniye s tochki zreniya praktika* [Deep learning from a practitioner's perspective]. Moscow, Lan Publ., 2018. Available at: <https://e.lanbook.com/book/116122> (accessed: 03/22/2026).
12. Paramonov V.A., Smagin N.E., Kosykh A.D., Khomonenko A.D. *Metody i modeli issledovaniya slozhnykh sistem i obrabotki bol'shikh dannykh* [Methods and models for research of complex systems and big data processing]. Saint Petersburg, Lan Publ., 2024. Available at: <https://e.lanbook.com/book/425024> (accessed: 03/22/2026).
13. Nenashev V.A. *Komp'yuternoye zreniye. Analiz, obrabotka i modelirovaniye: uchebnoye posobiye* [Computer vision. Analysis, processing and modeling: textbook]. Saint Petersburg, Lan Publ., 2022. Available at: <https://e.lanbook.com/book/341057> (accessed: 03/22/2026).

Felagereva Evgeniya Yurievna. Bachelor's degree in Radiophysics. The author's main research areas are machine learning, neural networks, and radiophysics. AuthorID: 1337017, SPIN: 1499-0400, ORCID: 0009-0001-2406-1225. efelagereva@gmail.com. 350040, Krasnodar, st. Stavropol, 149.

Rudoman Nelly Radikovna. Senior Lecturer at the Department of Optoelectronics, Kuban State University. The author's main research areas are solid state physics, computer modeling of physical phenomena, machine learning, artificial intelligence, quantum computers, and quantum cryptography. AuthorID: 39466, SPIN: 5844-2932, ORCID: 0009-0003-4308-235X. rudnel@ftf.kubsu.ru. 350040, Krasnodar, st. Stavropol, 149.

Novikov Alexander Dmitrievich. Postgraduate student in optics. The author's main research areas are machine learning, neural networks, and radiophysics. ORCID: 0009-0003-1596-8719. nad1100@yandex.ru. 350040, Krasnodar, st. Stavropol, 149.

Статья поступила в редакцию 24.01.2026; одобрена после рецензирования 16.04.2026; принята к публикации 01.08.2026.

The article was submitted 01/24/2026; approved after reviewing 04/16/2026; accepted for publication 08/01/2026.

Миварная система обработки информации и принятия решений для группового управления гетерогенными складскими роботами

Гун Шэншо

Московский государственный технический университет им. Н. Э. Баумана,
Россия, Москва, hiteyeb@163.com

Аннотация. В статье рассматривается проблема группового управления функционально гетерогенными складскими роботами – роботами-погрузчиками (РП), роботами-транспортёрщиками (РТ) и роботами-разгрузчиками (РР). Целью исследования является разработка миварной системы обработки информации и принятия решений, обеспечивающей динамическую классификацию транспортных задач, выбор исполнителей по критерию совместимости и адаптивное управление энергетическими ресурсами парка при сохранении линейной вычислительной сложности логического вывода и полной объяснимости решений. Предложена классификация транспортных задач на три типа – индивидуальную, кооперативную и коллективную транспортировку – на основе системы булевых индикаторов, динамического приоритета и проверки габаритно-весовых ограничений. Для коллективной транспортировки разработан механизм пространственной кластеризации грузов с двухрежимной маршрутной оптимизацией (полный перебор и комбинация ближайшего соседа с 2-opt). Для кооперативной транспортировки крупногабаритных грузов введён аукционный механизм формирования команд РТ, минимизирующий составную целевую функцию с учётом избыточности грузоподъёмности, гетерогенности скоростей и путевых затрат. Для каждого типа роботов сформулированы уравнения состояния, работоспособности и энергетической достаточности. Математическая модель преобразована в миварную базу знаний из 302 правил «ЕСЛИ–ТО» и верифицирована на платформе КЭСМИ. Новизна работы состоит в одновременном учёте функциональной гетерогенности парка, трёхкатегорийной классификации задач и адаптивного энергетического управления в рамках единой миварной базы знаний – аспектов, ранее рассматривавшихся изолированно. Экспериментальная проверка на имитационной системе (склад 1000×1000 м, 140 роботов, 300 задач) показала коэффициент выполнения задач 98,9–99,9 %, что на 14–18 п.п. превышает базовые алгоритмы. Метод сохраняет работоспособность при отказе до 30 % роботов и при увеличении доли кооперативных задач до 30 %.

Ключевые слова: мивар, миварная система принятия решений, миварная база знаний, логический искусственный интеллект, складской робот, групповое управление, распределение задач

Цитирование: Гун Ш. Миварная система обработки информации и принятия решений для группового управления гетерогенными складскими роботами / Ш. Гун // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 178-196. – DOI:10.25729/ESI.2026.43.3.015.

Введение. Непрерывный рост объёмов электронной коммерции, увеличение номенклатуры хранимых единиц и ужесточение требований к срокам доставки превращают интеллектуальный склад [1] в один из ключевых узлов современной логистики [2]. Традиционные автоматизированные стеллажные системы (AS/RS) и ранние поколения автоматически управляемых транспортных средств (AGV) с фиксированными маршрутами оказываются недостаточно гибкими при обработке высокодинамичных и разнородных потоков заказов [3]. На смену им приходят группы автономных мобильных роботов (AMR), обеспечивающие масштабируемость, гибкость и отказоустойчивость складских операций [4].

Организация совместной работы такой группы требует модели обработки информации и принятия решений [5], способной в реальном времени учитывать положение роботов, уровень заряда их аккумуляторов, очереди задач, физические характеристики грузов и состояние погрузочных платформ, а на основании этих данных формировать распределение ресурсов, близкое к оптимальному [6]. В российской литературе вопросы планирования действий и распределения заданий в мультиагентных робототехнических системах рассматривались в работе И. М. Макарова [7], проблема группового управления роботами с учётом ограничения

по запасу хода – в работе В. С. Лазарева [8], а управление группой наземных роботов с перераспределением энергетических ресурсов – в работе К. Д. Крестовникова [9].

Существующие зарубежные подходы к проблеме распределения принято подразделять на две парадигмы. Централизованные оптимизационные модели на базе смешанного целочисленного программирования, эвристик переменных окрестностей и аукционных схем [10] теоретически способны находить глобальный оптимум, однако их вычислительная сложность быстро возрастает с увеличением числа роботов и задач. Децентрализованные реактивные модели и методы обучения с подкреплением [11] обеспечивают высокое быстроедействие, однако лишены глобального обзора и склонны к застреванию в локальных оптимумах, особенно при выполнении задач, требующих точной кооперации нескольких роботов – например, совместной транспортировки крупногабаритных грузов [12].

Одним из перспективных направлений, позволяющих преодолеть указанное противоречие, является использование логического искусственного интеллекта и, в первую очередь, миварного подхода [13]. Исторически развитие экспертных систем прошло путь от первых продукционных архитектур [14] до промышленных приложений в горных работах, медицине и других областях [15]. Миварный подход, предложенный О. О. Варламовым в 2002 году [16], основан на трехмерном информационном пространстве «вещь–свойство–отношение» и обеспечивает линейную вычислительную сложность логического вывода на правилах «ЕСЛИ–ТО» как для небольшого, так и для большого числа правил (например, при числе правил, превышающем три миллиона). Современная инструментальная среда КЭСМИ обеспечивает разработку, верификацию и эксплуатацию миварных баз знаний промышленного масштаба [17].

Применение миварных технологий в робототехнике активно развивается [18]. В работе Д. В. Аладина впервые предложена миварная система для планирования действий роботов [19], в работах А. А. Коценко и соавторов исследуется методика построения миварных систем принятия решений для планирования маршрутов робототехнических комплексов в трёхмерном логическом пространстве [20], применение миварных экспертных систем (МЭС) для автоматизированного поиска нескольких траекторий робота и описание миварных сетей в формате двудольных и трёхдольных ориентированных графов [21]. В работе [22] обоснована возможность создания автономных групп роботов на основе миварной системы принятия решений; в [23] предложена миварная экспертная система для контроля соблюдения правил дорожного движения. Применительно к производственным системам вопросы распределения ресурсов средствами логического искусственного интеллекта рассмотрены в [24], а в [25] представлен алгоритм динамического планирования траектории робота на основе семантического обнаружения объектов. Примеры применения МЭС в распознавании образов и смежных задачах приведены в [26, 27].

Непосредственно к складской проблематике миварный подход применялся в предшествующих работах автора: в [28] была сформулирована миварная система принятия решений для распределения и перевозки грузов командой складских роботов; в [29] рассмотрено оптимизационное распределение грузов для групп складских роботов; в [30] предложена модель принятия решений для многозадачных складских роботов на основе четырёхуровневой архитектуры совместной оптимизации. Вместе с тем проведённый анализ показывает, что существующие решения, как правило, ориентированы на однородные группы роботов и не учитывают одновременно три ключевых аспекта: функциональную гетерогенность состава группы (наличие роботов-погрузчиков, роботов-транспортировщиков и роботов-разгрузчиков), трёхкатегорийную классификацию задач (индивидуальная, кооперативная и коллективная транспортировка) и энергетическое состояние роботов при

динамическом распределении. Отсутствует и единая формализация, позволяющая перевести эти аспекты в миварную базу знаний, пригодную для эксплуатации в КЭСМИ.

Цель настоящей работы – разработка миварной системы обработки информации и принятия решений для группового управления гетерогенными складскими роботами, обеспечивающей динамическую классификацию транспортных задач, выбор исполнителей по критерию совместимости и адаптивное управление энергетическими ресурсами. Для достижения поставленной цели решаются следующие задачи: построение математической модели классификации задач и выбора роботов трёх типов; преобразование модели в миварную базу знаний и тестирование на платформе КЭСМИ; разработка имитационной системы и экспериментальная проверка эффективности, адаптивности, надёжности и расширяемости предложенной системы.

Материалы и методы. В складской среде функционируют три типа гетерогенных роботов: робот-погрузчик (РП, RP), робот-транспортировщик (РТ, RT) и робот-разгрузчик (РР, RR). Множество грузов, поступающих в систему, обозначим G :

$$G = \{g_1, g_2, \dots, g_j, \dots, g_J\}$$

Каждый груз g_j характеризуется следующими параметрами:

$$g_j = \{P_{pickup_j}, P_{destination_j}, m_j, v_j: [l_j, w_j, h_j], \alpha_j, t_j, storey_j\}$$

где P_{pickup_j} и $P_{destination_j}$ – координаты места погрузки и назначения; m_j – масса; $v_j: [l_j, w_j, h_j]$ – габаритные размеры; α_j – степень срочности; t_j – момент поступления; $storey_j$ – номер целевого стеллажа.

Классификация транспортных задач. Для каждого груза проверяются ограничения по габаритам $l_j < l_{max} \cap w_j < w_{max} \cap h_j < h_{max}$ и массе $m_j < M_{max}$. При их выполнении груз может быть доставлен одним РТ. Динамический приоритет $Priority_j$ задаётся взвешенной свёрткой степени срочности и времени ожидания:

$$Priority_j = 0.6\alpha_j + 0.4(1 - e^{-\lambda\Delta t_j}) \quad (1)$$

где $\alpha_j \in [0,1]$ – степень срочности груза; $\lambda = 0.02 \text{ min}^{-1}$ – коэффициент затухания срочности; $\Delta t_j = t_{current} - t_j$ – время ожидания груза; $(l_{max}, w_{max}, h_{max})$ – размеры грузового ящика RT_i .

Динамические параметры задаются следующим образом:

$$C_V = 0.6(1 + 0.1\gamma) \quad (2)$$

$$C_M = 0.7(1 + 0.1\gamma) \quad (3)$$

где γ характеризует текущую загруженность системы.

Условие отнесения груза к крупногабаритному:

$$B_j = \begin{cases} 1 & \text{if } \frac{v_j}{V_{max}} > C_V \cup \frac{m_j}{M_{max}} > C_M \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

где v_j и m_j – объем и масса груза; V_{max} и M_{max} – максимальный объем грузового ящика и максимальная грузоподъемность РТ; C_V и C_M – динамические постоянные объема грузового ящика и грузоподъемности РТ.

Решение об индивидуальной транспортировке:

$$X_j = \begin{cases} 1 & \text{if } B_j = 1 \cup Priority_j > 0.6 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Необходимость кооперативной транспортировки определяется как

$$isCoop_j = \begin{cases} 1 & \text{if } l_j \geq l_{max} \cup w_j \geq w_{max} \cup m_j \geq m_{max} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

Если $isCoop_j = 1$, требуется рассчитать необходимое количество РТ для кооперативной транспортировки:

$$N_j = \max\left(\left\lceil \frac{l_j}{l_{\max}} \right\rceil, \left\lceil \frac{w_j}{w_{\max}} \right\rceil, \left\lceil \frac{m_j}{m_{\max}} \right\rceil, 2\right) \quad (7)$$

Если $isCoop_j = 0$ и $X_j = 0$, груз относится к коллективной транспортировке и включается в множество $G_{res} = \{g_j | X_j = 0\}$. Таким образом, каждый поступающий груз относится ровно к одной из трёх категорий: индивидуальной ($X_j = 1, isCoop_j = 0$), кооперативной ($isCoop_j = 1$) или коллективной ($X_j = 0, isCoop_j = 0$). Далее для каждой категории описываются соответствующие процедуры обработки: группировка и маршрутизация коллективных грузов, выбор исполнителей для индивидуальных и коллективных задач, а также формирование команд для кооперативной транспортировки.

Группировка грузов коллективной транспортировки. На данном уровне проводится группировка грузов множества G_{res} с учётом ограничений по массе, объёму и расстоянию. Если груз можно добавить в текущую группу G_i , он включается в неё; иначе формируется новая группа G_{i+1} . Каждая группа содержит не более 6 грузов (ограничение $n_{G_i} \leq 6$ обусловлено, с одной стороны, физическим пределом загрузки одного робота (при средней массе груза 20 кг и $M_{max} = 100$ кг), а с другой – необходимостью сохранения полиномиальной сложности маршрутной оптимизации методом 2-opt.):

$$\begin{cases} \text{if } \sum_{g_k \in G_i} m_k + m_j \leq M_{max} \cap \sum_{g_k \in G_i} V_k + V_j \leq V_{max} \cap \|S_i - S_j\| \leq P_0 \\ \quad \rightarrow g_j \in G_i \\ \text{else} \quad \text{create } G_{i+1} \end{cases} \quad (8)$$

Пороговое значение расстояния задаётся динамически:

$$P_0 = \begin{cases} 70m & \text{if } \sum m_{G_i} \leq 0.7M_{max} \cap \sum V_{G_i} \leq 0.7V_{max} \\ 35m & \text{otherwise} \end{cases} \quad (9)$$

Для каждой полученной группы G_i определяется последовательность посещения π , минимизирующая суммарное расстояние транспортировки:

$$\min_{\pi} \left(d(P_{pickup}, S_{\pi(0)}) + \sum_{k=1}^{n_{G_i}-1} d(S_{\pi(k)}, S_{\pi(k+1)}) \right) \quad (10)$$

где $d(a, b)$ – манхэттенское расстояние между точками a и b ; π – текущий порядок транспортировки грузов; P_{pickup} – место загрузки первого груза группы; S_k – место хранения груза; n_{G_i} – количество грузов в текущей группе. Для $n_{G_i} \leq 2$ используется полный перебор; для $3 \leq n_{G_i} \leq 6$ – жадный алгоритм ближайшего соседа для построения начального маршрута:

$$\begin{cases} \text{Current point } c \leftarrow P_{pickup} \\ \text{iteration } t = 1 \rightarrow n \\ k^* = \arg \min d(c, S_k) \quad k \in U = \{0, 1, \dots, n-1\} \\ \pi_{init} \leftarrow \pi_{init} \oplus k^*, U \leftarrow U \setminus \{k^*\}, c \leftarrow S_{k^*} \end{cases} \quad (11)$$

где символ $\oplus$ обозначает операцию конкатенации (присоединения элемента или подпоследовательности к упорядоченной последовательности маршрута).

Затем применяется алгоритм 2-opt для локальной оптимизации:

$$\begin{cases} \forall (i, j) \in \{1 \leq i < j \leq n-1\}: \\ \pi_{new} = \pi[0:i] \oplus reverse(\pi[i:j+1]) \oplus \pi[j+1:] \\ \text{if } D(\pi_{new}) < D(\pi) - 10^{-5}: \\ \pi^* \leftarrow \pi_{new} \end{cases} \quad (12)$$

Порядок обработки задач. Индивидуальные и кооперативные задачи обрабатываются в порядке поступления. Для коллективных задач проводится консолидация по 30 грузов в

партию ($n_{G_{res}} = 30$, выбран как компромисс между эффективностью пространственной кластеризации и допустимым временем ожидания грузов в очереди) и группировка, после чего для каждой группы рассчитывается агрегированный показатель срочности:

$$W_{G_i} = \sum_j^{n_{G_i}} \left(\alpha_1 \frac{\Delta t_j}{60} + \beta_1 Priority_j \right) \quad (13)$$

где α_1, β_1 – веса по времени ожидания и приоритету груза, $\alpha_1 = 0.6, \beta_1 = 0.4$.

Группы обслуживаются в порядке убывания W_{G_i} .

Уравнения работа-транспортника (РТ). Уравнение состояния RT_i :

$$s_i(t) = \begin{cases} 0, & \text{if } \exists j x_{ij}(t) = 1 \\ 1, & \text{otherwise} \end{cases} \quad (14)$$

при $\forall j \in \{1, 2, \dots, J\}, \forall i \in \{1, 2, \dots, I\}$. Когда $s_i(t) = 0$, РТ занят; при $s_i(t) = 1$ – свободен.

Уравнение работоспособности:

$$f_i(t) = \begin{cases} 0, & \text{if } RT_i \text{ needs repair} \\ 1, & \text{otherwise} \end{cases} \quad (15)$$

Динамический порог оставшегося заряда РТ:

$$\tau_{RT} = \frac{15}{e^V} \quad (16)$$

где 15 – базовый пороговый уровень заряда (в процентах) при единичной загрузке системы.

Если $E_i(t) \leq \tau_{RT}$, робот направляется на зарядную станцию. Целевой уровень заряда задаётся адаптивно:

$$E_{RT_target} = 80 + \tau_{RT} \quad (17)$$

где 80 – минимальный целевой уровень заряда (%).

При высокой нагрузке (большие γ) порог снижается, что позволяет роботу быстрее вернуться к работе; при низкой нагрузке порог увеличивается, что продлевает ресурс аккумулятора. Скорость зарядки задаётся сегментированной линейной функцией:

$$\frac{dE}{dt} = \begin{cases} k_{fast1}, & \text{if } E_i(t) < 0.8E_{RT_target} \\ k_{slow1}, & \text{otherwise} \end{cases} \quad (18)$$

где $k_{fast1} = 0.15 \text{ min}^{-1}$ и $k_{slow1} = 0.08 \text{ min}^{-1}$.

Время завершения зарядки:

$$T_{RT_charge} = \frac{0.8E_{RT_target} - E_i(0)}{k_{fast1}} + \frac{0.2E_{RT_target}}{k_{slow1}} \quad (19)$$

Способность RT_i к перевозке груза j :

$$c_{ij}(t) = \begin{cases} 1, & \text{if } E_i(t) > \frac{D_{ij}}{200} + \tau_{RT} \\ 0, & \text{if } RT_i \text{ needs to be charged} \end{cases} \quad (20)$$

где общее манхэттенское расстояние для перевозки груза определяется как:

$$D_{ij} = d_{ij} + d_j + d_{j0} \quad (21)$$

с составляющими $d_{ij} = |x_i - x_{sj}| + |y_i - y_{sj}|$, $d_j = |x_{ej} - x_{sj}| + |y_{ej} - y_{sj}|$, $d_{j0} = |x_0 - x_{ej}| + |y_0 - y_{ej}|$.

Функция динамической совместимости RT_i и груза j :

$$Score_{ij}(t) = s_i(t)f_i(t)c_{ij}(t) \left(\frac{\alpha_2}{1 + \sqrt{d_{ij}}} + \beta_2 \frac{E_i(t)}{100} \right) \quad (22)$$

где α_2, β_2 – веса по расстоянию до груза и заряду робота, $\alpha_2 = 0.8, \beta_2 = 0.2$.

Переменная решения:

$$x_{ij}(t) = \begin{cases} 1, & \text{if } i^*, j^* = \operatorname{argmax}_j Score_{ij}(t) \\ 0, & \text{otherwise} \end{cases} \quad (23)$$

Если в момент t у RT_i максимальная совместимость к грузу j , то RT_i назначается на этот груз и $x_{ij}(t) = 1$.

Уравнения робота-погрузчика (РП) и робота-разгрузчика (РР). Состояние и работоспособность РП описываются аналогично уравнениям (14) и (15). Функция динамической совместимости RP_i и груза j :

$$Score_{ij}(t) = s_i(t)f_i(t)\frac{1}{1 + \sqrt{d_{ij}}} \quad (24)$$

РП назначается на груз по аналогии с (23). Состояние, работоспособность, динамический порог заряда $\tau_{RR} = \frac{20}{eV}$, целевой уровень $E_{RR_target} = 80 + \tau_{RR}$ и способность c_{pj} робота-разгрузчика определяются аналогично уравнениям (14)–(20) с учётом того, что РР выполняет только операции разгрузки и складирования.

Функция динамической совместимости RR_p и груза j :

$$Score_{pj}(t) = s_p(t)f_p(t)c_{pj}(t)\left(\frac{\alpha_2}{1 + \sqrt{d_{pj}}} + \beta_2 \frac{E_p(t)}{100}\right) \quad (25)$$

Совокупность переменных команд описывается булевыми переменными.

Выбор команд РТ для кооперативной транспортировки. Если для индивидуальной и коллективной транспортировки выбор РТ определяется функцией динамической совместимости $Score_{ij}$ (формула 22), то для кооперативной транспортировки, требующей согласованного действия нескольких роботов, применяется механизм комбинаторного аукциона, участниками которого выступают команды-кандидаты $C_r = \{RT_1, \dots, RT_i, \dots, RT_n\}$, состоящие из нескольких гетерогенных РТ. Кандидатные команды формируются следующим образом: сначала исключаются все роботы, не удовлетворяющие базовым условиям (низкий заряд, занятость); далее выбираются $K_j = n_{G_i} + 2$ ближайших к точке погрузки РТ, образующих группу $Group_j$; по $Group_j$ генерируются все возможные комбинации размера n_{G_i} . Общая стоимость выполнения задачи каждой командой рассчитывается как:

$$TotalCost(C_r) = \alpha_3 \Phi_{team}(C_r) + \beta_3 \Phi_{path}(C_r) \quad (26)$$

где α_3, β_3 – веса стоимости команды и зависимой от пути стоимости, $\alpha_3 = 0.4, \beta_3 = 0.6$.

Стоимость команды количественно оценивает её внутренние свойства:

$$\phi_{team}(C_r) = \omega_1 Redundancy + \omega_2 Heterogeneity \quad (27)$$

где ω_1, ω_2 – веса избыточности и гетерогенности, $\omega_1 = \omega_2 = 0.5$.

Redundancy штрафует неэффективные ситуации, когда совокупная грузоподъёмность команды существенно превышает требуемую:

$$Redundancy = 10 \frac{\sum \{RT_i \in C_r\} (Cap_i - m_j)}{m_j} \quad (28)$$

где Cap_i – грузоподъёмность RT_i ; 10 – коэффициент масштабирования, обеспечивающий сопоставимость вкладов *Redundancy* и *Heterogeneity* в составную функцию стоимости.

Heterogeneity штрафует различия в скорости роботов команды, усложняющие синхронное управление:

$$Heterogeneity = 10 \sigma_{speed}^2(C_r) \quad (29)$$

Путевые затраты:

$$\Phi_{path}(C_r, P_c) = \frac{1}{n_{C_r}} \sum \left(\{RT_i \in C_r\} d(Loc_{RT_i}, Loc_{p_j}) \right) + d(Loc_{p_j}, Loc_{d_j}) \quad (30)$$

Оптимизация проводится при двух ограничениях. Ограничение по совокупной грузоподъёмности:

$$\sum \{RT_i \in C_r\} Cap_i > Weight_j \quad (31)$$

Ограничение по энергии каждого робота команды:

$$\forall RT_i \in C_r: E_i \geq \gamma_E E(d_{path}) \quad (32)$$

Победителем аукциона становится команда C_r^* с минимальным значением $TotalCost(C_r)$. Сценарий назначения отличается от индивидуального тем, что сначала все РТ команды должны собраться у места погрузки ($s_{all_RT_ready} = 1$), после чего по команде системы начинается совместная загрузка.

Миварная база знаний. Описанная математическая модель была преобразована в миварную базу знаний (МБЗ), содержащую 302 правила вида «ЕСЛИ–ТО». Миварная база знаний организована в виде двудольной миварной сети [7-9], в которой вершины-переменные описывают состояния роботов, грузов и погрузочных платформ, а вершины-правила реализуют вычислительные и логические процедуры. Такая структура обеспечивает линейную вычислительную сложность логического вывода и позволяет автоматически перестраивать траекторию принятия решений при изменении исходных данных, что принципиально отличает предложенный метод от статических продукционных систем. Схема МБЗ приведена на рис. 1.

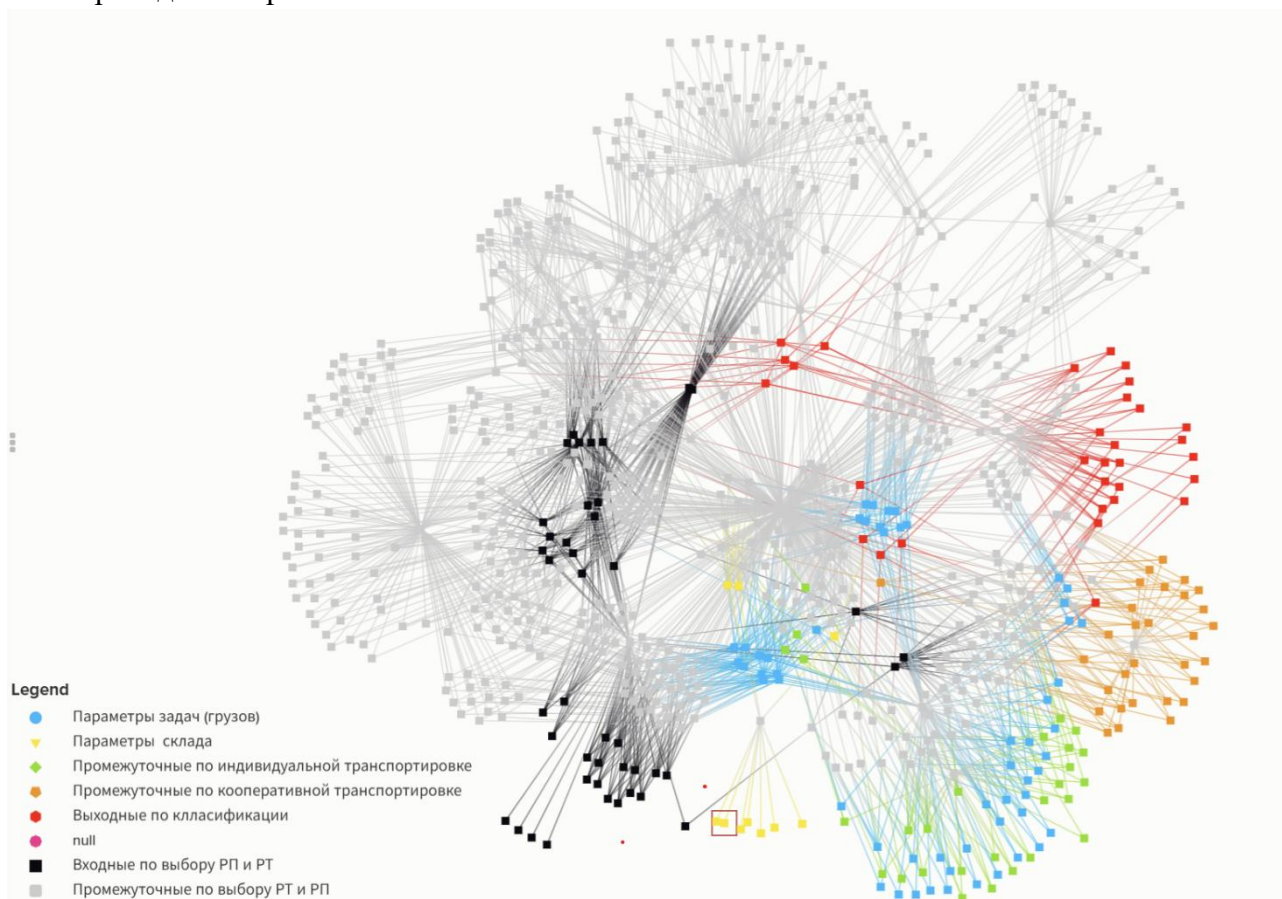


Рис. 1. Схема миварной базы знаний системы

Для наглядности сформированных с помощью миварного подхода правил приведены некоторые правила «ЕСЛИ–ТО» в таблице 1, используемые при групповом управлении гетерогенными складскими роботами.

Таблица 1. Фрагмент миварной базы знаний

Правило	ЕСЛИ	ТО
...	...	...
176	$l^*, j^* = \operatorname{argmax}_j \operatorname{Score}_{lj}(t)$ для $\forall j \in \{1, 2, \dots, J\}, \forall l \in \{1, 2, \dots, L\}$	$x_{l^*j^*}(t) = 1$, РП l^* назначается на груз j^* в момент t
177	$l^*, j^* \neq \operatorname{argmax}_j \operatorname{Score}_{lj}(t)$	$x_{l^*j^*}(t) = 0$, РП l^* не назначается на груз

		j^* в момент t
...	...	...

Результаты и обсуждение. На первом этапе была проведена верификация части разработанной МБЗ (для обработки индивидуальной задачи) на платформе КЭСМИ. Для проверки корректности логического вывода тестировали несколько сценариев. Результаты подтвердили полноту и непротиворечивость разработанной МБЗ и возможность её использования в качестве ядра системы принятия решений. Результат одного сценария (задача индивидуальная, нужен 1 РТ) приведен на рис. 2.

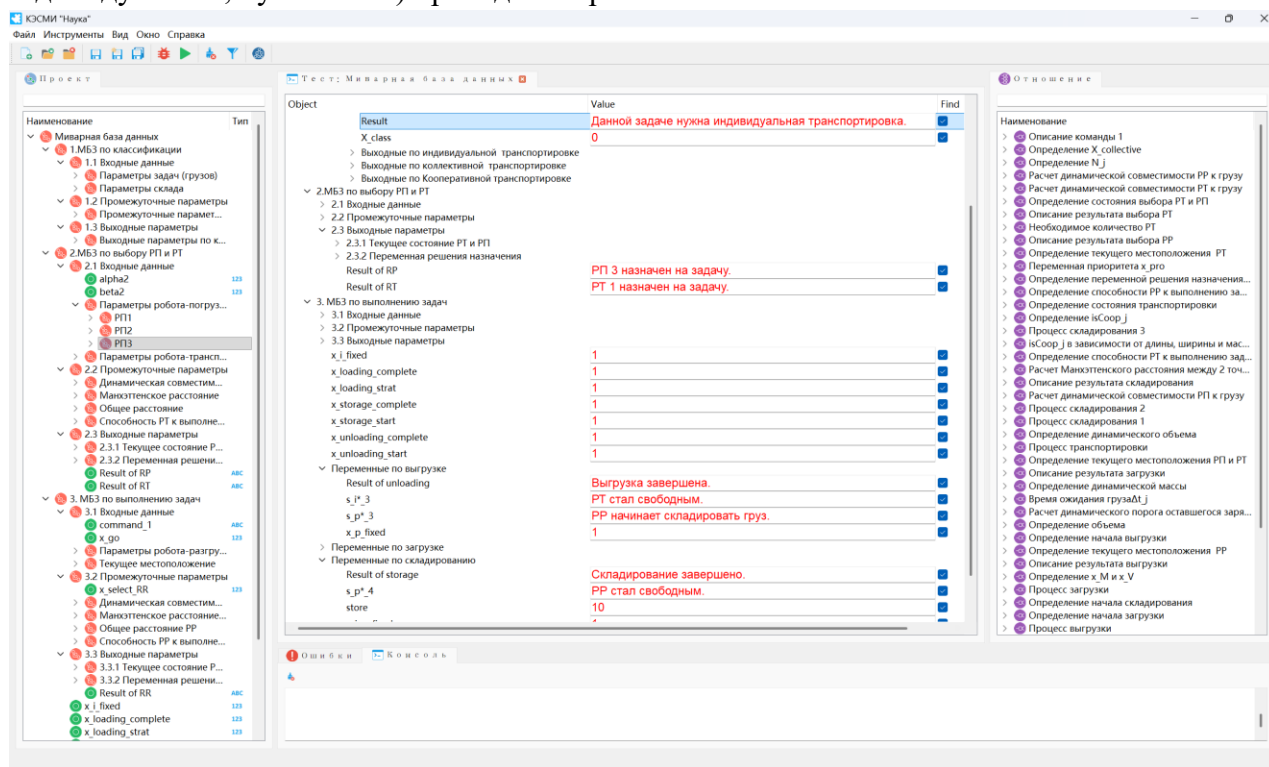


Рис. 2. Результат тестирования на КЭСМИ

Имитационная система. Для полноценной проверки предложенной системы разработана имитационная система на языке Python, воспроизводящая работу склада квадратной формы размером 1000×1000 м. В имитационной системе функционируют три типа роботов. Робот-погрузчик (РП) имеет грузоподъемность 350 кг, допустимые размеры груза 3×3×3 м и скорость 1 м/с; одна операция погрузки занимает 10 с, рабочая зона РП строго ограничена назначенной платформой. Робот-транспортировщик (РТ) – аккумуляторный робот с номинальной грузоподъемностью 100 кг, габаритами груза до 1×1×1 м, скоростью 2 м/с и расходом заряда 0,005 %/м; реализована возможность совместной транспортировки несколькими РТ. Робот-разгрузчик (РР) – аккумуляторный робот с грузоподъемностью 350 кг, скоростью 1 м/с и расходом 0,01 %/м; каждая операция разгрузки и последующего складирования занимает 10 с. При снижении заряда РТ или РР ниже динамических порогов τ_{RT} и τ_{RR} робот прерывает текущую задачу и направляется на ближайшую зарядную станцию. Все события регистрируются в системном журнале.

Графический интерфейс имитационной системы состоит из четырёх функциональных областей. Область визуализации склада отображает зоны стеллажей, магистральные дороги, зарядные станции и погрузочные платформы; цвет рамки символа робота указывает на его состояние (свободен / занят / на зарядке), текущие планируемые маршруты отображаются пунктирными линиями. Панель управления включает выбор автоматического или ручного режима, настройку количества роботов, регулировку скорости моделирования и кнопки

«Запуск/Пауза/Сброс»; в ручном режиме пользователь может задать параметры груза и быстро загрузить типовые тестовые сценарии. Панель состояния задач и роботов отображает в виде списка информацию о текущих и завершённых задачах, а также в формате вкладок – данные о состоянии РП, РТ и РР в реальном времени. Область системного журнала в реальном времени отображает ключевые события: классификацию задач, назначения роботов, переходы состояний и запуски зарядки. Для проверки корректности системы использовался набор из 50 тестовых сценариев, охватывающих все три категории транспортных задач и включающих пограничные случаи: минимальный заряд РТ, одновременное поступление нескольких срочных грузов, недоступность погрузочной платформы. Во всех сценариях система быстро определила категорию задачи, выбрала роботы и сформировала согласованную траекторию вывода. Результаты некоторых сценариев представлены на рис. 3 и 4.

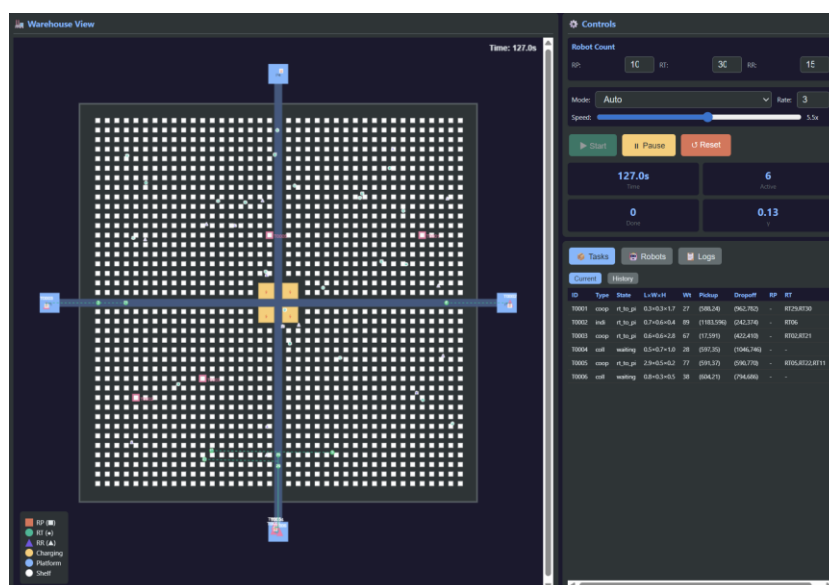


Рис. 3. Результат тестирования (автоматический режим)

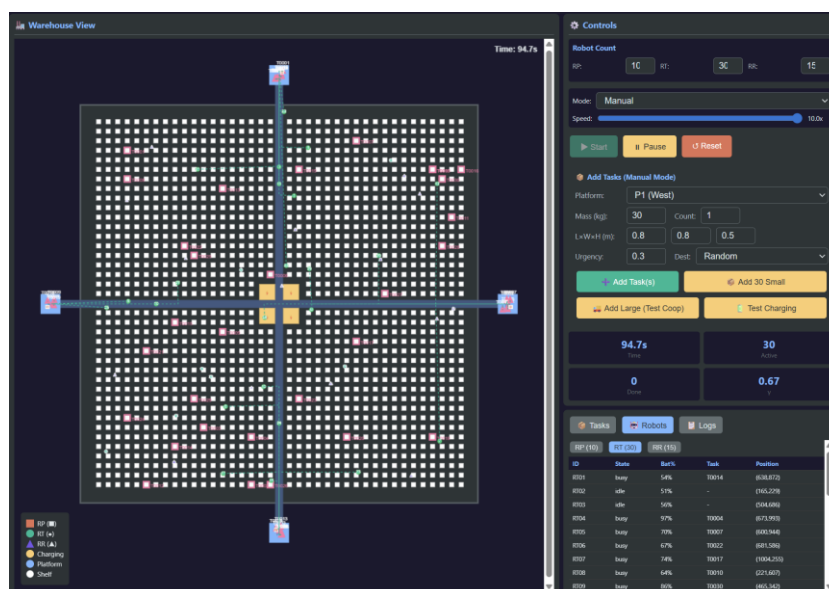


Рис. 4. Результат тестирования (ручной режим)

Экспериментальная проверка предложенного метода. Экспериментальная часть включает четыре серии исследований: сравнение с базовыми алгоритмами, анализ адаптивности, анализ надёжности и анализ расширяемости. Во всех экспериментах использовалась стандартная конфигурация: склад 1000×1000 м, 80 РТ, 20 РП, 40 РР, 300 задач (70 % индивидуальных, 15 % кооперативных, 15 % коллективных). Каждая серия повторялась

не менее 10 раз; результаты представлены в виде среднего значения и стандартного отклонения. Для всесторонней оценки эффективности алгоритмов распределения задач использовались следующие три ключевых показателя.

(1) Коэффициент выполнения задач CR :

$$CR = \frac{N_{\text{completed}}}{N_{\text{total}}} \times 100\% \quad (33)$$

Где $N_{\text{completed}}$ – количество успешно выполненных задач, N_{total} – общее количество задач.

(2) Среднее энергопотребление на задачу E_{ave} :

$$E_{\text{ave}} = \frac{\sum_{i=1}^N E_i}{CR \times 300} \quad (34)$$

Где E_i – энергопотребление i -го робота, N – общее количество роботов.

(3) Коэффициент полезного использования транспортных роботов U_{RT} :

$$U_{RT} = \frac{\sum_{i=1}^{N_{RT}} T_{\text{transport},i}}{T_{\text{total}} \times N_{RT}} \times 100\% \quad (35)$$

Где $T_{\text{transport},i}$ – фактическое время транспортировки i -го транспортного робота, T_{total} – общая продолжительность моделирования, N_{RT} – общее количество транспортных роботов.

Сравнение с базовыми алгоритмами. Предложенный метод (Mivar) сравнивался с двумя базовыми алгоритмами – «ближайший сосед» (Nearest) и «жадный» (Greedy) – при трёх значениях интенсивности поступления задач $\lambda = 0,05; 0,2; 0,5$. Результаты по коэффициенту выполнения задач представлены на рис. 5.

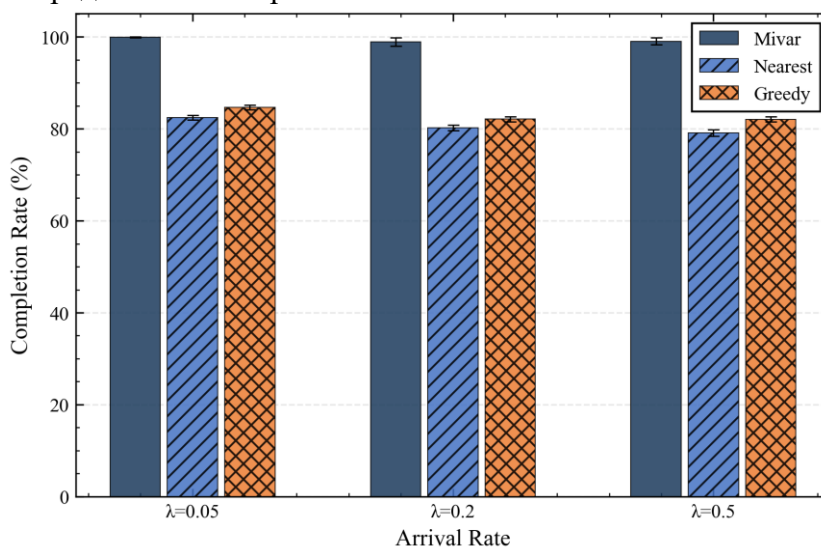


Рис. 5. Коэффициент выполнения задач CR при различной интенсивности поступления

Метод Mivar демонстрирует исключительно высокий и стабильный CR при всех трёх уровнях интенсивности (99,97 %, 98,93 % и 99,07 % при стандартном отклонении не выше 1 %). Базовые алгоритмы отстают на 14–18 процентных пунктов (79–85 %), что объясняется двумя причинами: отсутствием механизма формирования команд РТ для кооперативных задач и отсутствием проверки остаточного заряда роботов перед назначением.

Из рис. 6 следует, что алгоритм Nearest характеризуется наименьшим средним энергопотреблением (10,80–14,76 усл. ед.), поскольку назначает задачу ближайшему РТ и минимизирует длину маршрута; однако именно из-за этой «близорукости» он не справляется с удалёнными и кооперативными задачами, что и обуславливает самый низкий CR . Энергопотребление метода Mivar (11,96–15,66 усл. ед.) лишь немного превышает Nearest и существенно ниже значений алгоритма Greedy ($\approx 18,5$ усл. ед.). Небольшое превышение относительно Nearest связано с необходимостью совместной перевозки крупногабаритных

грузов несколькими РТ; при этом пакетная группировка и оптимизация маршрутов коллективных задач позволяют сократить число повторных пробегов, благодаря чему Mivar оказывается существенно экономичнее Greedy.

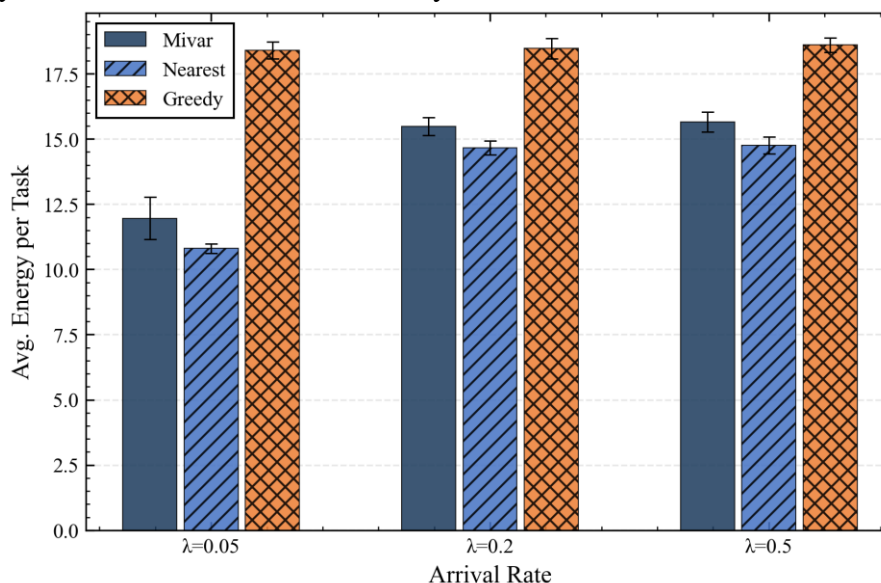


Рис. 6. Среднее энергопотребление при различной интенсивности поступления

Из рис. 7 видно, что коэффициент полезного использования РТ при использовании метода Mivar существенно превышает показатели базовых алгоритмов и возрастает с ростом интенсивности: с 44,80 % при $\lambda = 0,05$ до 74,95 % при $\lambda = 0,5$. Повышенная эффективность обусловлена тремя факторами: энерго-ориентированное планирование, механизм группирования для кооперативных задач и группировка коллективных задач.

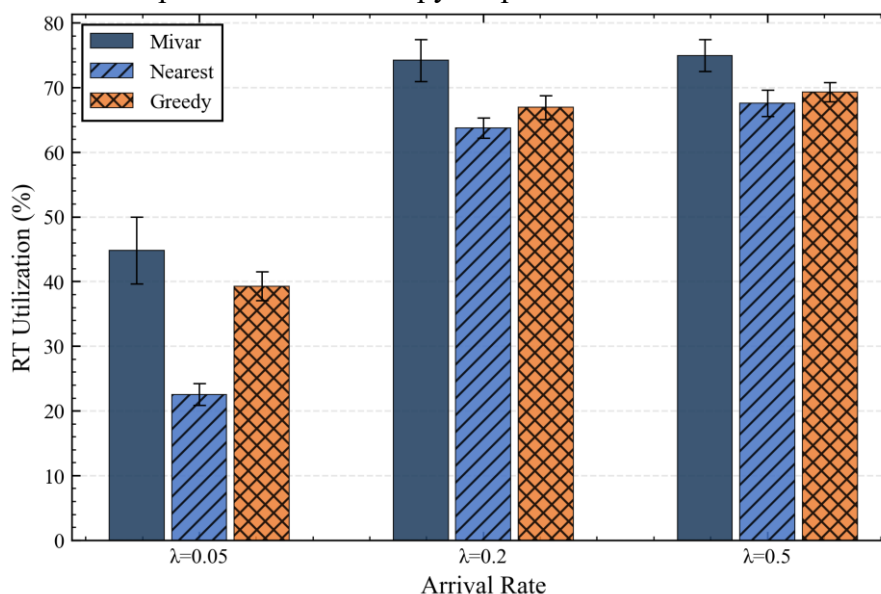


Рис. 7. Коэффициент полезного использования РТ при различной интенсивности поступления

Анализ адаптивности. Целью данного эксперимента является анализ адаптивности метода Mivar при различных распределениях типов задач и конфигурациях роботов. Метод исследовался при трёх распределениях типов задач – А (80/10/10), В (60/20/20) и С (50/30/20) – и при трёх конфигурациях – ограниченной, стандартной и расширенной. Результаты приведены в таб. 2 и 3.

Таблица 2. Показатели метода Mivar при различных распределениях задач

Распределение задач	Коэффициент выполнения (%)	Общее энергопотребление	Коэффициент использования РТ (%)
A (80/10/10)	99,77±0,42	4722,20±78,60	76,54±3,31
B (60/20/20)	98,03±1,41	5053,44±115,65	71,13±2,99
C (50/30/20)	94,37±1,04	5161,55±122,66	71,10±2,68

Таблица 3. Показатели метода Mivar при различных конфигурациях роботов

Конфигурация роботов	Коэффициент выполнения (%)	Общее энергопотребление	Коэффициент использования РТ (%)
Ограниченная	39,37±2,16	1732,38±98,52	65,35±18,40
Стандартная	98,93±0,87	4852,92±102,67	74,21±3,24
Расширенная	100,00	4795,74±131,54	73,97±2,36

При увеличении доли кооперативных и коллективных задач наблюдается плавное снижение CR с 99,77 % до 94,37 %, связанное с ростом затрат на сбор команд РТ в одной точке и увеличением холостого пробега; тем не менее коэффициент выполнения остаётся выше 94 % даже при доле кооперативных задач 30 %, что подтверждает адаптивность метода к изменению структуры нагрузки. При изменении конфигурации в ограниченной конфигурации CR падает до 39,37 % вследствие острого дефицита РТ, однако в стандартной и расширенной конфигурациях достигает 98,93–100,00%. Таким образом, производительность предложенного метода ограничивается прежде всего физическим ресурсом группы, а не вычислительными возможностями алгоритма, что свидетельствует о способности метода максимизировать эффективность в пределах доступных ресурсов.

Анализ надёжности. Для оценки надёжности проведены эксперименты с внесением отказов при пяти условиях: без отказов, 10 %, 20 %, 30 % отказов и 20 % отказов с активированным механизмом автоматического восстановления. Отказы моделировались как случайный выход роботов из строя во время выполнения задачи. Результаты сведены на рис. 8.

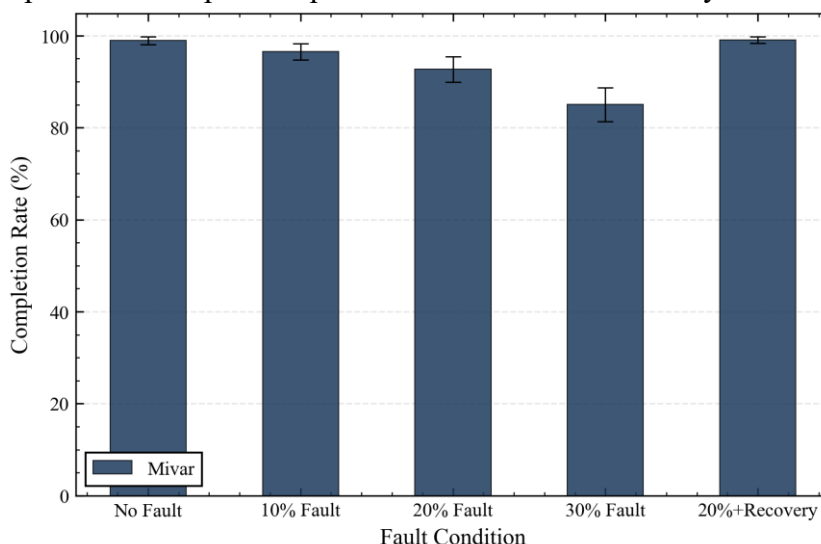


Рис. 8. Надёжность метода Mivar при различных уровнях отказов

С увеличением уровня отказов с 10 % до 30 % CR снижается с 96,53 % до 85,03 %, а количество невыполненных задач возрастает с 10 до 45: внесение отказов напрямую уменьшает число доступных роботов и увеличивает нагрузку на оставшихся. Принципиально важным является результат, полученный при 20 % отказов с автоматическим восстановлением:

CR возвращается к 99,07 %, а число невыполненных задач снижается до 2,8 – практически до уровня нормального режима.

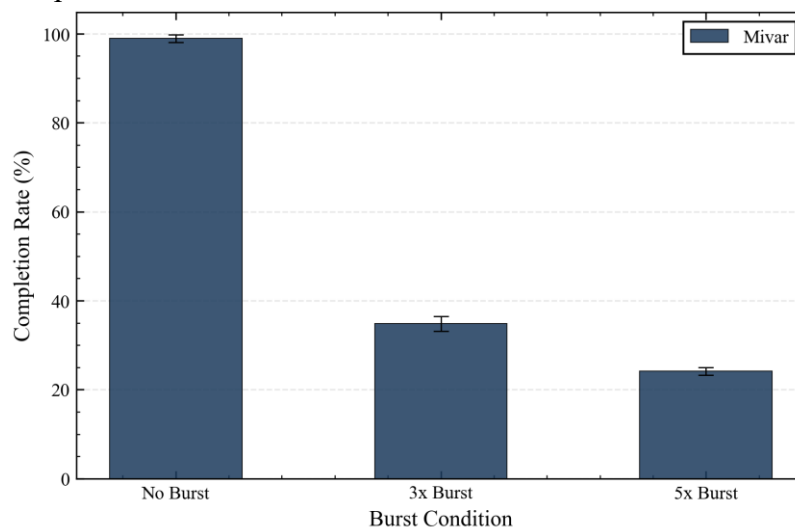


Рис. 9. Коэффициент выполнения задач метода Mivar при различных условиях внезапной нагрузки

Дополнительно проведены эксперименты с внезапным увеличением нагрузки в 3 и 5 раз (общее число задач возрастает с 300 до 900 и 1500 соответственно). Из рис. 9 следует, что CR резко снижается до 34,83 % и 24,13 %, а количество невыполненных задач достигает 782 и 1366. Данный результат отражает физический предел пропускной способности системы при фиксированном числе роботов: в условиях экстремальной нагрузки одной лишь алгоритмической оптимизации оказывается недостаточно, и требуется применение стратегий динамического масштабирования парка.

Анализ расширяемости. Целью данного эксперимента является оценка расширяемости метода Mivar в условиях различных масштабов с систематическим исследованием по двум направлениям: размер склада (500×500, 1000×1000 и 1500×1500) и плотность задач ($\lambda = 0,05; 0,2; 0,4$). Результаты эксперимента приведены в таб. 4 и 5.

Таблица 4. Показатели метода Mivar при различных масштабах склада

Масштаб склада	Коэффициент выполнения (%)	Общее энергопотребление	Среднее время ожидания (с)
500×500	72,47±5,43	1463,14±120,19	2934,04±247,53
1000×1000	98,93±0,87	4852,92±102,67	1408,70±97,24
1500×1500	100,00	6974,95±280,10	140,62±39,64

В малом складе 500×500 CR составил 72,47 %, а среднее время ожидания достигло 2934 с: 80 РТ, 20 РП и 40 РР функционируют в ограниченном пространстве с высокой плотностью, что приводит к пространственной загрузке, частым конфликтам маршрутов, перегрузке зарядных станций и конкуренции за погрузочно-разгрузочные посты. В стандартном складе 1000×1000 CR достигает 98,93 %, а в крупном 1500×1500 – практически 100 %: достаточное пространство позволяет избежать загроможденности, однако увеличение расстояний транспортировки вызывает закономерный рост общего энергопотребления.

Таблица 5. Показатели метода Mivar при различных плотностях задач

Плотность задач	Коэффициент выполнения (%)	Общее энергопотребление	Среднее время ожидания (с)
$\lambda=0,05$	99,97±0,10	4189,48±240,74	294,48±31,67

$\lambda=0,2$	$98,93\pm0,87$	$4852,92\pm102,67$	$1408,70\pm97,24$
$\lambda=0,4$	$99,13\pm0,76$	$4910,74\pm115,67$	$1673,20\pm91,99$

Из таб. 4 видно, что при изменении плотности задач от $\lambda = 0,05$ до $\lambda = 0,4$ CR сохраняется выше 99 %, а основное влияние проявляется в росте среднего времени ожидания – с 294 с до 1673 с. Полученные результаты показывают, что узкие места производительности метода Mivar связаны главным образом с ограничениями физической среды, а не с самим алгоритмом, что является важным аргументом в пользу его применимости в реальных складских сценариях различных масштабов.

Заключение. В настоящей работе предложена миварная система обработки информации и принятия решений для группового управления гетерогенными складскими роботами, объединяющая трёх функционально различных участников – роботов-погрузчиков, роботов-транспортировщиков и роботов-разгрузчиков. Разработана математическая модель, включающая классификацию транспортных задач на индивидуальные, кооперативные и коллективные, динамический расчёт приоритета грузов, оптимизацию маршрутов внутри группы коллективных задач и выбор исполнителей по критерию совместимости, учитывающему состояние, работоспособность, запас энергии и расстояние до точки погрузки. Для кооперативной транспортировки введён аукционный алгоритм формирования команд с учётом избыточности грузоподъёмности, гетерогенности скоростей и путевых затрат.

Математическая модель преобразована в миварную базу знаний, содержащую 302 правила вида «ЕСЛИ–ТО», и реализована на платформе КЭСМИ. Разработанная на языке Python имитационная система воспроизводит работу склада размером 1000×1000 м со 140 роботами трёх типов и служит основой для воспроизводимой экспериментальной проверки.

Экспериментальные исследования подтвердили эффективность, адаптивность, надёжность и расширяемость предложенного метода. В стандартных условиях коэффициент выполнения задач достигает $98,93 \pm 0,87$ %, что превышает показатели базовых алгоритмов «ближайший сосед» и «жадный» на 14–18 процентных пунктов при более высоком коэффициенте полезного использования РТ. Метод сохраняет CR выше 94 % при изменении доли кооперативных задач до 30 % и практически полностью восстанавливает работоспособность после отказа 20 % парка роботов благодаря механизму автоматического восстановления. При изменении масштаба склада и плотности задач узкие места производительности оказываются связанными главным образом с ограничениями физической среды, а не с самим алгоритмом.

Полученные результаты показывают, что миварные технологии представляют собой перспективный инструмент построения интеллектуальных складских систем нового поколения. Дальнейшие исследования будут направлены на интеграцию предложенной модели с методами обучения с подкреплением для автоматической настройки весовых коэффициентов функции совместимости, на расширение миварной базы знаний правилами обработки непредвиденных событий, а также на апробацию метода в реальной складской среде с участием физических роботов.

Список литературы

1. Tubis A.A., Rohman J. Intelligent warehouse in industry 4.0 – systematic literature review. *Sensors*, 2023, vol. 23, no. 8, p. 4105, DOI: 10.3390/s23084105..
2. An X., Wu C., Lin Y. et al. Multi-robot systems and cooperative object transport: communications, platforms, and challenges. *IEEE Open Journal of the Computer Society*, 2023, vol. 4, pp. 23-36, DOI: 10.1109/ojcs.2023.3238324.
3. Li Z., Barenji A.V., Jiang J. et al. A mechanism for scheduling multi robot intelligent warehouse system face with dynamic demand. *Journal of Intelligent Manufacturing*, 2020, vol. 31, no. 2, pp. 469-480, DOI: 10.1007/s10845-018-1459-y.

4. Fragapane G., De Koster R., Sgarbossa F., Strandhagen J.O. Planning and control of autonomous mobile robots for intralogistics: literature review and research agenda. *European Journal of Operational Research*, 2021, vol. 294, no. 2, pp. 405-426, DOI: 10.1016/j.ejor.2021.01.019.
5. Малышев М.И. Использование цифровых аналитических инструментов в специализированных интеллектуальных транспортно-логистических и складских системах // *Экономика. Общество. Человек: материалы национальной научно-практической конференции с международным участием*. – Белгород: БГТУ им. В.Г. Шухова, 2023. – С. 271-276. – EDN: EFOTCA.
6. Белоусов Е.А. Рекомендательная система диагностики сахарного диабета на основе механизма миварного вывода / Е.А. Белоусов, И.А. Попов, А.А. Евдокимов и др. // *Естественные и технические науки*, 2021. – № 7(158). – С. 169-174. – EDN: JSFUSI.
7. Макаров И.М. Модели и алгоритмы планирования действий и распределения заданий в мультиагентных робототехнических системах / И.М. Макаров, В.М. Лохин, С.В. Манько и др. // *Мехатроника, автоматизация, управление*, 2012. – № 5. – С. 44-50. – EDN: OYDIOV.
8. Лазарев В.С. Групповое управление роботами в условиях противодействия с учетом ограничения по запасу хода / В.С. Лазарев, И.В. Черногоров, В.А. Крухмалев // *Проблемы современной системотехники: сборник научных статей*. – Таганрог: ИП Ступин С.А., 2017. – Вып. 11. – С. 103-108. – EDN: YRADWQ.
9. Крестовников К.Д. Математическая модель и алгоритмы управления группой наземных роботов с перераспределением энергетических ресурсов // *Информационно-управляющие системы*, 2023. – № 6(127). – С. 20-34.
10. Yang S., Zhang Y., Ma L. et al. A novel maximin-based multi-objective evolutionary algorithm using one-by-one update scheme for multi-robot scheduling optimization. *IEEE Access*, 2021, vol. 9, pp. 121316-121328, DOI: 10.1109/access.2021.3105102.
11. Orr J., Dutta A. Multi-agent deep reinforcement learning for multi-robot applications: a survey. *Sensors*, 2023, vol. 23, no. 7, p. 3625, DOI: 10.3390/s23073625.
12. Tejada J.C., Toro-Ossaba A., López-Gonzalez A. et al. A review of multi-robot systems and soft robotics: challenges and opportunities. *Sensors*, 2025, vol. 25, no. 5, p. 1353, DOI: 10.3390/s25051353.
13. Чиженков Б.М. Миварный подход логического искусственного интеллекта / Б.М. Чиженков, А.В. Силаев, Н.А. Казаков и др. // *Искусственный интеллект: теоретические аспекты, практическое применение: материалы Донецкого международного научного круглого стола*. – Донецк: ИПИИ, 2022. – С. 199-203. – EDN: ZQVHHV.
14. Дошина А.Д. Экспертная система. Классификация. Обзор существующих экспертных систем // *Молодой ученый*, 2016. – № 21(125). – С. 756-758. – EDN: WYDQIL.
15. Халидуллин А.С. Обзор экспертных систем и перспективы их применения в подземных и открытых горных работах // *StudNet*, 2022. – Т. 5, № 6. – С. 165. – EDN: GOKPIB.
16. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. – М.: Радио и связь, 2002. – 286 с. – ISBN 5-256-01650-4. – EDN: RWTCOP.
17. Шэнь Ц. Динамическое планирование траектории робота на основе семантического обнаружения объектов с использованием миварной экспертной системы / Ц. Шэнь, Ш. Гун, О.О. Варламов и др. // *Проблемы искусственного интеллекта*, 2024. – № 4(35). – С. 164-176. – DOI: 10.24412/2413-7383-2024-4-164-176.
18. Linghan D., Shengshuo G., Qiujie S., Varlamov O. Wind turbine fault diagnosis based on a mivar expert system. *E3S Web of Conferences*, 2025, vol. 646, p. 00034, DOI: 10.1051/e3sconf/202564600034.
19. Varlamov O., Aladin D. A new generation of rules-based approach: mivar-based intelligent planning of robot actions (MIPRA) and brains for autonomous robots. *Machine Intelligence Research*, 2024, vol. 21, no. 5, pp. 919-940, DOI: 10.1007/s11633-023-1473-1.
20. Kotsenko A., Varlamov O., Denisov A. et al. Mivar-based route planning simulation model for obstacle-aware autonomous agricultural machinery. *Bio Web of Conferences: International Scientific Conference on Biotechnology and Food Technology (BFT-2024)*, Les Ulis, EDP Sciences, 2024, vol. 130, p. 08031, DOI: 10.1051/bioconf/202413008031.
21. Коценко А.А. Модель описания миварных сетей в формате двудольных и трехдольных ориентированных графов для принятия решений и обработки информации в машиностроительном ИИ / А.А. Коценко, Б.С. Горячкин, А.Г. Базанова и др. // *Динамика сложных систем - XXI век*, 2024. – Т. 18, № 1. – С. 5-17. – DOI: 10.18127/j19997493-202401-01.
22. Чувилов Д.А. Автоматизация системы диспетчерского контроля на основе экспертной системы городского пассажирского транспорта / Д.А. Чувилов, П. Джха // *Автоматизация и управление в технических системах*, 2017. – № 2(23). – С. 2. – EDN: YPHNMU.

23. Аладин Д.В. Миварная экспертная система для контроля соблюдения правил дорожного движения / Д.В. Аладин, Е.В. Ромичева, М.Ю. Чекулина // Интеллектуальные технологии и проблемы математического моделирования: тезисы докладов III Всероссийской научной конференции. – Ростов-на-Дону: ДГТУ, 2020. – С. 28. – EDN: RXZRXC.
24. Варламов О.О. Применение миварных технологий логического искусственного интеллекта для решения задач распределения ресурсов производственных систем / О.О. Варламов, О.В. Кривошеев // Системы управления и информационные технологии, 2022. – № 1(87). – С. 49-56. – DOI: 10.36622/VSTU.2022.87.1.011.
25. Аладин Д.В. Метод автоматической генерации миварных баз знаний для решения управленческих задач в пространстве состояний // Проблемы искусственного интеллекта, 2025. – № 3(38). – С. 88-99. – DOI: 10.24412/2413-7383-2025-3-38-88-99.
26. Доу Л. Модель принятия решений для обнаружения пожаров на основе распознавания образов и миварной экспертной системы // Системы управления и информационные технологии, 2025. – № 3(101). – С. 59-65. – EDN: NJVEWG.
27. Шэнь Ц. Миварные модель и алгоритм обработки информации для избегания динамических препятствий с изменяющейся зоной безопасности // Моделирование, оптимизация и информационные технологии, 2026. – Т. 14, № 1(52). – DOI: 10.26102/2310-6018/2026.52.1.004.
28. Гун Ш. Миварная система принятия решений для распределения и перевозки грузов командой складских роботов // Системы управления и информационные технологии, 2025. – № 2(100). – С. 23-29. – EDN: XZHHBF.
29. Гун Ш. Миварная система принятия решений по оптимизационному распределению грузов для групп складских роботов // Моделирование, оптимизация и информационные технологии, 2025. – Т. 13, № 3(50). – DOI: 10.26102/2310-6018/2025.50.3.047.
30. Гун Ш. Модель принятия решений для многозадачных складских роботов на основе четырехуровневой архитектуры совместной оптимизации // Проблемы искусственного интеллекта, 2025. – № 3(38). – С. 100-112. – DOI: 10.24412/2413-7383-2025-3-38-100-112.

Гун Шэншо. Аспирант Московского государственного технического университета имени Н.Э. Баумана (МГТУ им. Н.Э. Баумана), автор теоретических и экспериментальных исследований по применению логического искусственного интеллекта для группового управления роботами. AuthorID: 1264893; SPIN: 9392-9017, ORCID: 0000-0001-6177-7989. hiteyeb@163.com. 105005, Россия, Москва.

UDC 004.89 + 007.52 + 681.518

DOI:10.25729/ESI.2026.43.3.015

A Mivar information processing and decision-making system for group control of heterogeneous warehouse robots

Shengshuo Gong

Bauman Moscow state technical university, Russia, Moscow, hiteyeb@163.com

Abstract. This paper addresses the problem of group control of functionally heterogeneous warehouse robots – loading robots (LR), transport robots (TR), and unloading robots (UR). The aim is to develop a Mivar-based information processing and decision-making system that provides dynamic classification of transport tasks, executor selection via a compatibility criterion, and adaptive fleet energy management while preserving linear inference complexity and full decision explainability. A three-category task classification – individual, cooperative, and collective transportation – is proposed, implemented through Boolean indicators, dynamic priority computation, and dimensional-weight constraint verification. For collective transportation, a spatial clustering mechanism with dual-mode route optimization (exhaustive search and nearest-neighbor combined with 2-opt) is developed. For cooperative transportation of oversized cargo, a combinatorial auction mechanism for TR team formation is introduced, minimizing a composite cost function accounting for payload redundancy, speed heterogeneity, and travel costs. State, operability, and energy sufficiency equations are formulated for each robot type. The mathematical model is transformed into a Mivar knowledge base of 302 IF-THEN rules and verified on

the KESMI platform. The novelty lies in the simultaneous integration of fleet functional heterogeneity, three-category task classification, and adaptive energy management within a unified Mivar knowledge base – aspects previously addressed only in isolation. Experimental validation on a simulation system (1000×1000 m warehouse, 140 robots, 300 tasks) demonstrated a task completion rate of 98.9–99.9%, exceeding baseline Nearest Neighbor and Greedy algorithms by 14–18 percentage points. The method maintains operability under failure of up to 30% of robots and with cooperative task shares reaching 30%.

Keywords: mivar, mivar decision-making system, mivar knowledge base, logical artificial intelligence, warehouse robot, group control, task allocation

References

1. Tubis A.A., Rohman J. Intelligent warehouse in industry 4.0 – systematic literature review. *Sensors*, 2023, vol. 23, no. 8, p. 4105, DOI: 10.3390/s23084105..
2. An X., Wu C., Lin Y. et al. Multi-robot systems and cooperative object transport: communications, platforms, and challenges. *IEEE Open Journal of the Computer Society*, 2023, vol. 4, pp. 23-36, DOI: 10.1109/ojcs.2023.3238324.
3. Li Z., Barenji A.V., Jiang J. et al. A mechanism for scheduling multi robot intelligent warehouse system face with dynamic demand. *Journal of Intelligent Manufacturing*, 2020, vol. 31, no. 2, pp. 469-480, DOI: 10.1007/s10845-018-1459-y.
4. Fragapane G., De Koster R., Sgarbossa F., Strandhagen J.O. Planning and control of autonomous mobile robots for intralogistics: literature review and research agenda. *European Journal of Operational Research*, 2021, vol. 294, no. 2, pp. 405-426, DOI: 10.1016/j.ejor.2021.01.019.
5. Malyshev M.I. Ispol'zovaniye tsifrovyykh analiticheskikh instrumentov v spetsializirovannykh intellektual'nykh transportno-logisticheskikh i skladskikh sistemakh [Use of digital analytical tools in specialized intelligent transport-logistics and warehouse systems]. *Ekonomika. Obshchestvo. Chelovek: materialy natsional'noy nauchno-prakticheskoy konferentsii s mezhdunarodnym uchastiyem [Economics. Society. Human: Proceedings of the National Scientific and Practical Conference with International Participation]*. Belgorod, BSTU Publ., 2023, pp. 271-276, EDN: EFOTCA.
6. Belousov E.A., Popov I.A., Evdokimov A.A. et al. Rekomendatel'naya sistema diagnostiki sakharnogo diabeta na osnove mekhanizma mivarnogo vyvoda [Recommendation system for diabetes diagnosis based on the mivar inference mechanism]. *Estestvennyye i tekhnicheskiye nauki [Natural and Technical Sciences]*, 2021, no. 7(158), pp. 169-174, EDN: JSFUSI.
7. Makarov I.M., Lokhin V.M., Manko S.V. et al. Modeli i algoritmy planirovaniya deystviy i raspredeleniya zadaniy v mul'tiagentnykh robototekhnicheskikh sistemakh [Models and algorithms for action planning and task allocation in multi-agent robotic systems]. *Mekhatronika, avtomatizatsiya, upravleniye [Mechatronics, Automation, Control]*, 2012, no. 5, pp. 44-50, EDN: OYDIOV.
8. Lazarev V.S., Chernogorov I.V., Krukhmalyov V.A. Gruppovoye upravleniye robotami v usloviyakh protivodeystviya s uchedom ogranicheniya po zapasu khoda [Group control of robots under adversarial conditions with limited range]. *Problemy sovremennoy sistemotekhniki [Problems of Modern Systems Engineering]*, Taganrog, Stupin S.A. Publ., 2017, iss. 11, pp. 103-108, EDN: YRADWQ.
9. Krestovnikov K.D. Matematicheskaya model' i algoritmy upravleniya gruppoy nazemnykh robotov s pereraspredeleniyem energeticheskikh resursov [Mathematical model and algorithms for controlling a group of ground robots with redistribution of energy resources]. *Informatsionno-upravlyayushchiye sistemy [Information and Control Systems]*, 2023, no. 6(127), pp. 20-34, DOI: 10.31799/1684-8853-2023-6-20-34.
10. Yang S., Zhang Y., Ma L. et al. A novel maximin-based multi-objective evolutionary algorithm using one-by-one update scheme for multi-robot scheduling optimization. *IEEE Access*, 2021, vol. 9, pp. 121316-121328, DOI: 10.1109/access.2021.3105102.
11. Orr J., Dutta A. Multi-agent deep reinforcement learning for multi-robot applications: a survey. *Sensors*, 2023, vol. 23, no. 7, p. 3625, DOI: 10.3390/s23073625.
12. Tejada J.C., Toro-Ossaba A., López-Gonzalez A. et al. A review of multi-robot systems and soft robotics: challenges and opportunities. *Sensors*, 2025, vol. 25, no. 5, p. 1353, DOI: 10.3390/s25051353.
13. Chizhenkov B.M., Silaev A.V., Kazakov N.A. et al. Mivarnyy podkhod logicheskogo iskusstvennogo intellekta [Mivar approach of logical artificial intelligence]. *Iskusstvennyy intellekt: teoreticheskiye aspekty, prakticheskoye primeneniye: materialy Donetskogo mezhdunarodnogo nauchnogo kruglogo stola [Artificial Intelligence: Theoretical Aspects, Practical Application: Proceedings of the Donetsk International Scientific Round Table]*. Donetsk, IPAI Publ., 2022, pp. 199-203, EDN: ZQVHHV.
14. Doshina A.D. Ekspertnaya sistema. Klassifikatsiya. Obzor sushchestvuyushchikh ekspertnykh sistem [Expert system. Classification. Review of existing expert systems]. *Molodoy uchenyy [Young Scientist]*, 2016, no. 21(125), pp. 756-758, EDN: WYDQIL.

15. Khalidullin A.S. Obzor ekspertnykh sistem i perspektivy ikh primeneniya v podzemnykh i otkrytykh gornyykh rabotakh [Review of expert systems and prospects for their application in underground and open-pit mining]. StudNet, 2022, vol. 5, no. 6, p. 165, EDN: GOKPIB.
16. Varlamov O.O. Evolyutsionnyye bazy dannykh i znaniy dlya adaptivnogo sinteza intellektual'nykh sistem. Mivarnoye informatsionnoye prostranstvo [Evolutionary databases and knowledge bases for adaptive synthesis of intelligent systems. Mivar information space]. Moscow, Radio i svyaz [Radio and Communication] Publ., 2002, 286 p. ISBN 5-256-01650-4, EDN: RWTCOP.
17. Shen C., Gong S., Varlamov O.O. et al. Dinamicheskoye planirovaniye trayektorii robota na osnove semanticheskogo obnaruzheniya obyekтов s ispol'zovaniyem mivarnoy ekspertnoy sistemy [Dynamic robot trajectory planning based on semantic object detection using a mivar expert system]. Problemy iskusstvennogo intellekta [Problems of Artificial Intelligence], 2024, no. 4(35), pp. 164-176, DOI: 10.24412/2413-7383-2024-4-164-176.
18. Linghan D., Shengshuo G., Qiujie S., Varlamov O. Wind turbine fault diagnosis based on a mivar expert system. E3S Web of Conferences, 2025, vol. 646, p. 00034, DOI: 10.1051/e3sconf/202564600034.
19. Varlamov O., Aladin D. A new generation of rules-based approach: mivar-based intelligent planning of robot actions (MIPRA) and brains for autonomous robots. Machine Intelligence Research, 2024, vol. 21, no. 5, pp. 919-940, DOI: 10.1007/s11633-023-1473-1.
20. Kotsenko A., Varlamov O., Denisov A. et al. Mivar-based route planning simulation model for obstacle-aware autonomous agricultural machinery. Bio Web of Conferences: International Scientific Conference on Biotechnology and Food Technology (BFT-2024), Les Ulis, EDP Sciences, 2024, vol. 130, p. 08031, DOI: 10.1051/bioconf/202413008031.
21. Kotsenko A.A., Goryachkin B.S., Bazanova A.G. et al. Model' opisaniya mivarnykh setey v formate dvudol'nykh i trekhdol'nykh orientirovannykh grafov dlya prinyatiya resheniy i obrabotki informatsii v mashinostroitel'nom II [A model for describing mivar networks in the format of bipartite and tripartite directed graphs for decision-making and information processing in engineering AI]. Dinamika slozhnykh sistem - XXI vek [Dynamics of Complex Systems - XXI Century], 2024, vol. 18, no. 1, pp. 5-17, DOI: 10.18127/j19997493-202401-01.
22. Chuvikov D.A., Jha P. Avtomatizatsiya sistemy dispetcherskogo kontrolya na osnove ekspertnoy sistemy gorodskogo passazhirskogo transporta [Automation of a dispatch control system based on an expert system for urban passenger transport]. Avtomatizatsiya i upravleniye v tekhnicheskikh sistemakh [Automation and Control in Technical Systems], 2017, no. 2(23), p. 2, EDN: YPHHMu.
23. Aladin D.V., Romicheva E.V., Chekulina M.Yu. Mivarnaya ekspertnaya sistema dlya kontrolya soblyudeniya pravil dorozhnogo dvizheniya [Mivar expert system for monitoring compliance with traffic rules]. Intellektual'nyye tekhnologii i problemy matematicheskogo modelirovaniya: tezisy dokladov III Vserossiyskoy nauchnoy konferentsii [Intelligent Technologies and Problems of Mathematical Modeling: Abstracts of the III All-Russian Scientific Conference]. Rostov-on-Don, DSTU Publ., 2020, p. 28, EDN: RXZRXC.
24. Varlamov O.O., Krivosheev O.V. Primeneniye mivarnykh tekhnologiy logicheskogo iskusstvennogo intellekta dlya resheniya zadach raspredeleniya resursov proizvodstvennykh sistem [Application of mivar technologies of logical artificial intelligence for solving resource allocation problems in production systems]. Sistemy upravleniya i informatsionnyye tekhnologii [Control Systems and Information Technology], 2022, no. 1(87), pp. 49-56, DOI: 10.36622/VSTU.2022.87.1.011.
25. Aladin D.V. Metod avtomaticheskoy generatsii mivarnykh baz znaniy dlya resheniya upravlencheskikh zadach v prostranstve sostoyaniy [Method of automatic generation of mivar knowledge bases for solving management problems in state space]. Problemy iskusstvennogo intellekta [Problems of Artificial Intelligence], 2025, no. 3(38), pp. 88-99, DOI: 10.24412/2413-7383-2025-3-38-88-99.
26. Dou L. Model' prinyatiya resheniy dlya obnaruzheniya pozharov na osnove raspoznavaniya obrazov i mivarnoy ekspertnoy sistemy [A decision-making model for fire detection based on pattern recognition and a mivar expert system]. Sistemy upravleniya i informatsionnyye tekhnologii [Control Systems and Information Technology], 2025, no. 3(101), pp. 59-65, EDN: NJVEWG.
27. Shen C. Mivarnyye model' i algoritm obrabotki informatsii dlya izbeganiya dinamicheskikh prepyatstviy s izmenyayushcheysoy zonoy bezopasnosti [Mivar model and information processing algorithm for avoiding dynamic obstacles with a changing safety zone]. Modelirovaniye, optimizatsiya i informatsionnyye tekhnologii [Modeling, Optimization and Information Technology], 2026, vol. 14, no. 1(52), DOI: 10.26102/2310-6018/2026.52.1.004.
28. Gong S. Mivarnaya sistema prinyatiya resheniy dlya raspredeleniya i perevozki грузов командой складских роботов [Mivar decision-making system for distribution and transportation of goods by a team of warehouse robots]. Sistemy upravleniya i informatsionnyye tekhnologii [Control Systems and Information Technology], 2025, no. 2(100), pp. 23-29, EDN: XZHHBF.

29. Gong S. Mivarnaya sistema prinyatiya resheniy po optimizatsionnomu raspredeleniyu грузов dlya grupp skladских robotov [Mivar decision-making system for optimal cargo distribution for groups of warehouse robots]. Modelirovaniye, optimizatsiya i informatsionnyye tekhnologii [Modeling, Optimization and Information Technology], 2025, vol. 13, no. 3(50), DOI: 10.26102/2310-6018/2025.50.3.047.
30. Gong S. Model' prinyatiya resheniy dlya mnogozadachnykh skladских robotov na osnove chetyrehurovnevoy arkhitektury sovmestnoy optimizatsii [Decision-making model for multi-task warehouse robots based on a four-level joint optimization architecture]. Problemy iskusstvennogo intellekta [Problems of Artificial Intelligence], 2025, no. 3(38), pp. 100-112, DOI: 10.24412/2413-7383-2025-3-38-100-112.

Gong Shengshuo. Postgraduate student at Bauman Moscow State Technical University (BMSTU), author of theoretical and experimental research on the application of logical artificial intelligence for group control of robots. AuthorID: 1264893; SPIN: 9392-9017; ORCID: 0000-0001-6177-7989. hiteyeb@163.com. 105005, Russia, Moscow.

Статья поступила в редакцию 26.06.2026; одобрена после рецензирования 22.07.2026; принята к публикации 11.08.2026.

The article was submitted 06/26/2026; approved after reviewing 07/22/2026; accepted for publication 08/11/2026.

УДК 004.89

DOI:10.25729/ESI.2026.43.3.016

Иерархическая модель структуры документа для решения задачи генерации конспектов лекций

Исмагулов Милан Ерикович

Югорский государственный университет,

Россия, Ханты-Мансийск, m_ismagulov@ugrasu.ru

Аннотация. В статье предлагается модель описания структуры документа, основанная на методах анализа иерархической структуры документа (HDSA) и формализме теории графов. В рамках предложенного подхода документ рассматривается как иерархическая система взаимосвязанных структурных элементов, включающая заголовки различных уровней, текстовые блоки, таблицы, формулы, изображения и иные типы контента. Каждый элемент интерпретируется, как вершина графа, а отношения вложенности и логической зависимости между элементами описываются с использованием ориентированных ребер графовой структуры.

Применение предложенной модели к корпусу документов из различных предметных областей позволило выявить устойчивые закономерности организации текстового материала, характерные для конкретных типов документов. На основе полученных характеристик был сформирован набор JSON-шаблонов, описывающих типовые структуры документов для различных предметных областей, включая технические, юридические и научно-образовательные материалы.

Каждый шаблон содержит формализованное описание структуры документа, включая допустимые уровни заголовков, ограничения на типы содержимого, а также параметры стилевого и композиционного оформления. Дополнительно шаблон может включать сведения о предпочтительном формате представления данных, правилах генерации таблиц, формул и специализированных текстовых конструкций. Использование JSON-формата обеспечивает машинно-ориентированное представление структуры документа и возможность интеграции шаблонов в программные системы автоматической генерации текста.

Разработанные шаблоны предназначены для использования в процессе генерации текстов большими языковыми моделями (Large Language Models – LLM). В рамках предложенного подхода шаблон выступает как формализованное структурное ограничение, определяющее допустимую организацию итогового документа и последовательность формирования его разделов. Это позволяет снизить вариативность структуры генерируемого текста, повысить согласованность взаимосвязанных разделов и обеспечить соответствие результата требованиям конкретной предметной области.

Предложенный подход ориентирован на реализацию управляемой генерации документов, при которой большая языковая модель выполняет генерацию содержимого в рамках заранее заданной структурной схемы. За счет этого обеспечивается более стабильное качество формирования документа, уменьшается вероятность нарушения логики построения текста и повышается воспроизводимость результатов генерации. Кроме того, использование структурных шаблонов позволяет адаптировать процесс генерации под различные типы документов без изменения архитектуры самой языковой модели.

Ключевые слова: иерархическая структура документа, теория графов, JSON-шаблоны, большие языковые модели (LLM), структурированная генерация текста, структурные паттерны, модель представления документа

Цитирование: Исмагулов М.Е. Иерархическая модель структуры документа для решения задачи генерации конспектов лекций / М.Е. Исмагулов // Информационные и математические технологии в науке и управлении, 2026. – № 3(43). – С. 197-210. – DOI:10.25729/ESI.2026.43.3.016.

Введение. В современных задачах генерации текста всё большую роль играет формирование структурированного вывода, согласованного с требованиями конкретной предметной области. Вместе с тем большинство существующих подходов опираются на универсальные шаблоны структуры документа и недостаточно учитывают реальные иерархические закономерности текстов. В настоящем исследовании предлагается модель оценки и использования иерархической структуры документа на основе метода HDSA и теории графов, позволяющая формировать JSON-шаблоны и обеспечивать контролируемую генерацию текста большими языковыми моделями [1, 2, 3].

1. Метод анализа иерархической структуры документа. Метод HDSA (Hierarchical Document Structure Analysis – иерархический анализ структуры документов) был разработан для решения задачи структурного разделения содержимого различных текстовых документов. Его основная цель заключается в выделении семантических блоков текста (заголовков, разделов, абзацев, параграфов с формулами и др.) с последующим построением иерархии и установлением связей между ними [4]. Наиболее простой и распространённый тип связи – отношение «родитель → ребёнок», отражающее вложенность блоков и обеспечивающее представление документа в виде дерева.

Формально документ можно представить, как дерево:

$$T = (V, E, r)$$

где V – множество вершин, соответствующих структурным блокам, $E \subseteq V \times V$ – множество рёбер, отражающих иерархические связи, а $r \in V$ – корень, интерпретируемый, как вся страница документа. Для каждой вершины $v \in V$ задаётся её тип $\tau(v)$ (заголовок, абзац, колонтитул и др.) и глубина $d(v)$, равная расстоянию до корня.

1.1. Связь между методами анализа иерархической структуры документа и анализом макета документа. Метод HDSA близок к Document Layout Analysis (DLA), однако между ними существует принципиальное различие. DLA направлен на анализ пространственного расположения текстовых блоков на странице и применяется преимущественно для восстановления визуальной структуры документа, сегментации и макетирования [5, 6]. HDSA же рассматривает документы с позиции логико-семантических связей между структурными единицами текста. Таким образом, DLA полезен для задач оптического распознавания символов и реконструкции макета [7], а HDSA – для анализа логической организации содержания документа и построения моделей его структурного представления [8, 9].

Визуальное подтверждение иерархической структуры документа обычно представлено в оглавлении. Содержание является линейным отображением HDSA-дерева: каждая вершина иерархии (глава, раздел, подраздел и т.д.) представлена в виде строки с соответствующим уровнем вложенности. Уровень определяется средствами визуального форматирования – нумерацией, отступами или шрифтовыми акцентами [10].

1.2. Основная идея анализа. Анализ страницы документа начинается с её разбиения на текстовые блоки с присвоением каждому из них уникального индекса. Страница интерпретируется, как корень дерева, содержащий как содержательные, так и служебные элементы (колонтитулы, номера страниц, сноски), которые образуют отдельные ветви. Далее формируется иерархия содержательных блоков – разделов, подразделов, абзацев и специализированных элементов, в результате чего документ представляется в виде дерева, отражающего его логико-семантическую структуру (рисунок 1).

Анализ иерархической структуры документа начинается с визуальной сегментации страницы на текстовые блоки по пробелам между ними. Далее строится иерархия блоков: для одно колоночных документов – линейно сверху вниз, для многоколоночных – с предварительным определением направления чтения. Для упорядочивания элементов задаётся весовое правило, согласно которому заголовки обладают наибольшим иерархическим приоритетом по отношению к остальным типам блоков. Особенностью анализа, представленного в данном исследовании, заключается представление HDSA в табличной форме, обеспечивающей формализованное представление HDSA-структуры (таблица 1).

Анализ документов от логической иерархии к структурированному представлению

1. Введение
Сравнительные исследования в области обработки естественного языка демонстрируют устойчивый интерес к проблеме структурного анализа документов. В отличие от традиционного подхода, рассматривающего текст исключительно как линейную последовательность предложений, новые методы ориентированы на выявление иерархических зависимостей между различными элементами документа.

1.1. Ключевые направления анализа
В современной литературе можно выделить несколько ключевых направлений структурного анализа текста, каждое из которых имеет собственные методологические и прикладные особенности:

- Выделение абзацев и заголовков. На этом этапе документ сегментируется на основные структурные элементы, определяющие иерархию, исключая главы, разделы, подразделы и пункты.
- Определение элементов библиографии. Здесь формируются иерархии, включающие главы, разделы, подразделы и пункты.
- Работа со специализированными элементами. В их число входят формулы, таблицы, рисунки и списки, которые требуют особого представления.

Таким образом, можно отметить, что различные структурные единицы взаимодействуют друг с другом в рамках единой иерархической структуры, распределяя условную вероятность в логической модели:

$$p(y | x, \theta) = \frac{\exp(\theta^T \varphi(x, y))}{\sum_{y' \in Y} \exp(\theta^T \varphi(x, y'))}$$

Полученное выражение описывает условное распределение классов $y \in Y$ при фиксированном классе x . Вектор признаков $\varphi(x, y)$ описывает представление объекта, а параметры модели θ задают веса, оптимизируемые в процессе обучения. Нормированный максимум по элементу играет роль функции распределения вероятностей, обеспечивая выполнение условия $\sum_{y \in Y} p(y | x, \theta) = 1$. Следует отметить, что максимизация правдоподобия в данной модели сводится к задаче минимизации функции потерь вида:

$$L(\theta) = - \sum_{i=1}^n \log p(y_i | x_i, \theta),$$

что делает метод особенно удобным для практического применения в задачах классификации и структурированного представления.

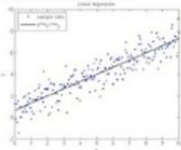


Рис.1 - Линейная регрессия

© Автор(ы), 2025. Все права защищены.

H
D
S
A

Анализ документов от логической иерархии к структурированному представлению

1. Введение
Сравнительные исследования в области обработки естественного языка демонстрируют устойчивый интерес к проблеме структурного анализа документов. В отличие от традиционного подхода, рассматривающего текст исключительно как линейную последовательность предложений, новые методы ориентированы на выявление иерархических зависимостей между различными элементами документа.

1.1. Ключевые направления анализа
В современной литературе можно выделить несколько ключевых направлений структурного анализа текста, каждое из которых имеет собственные методологические и прикладные особенности:

- Выделение абзацев и заголовков. На этом этапе документ сегментируется на основные структурные элементы, определяющие иерархию, исключая главы, разделы, подразделы и пункты.
- Определение элементов библиографии. Здесь формируются иерархии, включающие главы, разделы, подразделы и пункты.
- Работа со специализированными элементами. В их число входят формулы, таблицы, рисунки и списки, которые требуют особого представления.

Таким образом, можно отметить, что различные структурные единицы взаимодействуют друг с другом в рамках единой иерархической структуры, распределяя условную вероятность в логической модели:

$$p(y | x, \theta) = \frac{\exp(\theta^T \varphi(x, y))}{\sum_{y' \in Y} \exp(\theta^T \varphi(x, y'))}$$

Полученное выражение описывает условное распределение классов $y \in Y$ при фиксированном классе x . Вектор признаков $\varphi(x, y)$ описывает представление объекта, а параметры модели θ задают веса, оптимизируемые в процессе обучения. Нормированный максимум по элементу играет роль функции распределения вероятностей, обеспечивая выполнение условия $\sum_{y \in Y} p(y | x, \theta) = 1$. Следует отметить, что максимизация правдоподобия в данной модели сводится к задаче минимизации функции потерь вида:

$$L(\theta) = - \sum_{i=1}^n \log p(y_i | x_i, \theta),$$

что делает метод особенно удобным для практического применения в задачах классификации и структурированного представления.

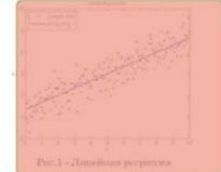


Рис.1 - Линейная регрессия

© Автор(ы), 2025. Все права защищены.

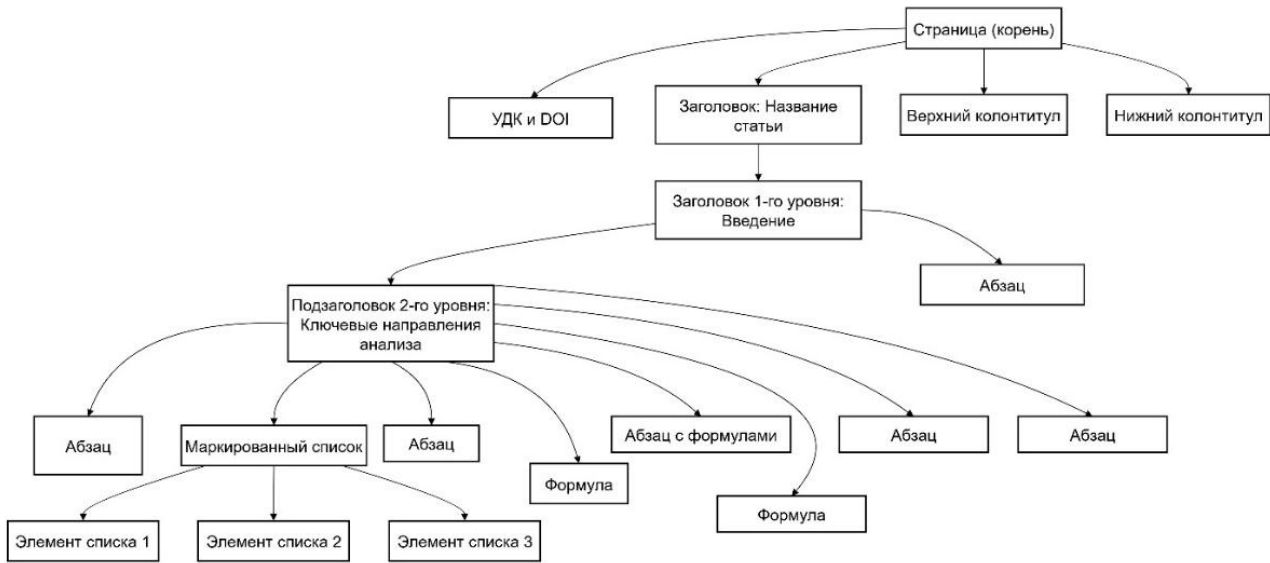
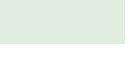
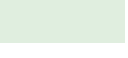
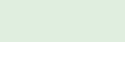
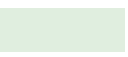


Рис. 1. Пример разбиения документа на блоки методом HDSA и построение дерева

Таблица 1. Пример HDSA таблицы анализа документа

Цвет элемента на схеме	Название элемента	Уровень вложенности	Горизонтальный вес (иерархический приоритет)	Родитель → Дочерние элементы
#e5eefe	Верхний колонтитул	0	0	Родитель (Корень) – Дочерний элемент отсутствует
#c3aeef	УДК и DOI	1	0	Родитель (Корень) – Дочерний элемент отсутствует
#fff6d9	Заголовок документа	2	0	Родитель (Корень) – Дочерний элемент заголовок 1-го уровня

Цвет элемента на схеме	Название элемента	Уровень вложенности	Горизонтальный вес (иерархический приоритет)	Родитель → Дочерние элементы
#e7dfee 	Заголовок 1-го уровня	3	1	Родитель (Заголовок документа) – Дочерние элементы абзац и заголовок 2-го уровня
#e0eedf 	Абзац	4	2	Родитель (Заголовок 1-го уровня) – Дочерний элемент отсутствует
#feedd9 	Заголовок 2-го уровня	5	2	Родитель (Заголовок 1-го уровня) – Дочерний элемент отсутствует
#e0eedf 	Абзац	6	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент отсутствует
#9cc7e9 	Маркированный список	7	3	Родитель (Заголовок 2-го уровня) – Дочерние элементы – элементы списка 1,2 и 3
#e0eedf 	Абзац	11	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент отсутствует
#d0b6ff 	Формула блочная	12	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент отсутствует
#84a882 	Абзац со строчными формулами	13	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент может быть представлен строчными формулами при дальнейшей декомпозиции
#d0b6ff 	Формула блочная	14	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент отсутствует
#e0eedf 	Абзац	15	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент отсутствует
#f4b4a8 	Изображение с подписью	17	3	Родитель (Заголовок 2-го уровня) – Дочерний элемент отсутствует
#ffeedc 	Нижний колонтитул	18	0	Родитель (Корень) – Дочерний элемент нумерация страницы

Подобное табличное представление удобно масштабируется на более сложные документы с большим числом элементов и связей. Каждому типу блока можно присвоить уникальный индекс, что обеспечивает компактное и формализованное описание структуры документа.

1.3. Предлагаемая модель. В отличие от традиционного применения метода HDSA для описания структуры, в настоящей работе он используется как основа для построения модели

оценки реальных документов различных предметных областей. На базе формализма теории графов вводятся количественные характеристики – глубина иерархии, показатели информативности блоков, меры посредничества и матрицы смежности. Эти характеристики позволяют выявить устойчивые структурные закономерности и использовать их для синтеза формализованных JSON-шаблонов [11, 12], интегрируемых в процессы генерации текста большими языковыми моделями [13, 14].

В настоящем исследовании модель, построенная по методу Hierarchical Document Structure Analysis (HDSA), основывается на ряде мер, заимствованных из теории графов:

1. Средняя глубина иерархии графа. Показатель глубины иерархии используется для количественной оценки степени вложенности структурных блоков текста и характеризует общую иерархическую сложность структуры документа. Средняя глубина графа не тождественна максимальному уровню заголовка, однако при отсутствии аномально глубоких ветвей может использоваться как устойчивая оценка эффективной глубины структуры документа.

Неясно, каким образом определяется принадлежность блока контента к той или иной макрогруппе при расчёте меры разнообразия блоков документа.

Средняя глубина иерархии для всего графа $G = (V, E)$ определяется формулой:

$$H(G) = \frac{1}{|V|} \sum_{v \in V} h(v)$$

где $h(v)$ – глубина вершины v , определяемая числом уровней вложенности до корневого элемента. $|V|$ – это мощность множества V , т. е. количество элементов этого множества.

2. Мера разнообразия блоков документа. Для описания общего состава документа D вводится вектор V_c присутствия типов контента в документе по макрогруппам:

$$V_c(D) = (n_{\text{math}}, n_{\text{struct}}, n_{\text{tabular}}, n_{\text{teorethical}}, n_{\text{visual}})$$

где каждый компонент n_i отражает наличие элементов соответствующего типа, присутствующих в документе. Принадлежность блоков к определенной макрогруппе определяется за счет семантики языка разметки, в данном исследовании языком разметки выступает LaTeX.

Для получения числовой меры разнообразия контента $M_{div}(D)$ вводится следующая функция:

$$M_{div}(D) = \frac{\text{число присутствующих макрогрупп в документе}}{\text{максимальное количество макрогрупп в векторе}}$$

- $M_{div}(D) \approx 0$ – документ однотипный (например, полностью текстовый);
- $M_{div}(D) \approx 1$ – документ содержит сбалансированное количество элементов разных типов, т.е. является разнообразным.

Согласно этой мере, документы условно можно разделить на 4 класса, представленные в таблице 2.

3. Мера объема текста в документе.

Для каждой вершины $v \in V$ определяются:

$$M_{total}(v) = \text{общее количество токенов в элементе}$$

$$M_{text}(v) = \text{количество текстовых токенов в элементе}$$

На их основе вводится отношение:

$$R_{\text{text}}(v) = \frac{M_{\text{text}}(v)}{M_{\text{total}}(v)},$$

которое характеризует долю текстовой информации в структуре конкретного элемента. Среднее значение этого отношения по всем вершинам может рассматриваться как показатель текстоцентричности документа:

$$\bar{R}_{\text{text}}(D) = \frac{1}{|V|} \sum_{v \in V} R_{\text{text}}(v)$$

Таблица 2. Примеры классов документа согласно мере разнообразия

Диапазон $M_{\text{div}}(D)$	Интерпретация
0.0 – 0.33	Низкое разнообразие
0.34 – 0.66	Среднее разнообразие
0.67 – 1	Высокое разнообразие

4. Мера центральности посредничества узлов (betweenness centrality). Эта мера необходима для оценки влияния отдельных типов вершин (заголовков, абзацев, формул и т.д.) на связанность графа документа. Она характеризует степень участия вершины v в кратчайших путях между другими вершинами и вычисляется по формуле:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}},$$

где σ_{st} – количество кратчайших путей между вершинами s и t , а $\sigma_{st}(v)$ – число таких путей, проходящих через вершину v .

В терминах анализа структуры документа высокая центральность посредничества означает, что элемент играет роль связующего звена между различными частями текста (например, раздел, объединяющий несколько подтем). Именно таким образом в рамках одного документа численно определяются элементы с высокой степенью посредничества, такие, как заголовки различных уровней, от элементов с низкой степенью посредничества, например, абзацы. В дополнение к основным количественным мерам оценки структуры документа рассматриваются вспомогательные показатели, в частности, показатель количества структурных блоков позволит оценить численно элементы графа.

2. Суть эксперимента. Экспериментальная часть исследования была направлена на проверку применимости модели и метода HDSA для автоматического извлечения структурных метрик документов и формирования на их основе формализованных JSON-шаблонов [15, 16]. Эксперименты проводились на корпусах объёмом 50 страниц для трёх предметных областей (математика, юриспруденция, экология), подготовленных в формате LaTeX, что позволило оценить устойчивость подхода при работе с текстами различной структурной организации. Выбор данных областей обеспечил анализ как универсальных, так и предметно-специфических структурных паттернов (рисунок 2).

В рамках исследования был сформирован специализированный корпус документов, содержащий 5 документов для каждой предметной области, участвующей в экспериментальном анализе.

Список документов по математике:

- Письменный Д. Т. «Конспект лекций по высшей математике»;
- Краснов М. Л., Киселёв А. И., Макаренко Г. И. и др. «Вся высшая математика»;
- Дейзенрот М. П., Фейзал А. А., Он Ч. С. «Математика для машинного обучения»;

- Демидович Б. П., Кудрявцев В. А. «Краткий курс высшей математики»;
- Будаков Б. М., Фомин С. В. «Кратные интегралы и ряды».



Рис. 2. Алгоритм проведения эксперимента

Список документов по юриспруденции:

- Мамедова С.Н. «Курс лекций по дисциплине правоведение»;
- Мелехин А.В. «Теория государства и права: Учебник с учебно-методическими материалами»;
- Исакова В. Б. «Основы права: учебник для неюридических вузов и факультетов»;
- Клишаса А.А. «Теория государства и права: учебник – Российский университет дружбы народов, Юридический институт»;
- Сулейменов М. «Право как система».

Список документов по экологии:

- Вороного А.А., Ситникова С.В. «Экология конспект лекций»;
- Гашев С. Н., «Конспекты лекций по системной экологии: учебное пособие»;
- Курмаева Н. М., Смирнов Д.Г. «Краткий курс экологии»;
- Панин В.Ф., Сечин А.И., Федосова В.Д. «Экология общеэкологическая концепция биосферы и экономические рычаги преодоления глобального экологического кризиса; обзор современных принципов и методов защиты биосферы»;
- Криксунов Е.А., Пасечник В.В. «Экология 10, 11 класс».

Результаты оценки каждого документа, полученные с применением предложенной модели, представлены в таблице 3.

2.1. Оценка документов из предметной области. На этапе предварительной визуальной оценки было выдвинуто предположение, что наибольшей структурной и содержательной

сложностью будут характеризоваться материалы по экологии вследствие их междисциплинарного характера и сочетания различных типов контента. Высокая структурная сложность также ожидается для математических материалов из-за частой смены текстовых и формализованных блоков. В то же время документы по юриспруденции предполагаются наиболее однотипными и текстоцентричными, что обусловлено преобладанием нормативного изложения.

Таблица 3. Результаты оценки документов

Предметная область	$H(G)$	$M_{div}(D)$	$R_{text}(D)$	Элементы с наибольшим $C_B(v)$	Количество блоков в графе
Математика (Письменный)	2,75	0,83	0,58	Заголовки 1 и 2 уровня	617
Математика (Краснов)	2,81	0,83	0,62	Заголовки 1 и 2 уровня	751
Математика (Дейзенрот)	3,1	0,83	0,7	Заголовки 1 и 2 уровня	433
Математика (Демидович)	2	0,83	0,51	Заголовки 1 и 2 уровня	303
Математика (Будак)	2,75	0,83	0,66	Заголовки 1 и 2 уровня	804
Юриспруденция (Мамедова)	2,81	0,33	1	Заголовки 1 и 2 уровня	239
Юриспруденция (Мелехин)	2,9	0,33	1	Заголовки 1 и 2 уровня	283
Юриспруденция (Клишаса)	2,97	0,33	1	Заголовки 1 и 2 уровня	237
Юриспруденция (Сулейменов)	2,55	0,33	1	Заголовки 1 и 2 уровня	172
Юриспруденция (Исакова)	2,67	0,33	1	Заголовки 1 и 2 уровня	258
Экология 1 (Вороной)	2,92	0,67	0,69	Заголовки 1 и 2 уровня	202
Экология 2 (Панин)	2,65	0,67	0,56	Заголовки 1 и 2 уровня	193
Экология 3 (Криксунов)	3,01	0,67	0,67	Заголовки 1 и 2 уровня	221
Экология 4 (Гашев)	2,89	0,67	0,72	Заголовки 1 и 2 уровня	284
Экология 5 (Курмаева)	3,41	0,67	0,84	Заголовки 1 и 2 уровня	306

Анализ данных таблицы 3 показывает, что для документов всех трёх предметных областей средняя глубина иерархического графа стремится к значению около 3, что согласуется с визуальной оценкой структуры и используемыми системами заголовков. Для всех документов также характерны высокие значения центральности узлов, соответствующих заголовкам. При этом наибольшая структурная и содержательная вариативность наблюдается в материалах по экологии и математике, что отражается в значениях показателя разнообразия и количестве блоков в графе. Документы по юриспруденции, напротив, отличаются выраженной текстоцентричностью и более однородной структурой.

2.3. Формирование шаблона на основе мер. В ходе эксперимента были сформулированы правила формирования JSON-шаблона, используемого при генерации структурированного документа. Шаблон включает функциональные блоки *metrics*, *structure* и *style*, а также параметр выходного формата *output_format*. Блок *metrics* содержит эталонные значения метрик и их интерпретации, *structure* задаёт иерархию заголовков и допустимые типы контентных блоков, а *style* определяет требования к стилю и профессиональному языку результирующего документа.

Листинг 1. Пример сокращенного шаблона, описывающий правила формирования на языке JSON

```
{
  "metrics": {
    "H(G)": {
      "value": 3.0,
      "interpretation": "Эффективная глубина иерархии документа"},
    "M_div": {
      "value": 0.67,
      "interpretation": "Разнообразие блоков контента"}},
    "R_text": {
      "value": 0.69,
      "interpretation": "Доля текстоцентричного контента"}},
    "structure": {
      "headings": [
        {"level": 1, "centrality": "high"},
        {"level": 2, "centrality": "medium"},
        {"level": 3, "centrality": "low"}],
      ...
    "blocks": {
      "text": "text text text",
      "inline_equation": "text $inline_equation$ text",
      "block_equation": "$$block_equation$$",
      "list": "- item"}},
    "style": {
      "domain_language": "professional",
      "tone": "academic",
      "verbosity": "moderate"
    },
    "output_format": "markdown"
  }
```

Анализ сформированных шаблонов показал, что их базовая структурная схема не зависит от предметной области и остаётся инвариантной. Различия между шаблонами проявляются не на уровне иерархии, а на уровне состава и разнообразия используемых контентных блоков. Для материалов по математике и экологии характерно расширение набора блоков и увеличение доли нетекстовых элементов, тогда как шаблоны для юриспруденции сохраняют преимущественно текстоцентричную структуру.

2.4. Результаты и обсуждение. Для обеспечения статистической устойчивости эксперимента на первом этапе было выполнено 10 генераций без использования шаблона, что позволило зафиксировать базовый уровень вариативности и структуры формируемого текста. На втором этапе были проведены 10 генераций с применением шаблона. В качестве большой языковой модели используется модель Mistral Large через API со стандартным значением температуры равное 0,3. Сводные результаты обеих серий экспериментов представлены в таблице 4, на которую далее даётся ссылка при анализе полученных различий.

Анализ результатов, представленных в таблице 4, показывает, что использование шаблона не оказывает влияния на показатели $M_{div}(D)$ и $R_{text}(D)$, что обусловлено спецификой предметной области (юриспруденция), характеризующейся доминированием однородных текстовых блоков и отсутствием выраженного структурного разнообразия. В то же время

применение шаблона приводит к снижению вариативности графовой структуры документа, что выражается в стабилизации значения $H(G)$ и уменьшении разброса количества блоков.

Таблица 4. Результат генерации для конспекта по праву

$H(G)$	$M_{div}(D)$	$R_{text}(D)$	Количество блоков в графе
Эталонные значения			
2,8	0,33	1	В рамках эксперимента не используется
10 генераций без шаблона			
2,63	0,33	1	46
2,31	0,33	1	36
2,13	0,33	1	23
2,2	0,33	1	40
2,86	0,33	1	46
1,98	0,33	1	51
2,06	0,33	1	33
3,27	0,33	1	52
3,27	0,33	1	45
2,2	0,33	1	40
2,13	0,33	1	23
10 генераций с использованием шаблона			
3,07	0,33	1	28
3,04	0,33	1	21
3,0	0,167	1	20
3,08	0,33	1	23
3,0	0,33	1	20
3,0	0,167	1	20
3,06	0,33	1	33
2,7	0,33	1	32
3,09	0,33	1	22
3,03	0,33	1	21

В отличие от результатов, полученных для лекций по праву, эксперименты по математике (таблица 5) демонстрируют выраженное влияние шаблона на формирование структуры документа. Аналогично предыдущему эксперименту, применение шаблона приводит к снижению структурной вариативности, стабилизации значения $H(G)$ и сокращению разброса числа блоков, обеспечивая более компактную и воспроизводимую организацию документа. В генерациях без шаблона наблюдается значительный разброс характеристик и вырожденные случаи с утратой математической модальности. Дополнительно выявлена устойчивая тенденция к избыточному дроблению содержания на структурные блоки, что указывает на повышенную фрагментацию генерируемого документа.

Таблица 5. Результат генерации для конспекта по математике

$H(G)$	$M_{div}(D)$	$R_{text}(D)$	Количество блоков в графе
Эталонные значения			
2,75	0,83	0,58	В рамках эксперимента не используется
10 генераций без шаблона			
3,51	0,55	0,81	75
2,47	0,33	1	15
3,99	0,55	0,79	91
2	0,167	1	4
3,51	0,55	0,81	75
2,89	0,55	0,62	21
3,77	0,55	0,93	70
3,48	0,55	0,79	75
3,85	0,55	0,88	89
10 генераций с использованием шаблона			
3,51	0,67	0,76	52
3,404	0,55	0,77	52
3,36	0,55	0,86	58
3,39	0,67	0,89	57
3,45	0,55	0,89	49
2,84	0,55	0,91	51
2,75	0,67	0,73	41
3,67	0,55	0,65	54
3,68	0,55	0,67	54
3,63	0,55	0,69	35

Результаты экспериментов для экологической предметной области подтверждают тенденцию, ранее выявленную для математики: независимо от использования шаблона модель склонна к избыточной фрагментации документа, что проявляется в увеличении числа блоков и значений $H(G)$. При этом показатели текстоцентричности остаются близкими к единице, отражая текстоцентричный характер экологического материала и обуславливая низкие значения показателя разнообразия. Использование шаблона снижает структурную вариативность и устраняет вырожденные случаи, однако не устраняет саму тенденцию к фрагментации, что указывает на систематическую особенность генеративного механизма модели (таблица 6).

Заключение. В работе предложена и апробирована модель анализа и использования иерархической структуры документа на основе метода HDSA и теории графов, позволяющая формировать формализованные JSON-шаблоны. Эксперименты на материалах по юриспруденции, математике и экологии показали, что применение шаблонов снижает структурную вариативность генерации и стабилизирует графовую организацию документа без потери полноты содержания. Одновременно выявлена систематическая тенденция к избыточной фрагментации без использования шаблона в структурно насыщенных предметных областях, что определяет направления дальнейшей доработки шаблонов и метода.

Таблица 6. Результат генерации для конспекта по экологии

$H(G)$	$M_{div}(D)$	$R_{text}(D)$	Количество блоков в графе
Эталонные значения			
2,92	0,67	0,69	В рамках эксперимента не используется
10 генераций с использованием шаблона			
3,16	0,33	1	69
3,11	0,33	1	63
2,46	0,167	1	24
2,43	0,167	1	30
2,41	0,167	1	22
2,7	0,167	1	30
3,36	0,55	0,99	124
3,25	0,33	1	87
3,13	0,33	1	71
3,08	0,33	1	72
10 генераций без шаблона			
3,64	0,55	1	45
2,58	0,55	1	35
3,09	0,55	0,98	63
3,12	0,33	1	67
3,13	0,33	1	46
2,71	0,55	1	41
3,06	0,33	1	53
2,88	0,33	1	32
3,12	0,33	1	49
2,98	0,33	1	54

Список источников

1. Liang X., Wang H., Wang Y. et al. Controllable text generation for large language models: a survey. Preprint arXiv:2408.12599, 2024, DOI: 10.48550/arXiv.2408.12599.
2. Liu Y., Li D., Wang K. et al. Are LLMs good at structured outputs? A benchmark for evaluating structured output capabilities in LLMs. Information Processing & Management, 2024, vol. 61, no. 5, art. no. 103809, DOI: 10.1016/j.ipm.2024.103809.
3. Lu Y., Li H., Cong X. et al. Learning to generate structured output with schema reinforcement learning. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025, pp. 4905–4918.
4. Xing H., Cheng C., Gao F. et al. DocHieNet: a large and diverse dataset for document hierarchy parsing. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Y. Al-Onaizan, M. Bansal, Y.-N. Chen (eds.), 2024, pp. 1129–1142, DOI:10.18653/v1/2024.emnlp-main.65.
5. Pfizmann B., Auer C., Dolfi M. et al. DocLayNet: a large human-annotated dataset for document-layout analysis. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), 2022, pp. 3530–3540, DOI: 10.1145/3534678.3539043.
6. Wang J., Krumdick M., Tong B. et al. A graphical approach to document layout analysis. Preprint arXiv:2308.02051, 2023, DOI: 10.48550/arXiv.2308.02051.
7. Jiovannini S., Marinai S. A survey on reading order, table of contents, and structure extraction in document analysis. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2025, pp. 7585–7594.

8. Rausch J., Martinez O., Bissig F. et al. DocParser: hierarchical document structure parsing from renderings. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, vol. 35, no. 5, pp. 4328–4338, DOI: 10.1609/aaai.v35i5.16558.
9. Wang J., Hu K., Zhong Z. et al. Detect-order-construct: a tree construction-based approach for hierarchical document structure analysis. Pattern Recognition, 2024, vol. 156, art. no. 110836, DOI: 10.1016/j.patcog.2024.110836.
10. Yang J., Jiang D., He L. et al. StructEval: benchmarking LLMs' capabilities to generate structural outputs. Preprint arXiv:2505.20139, 2026, DOI: 10.48550/arXiv.2505.20139
11. Chen Y., Liu R., Zheng J. et al. Graph-based document structure analysis. Preprint arXiv:2502.02501, 2025, DOI: 10.48550/arXiv.2502.02501
12. Kim D. Guided JSON with LLMs: from raw PDFs to structured intelligence. Available at: <https://medium.com/@kimdoil1211/structured-output-with-guided-json-a-practical-guide-for-llm-developers-6577b2eee98a> (accessed: 02/03/2025).
13. Geng S., Cooper H., Moskal M. et al. Generating structured outputs from language models: benchmark and studies. Preprint arXiv:2501.10868, 2025, DOI: 10.48550/arXiv.2501.10868.
14. Joshi I., Agarwal B., Rojkova V. Think inside the JSON: reinforcement strategy for strict LLM schema adherence. Preprint arXiv:2502.14905, 2025, DOI:10.48550/arXiv.2502.14905.
15. Balaskas G., Papadopoulos H., Pappa D. et al. A framework for domain-specific dataset creation and adaptation of large language models. Computers, 2025, vol. 14, no. 5, art. no. 172, DOI: 10.3390/computers14050172.
16. Taveekitworachai P., Dewantoro M.F., Xia Y. et al. BenchING: a benchmark for evaluating large language models in following structured output format instruction in text-based narrative game tasks. IEEE Transactions on Games, 2025, vol. 17, no. 3, pp. 665–675, DOI: 10.1109/TG.2025.3529117.

Исмагулов Милан Ерикович. Аспирант по направлению 2.3.1. – Системный анализ, управление и обработка информации, статистика. Основные направления исследований автора: мультимодальное преобразование данных, машинное обучение, методы обработки данных моделями машинного обучения. AuthorID: 1301221, SPIN: 6808-8045, ORCID: 0009-0007-3280-5259, m_ismagulov@ugrasu.ru. 628012, Ханты-Мансийский автономный округ - Югра, г. Ханты-Мансийск, ул. Чехова, д.16.

UDC 004.89

DOI:10.25729/ESI.2026.43.3.016

A Hierarchical model of document structure for the task of generating lecture summaries

Milan E. Ismagulov

Yugra State University, Russia, Khanty-Mansiysk, m_ismagulov@ugrasu.ru

Abstract. This article proposes a document structure description model based on Hierarchical Document Structure Analysis (HDSA) methods and the formalism of graph theory. Within the proposed approach, a document is considered as a hierarchical system of interconnected structural elements, including headings of different levels, text blocks, tables, formulas, images, and other types of content. Each element is interpreted as a graph vertex, while nesting and logical dependency relationships between elements are described using directed edges of the graph structure.

The application of the proposed model to a corpus of documents from various subject areas made it possible to identify stable patterns in the organization of textual material characteristic of specific document types. Based on the obtained characteristics, a set of JSON templates was developed to describe typical document structures for different subject areas, including technical, legal, and scientific and educational materials.

Each template contains a formalized description of the document structure, including permissible heading levels, constraints on content types, and parameters of stylistic and compositional formatting. Additionally, a template may include information on the preferred data presentation format, rules for generating tables and formulas, and specialized textual constructions. The use of the JSON format provides a machine-oriented representation of the document structure and enables the integration of templates into software systems for automatic text generation. The developed templates are intended for use in text generation by large language models (LLMs). Within the proposed approach, a template serves as a formalized structural constraint that defines the permissible organization of the resulting document and the sequence in which its sections are generated. This makes it possible to reduce the structural variability of generated text, improve the consistency of interrelated sections, and ensure that the resulting document complies with the requirements of a specific subject area.

The proposed approach is aimed at implementing controlled document generation, in which a large language model generates content within a predefined structural scheme. This provides more stable document generation quality, reduces the likelihood of violations in the logical organization of the text, and improves the reproducibility of generation results. Furthermore, the use of structural templates makes it possible to adapt the generation process to different document types without modifying the architecture of the language model itself.

Keywords: hierarchical document structure, graph theory, JSON templates, large language models (LLMs), structured text generation, structural patterns, document representation model

References

1. Liang X., Wang H., Wang Y. et al. Controllable text generation for large language models: a survey. Preprint arXiv:2408.12599, 2024, DOI: 10.48550/arXiv.2408.12599.
2. Liu Y., Li D., Wang K. et al. Are LLMs good at structured outputs? A benchmark for evaluating structured output capabilities in LLMs. Information Processing & Management, 2024, vol. 61, no. 5, art. no. 103809, DOI: 10.1016/j.ipm.2024.103809.
3. Lu Y., Li H., Cong X. et al. Learning to generate structured output with schema reinforcement learning. Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2025, pp. 4905–4918.
4. Xing H., Cheng C., Gao F. et al. DocHieNet: a large and diverse dataset for document hierarchy parsing. Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Y. Al-Onaizan, M. Bansal, Y.-N. Chen (eds.), 2024, pp. 1129–1142, DOI:10.18653/v1/2024.emnlp-main.65.
5. Pfitzmann B., Auer C., Dolfi M. et al. DocLayNet: a large human-annotated dataset for document-layout analysis. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '22), 2022, pp. 3530–3540, DOI: 10.1145/3534678.3539043.
6. Wang J., Krumdick M., Tong B. et al. A graphical approach to document layout analysis. Preprint arXiv:2308.02051, 2023, DOI: 10.48550/arXiv.2308.02051.
7. Jiovannini S., Marinai S. A survey on reading order, table of contents, and structure extraction in document analysis. Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, 2025, pp. 7585–7594.
8. Rausch J., Martinez O., Bissig F. et al. DocParser: hierarchical document structure parsing from renderings. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, vol. 35, no. 5, pp. 4328–4338, DOI: 10.1609/aaai.v35i5.16558.
9. Wang J., Hu K., Zhong Z. et al. Detect-order-construct: a tree construction-based approach for hierarchical document structure analysis. Pattern Recognition, 2024, vol. 156, art. no. 110836, DOI: 10.1016/j.patcog.2024.110836.
10. Yang J., Jiang D., He L. et al. StructEval: benchmarking LLMs' capabilities to generate structural outputs. Preprint arXiv:2505.20139, 2026, DOI: 10.48550/arXiv.2505.20139
11. Chen Y., Liu R., Zheng J. et al. Graph-based document structure analysis. Preprint arXiv:2502.02501, 2025, DOI: 10.48550/arXiv.2502.02501
12. Kim D. Guided JSON with LLMs: from raw PDFs to structured intelligence. Available at: <https://medium.com/@kimdoil1211/structured-output-with-guided-json-a-practical-guide-for-llm-developers-6577b2eee98a> (accessed: 02/03/2025).
13. Geng S., Cooper H., Moskal M. et al. Generating structured outputs from language models: benchmark and studies. Preprint arXiv:2501.10868, 2025, DOI: 10.48550/arXiv.2501.10868.
14. Joshi I., Agarwal B., Rojkova V. Think inside the JSON: reinforcement strategy for strict LLM schema adherence. Preprint arXiv:2502.14905, 2025, DOI:10.48550/arXiv.2502.14905.
15. Balaskas G., Papadopoulos H., Pappa D. et al. A framework for domain-specific dataset creation and adaptation of large language models. Computers, 2025, vol. 14, no. 5, art. no. 172, DOI: 10.3390/computers14050172.
16. Taveekitworachai P., Dewantoro M.F., Xia Y. et al. BenchING: a benchmark for evaluating large language models in following structured output format instruction in text-based narrative game tasks. IEEE Transactions on Games, 2025, vol. 17, no. 3, pp. 665–675, DOI: 10.1109/TG.2025.3529117.

Ismagulov Milan Erikovich. Postgraduate student in the field 2.3.1 – System Analysis, Management and Information Processing, Statistics. The author's main research areas: multimodal data conversion, machine learning, data processing methods using machine learning models. AuthorID: 1301221, SPIN: 6808-8045, ORCID: 0009-0007-3280-5259. m_ismagulov@ugrasu.ru. 628012, Khanty-Mansi Autonomous Okrug – Yugra, Khanty-Mansiysk, Chekhov St., 16.

Статья поступила в редакцию 27.03.2026; одобрена после рецензирования 01.06.2026; принята к публикации 05.08.2026.

The article was submitted 03/27/2026; approved after reviewing 06/01/2026; accepted for publication 08/05/2026.

